UNIVERSITY OF CAMBRIDGE

REACH

# Efficient Bayesian Modelling of Time Dependent and Transient RFI in 21cm Experiments

6th Global 21cm Workshop

Dominic Anstey

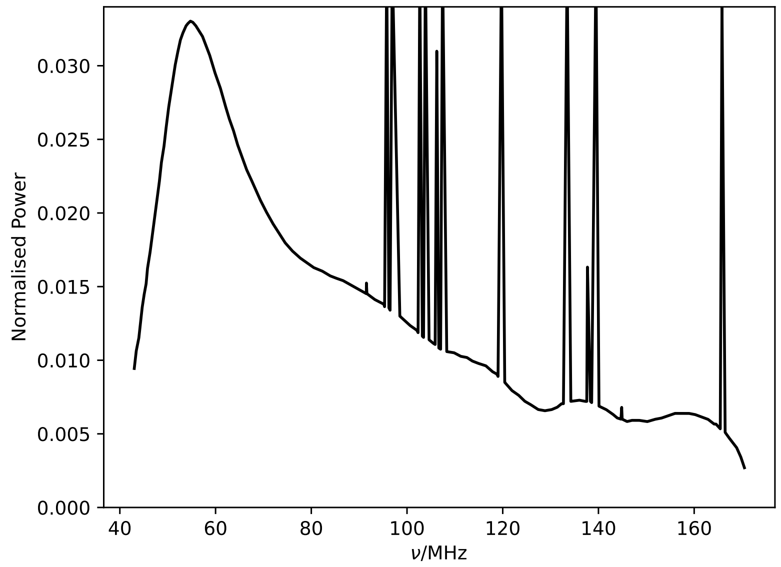
In collaboration with Sam Leeney

# The Challenge of RFI

The 21cm signal is predicted to fall in the ~50-150MHz band. Directly over FM radio band.

- Experiments in remote locations

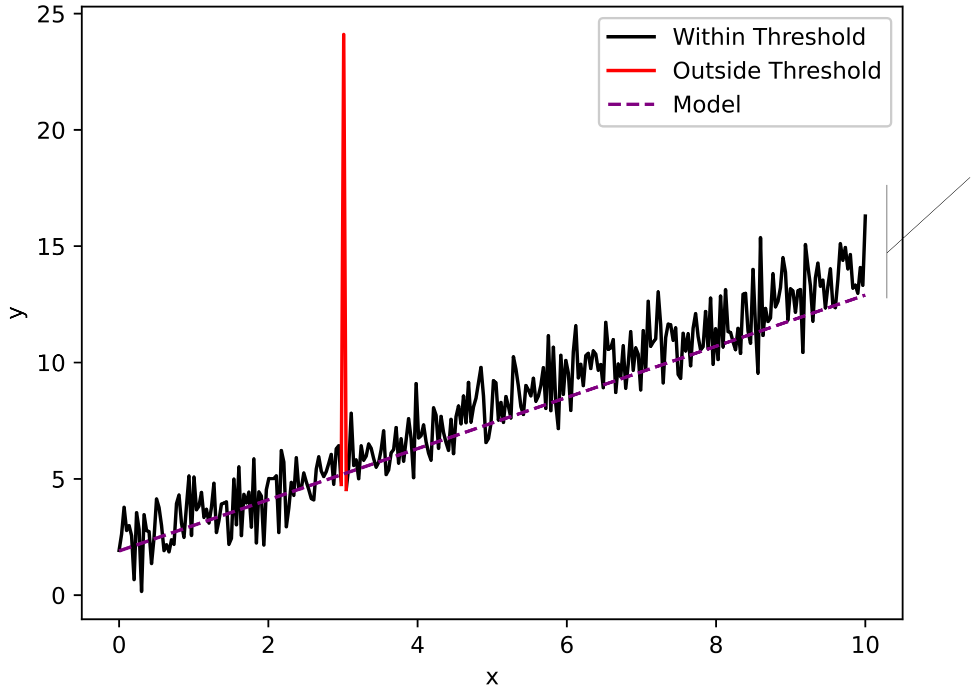
- RFI cleaning




Based on Leeney et al. 2023 (arXiv:2211.15448)

# Bayesian Anomaly Flagging

Leeney et al. 2023 (arXiv:2211.15448)




- Calculate likelihoods of each point

- Compare to a threshold

- If exceeds the threshold, include the point's likelihood, downweighted by a penalty

- If below the threshold, add a fixed penalty

# Bayesian Anomaly Flagging

Leeney et al. 2023 (arXiv:2211.15448)

$$\log \mathcal{L}_i = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\left(\frac{\mathcal{D}_i - \mathcal{M}_i(\boldsymbol{\theta})}{\sigma}\right)^2$$

$$\log \mathcal{L} = \sum_i \begin{cases} \log \mathcal{L}_i + \log(1-p) & \text{if} \quad \log \mathcal{L}_i + \log(1-p) > \log p - \log \Delta \\ \log(p) & \text{otherwise} \end{cases}$$

Probability of a point being contaminated

Approximate scale of the contamination

# Bayesian Anomaly Flagging

$$\log \mathcal{L}_{ij} = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\left(\frac{\mathcal{D}_{ij} - \mathcal{M}_{ij}\left(\boldsymbol{\theta}\right)}{\sigma}\right)^2$$

$$\log \mathcal{L} = \sum_{ij} \begin{cases} \log \mathcal{L}_{ij} + \log(1-p) & \text{if } \log \mathcal{L}_{ij} + \log(1-p) > \log p - \log \Delta \\ \log(p) & \text{otherwise} \end{cases}$$

# Toy Model

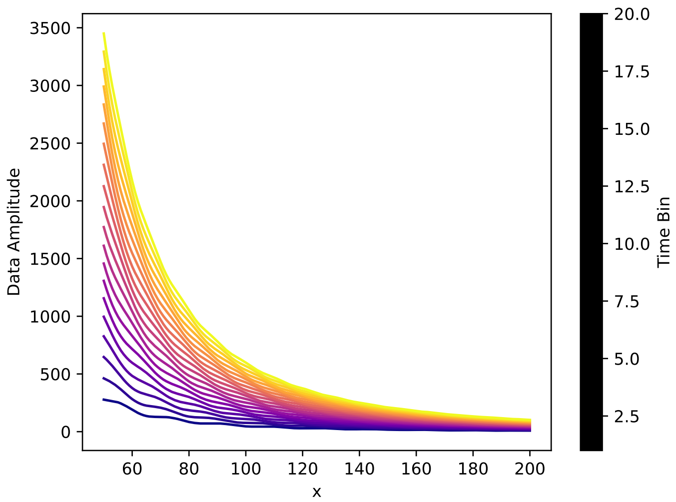
Sinusoid with time-varying parameters, multiplied by a power law

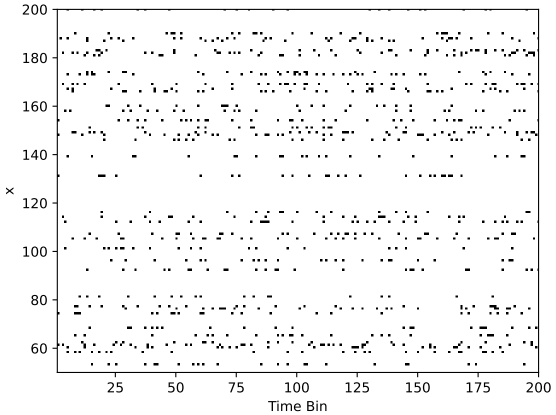
$$\mathcal{D}_{ij} = G_{ij} \times x_i^{-2.55} + \hat{\sigma} + \text{anomalies}$$

$$G_{ij} = \alpha_j \sin\left(\omega_j x_i + \phi_j\right) + \gamma_j$$

With randomly generated values for the sinusoid parameters

$$\mathcal{M}_{ij}\left(\theta\right) = G_{ij} \times x_i^{-\theta}$$



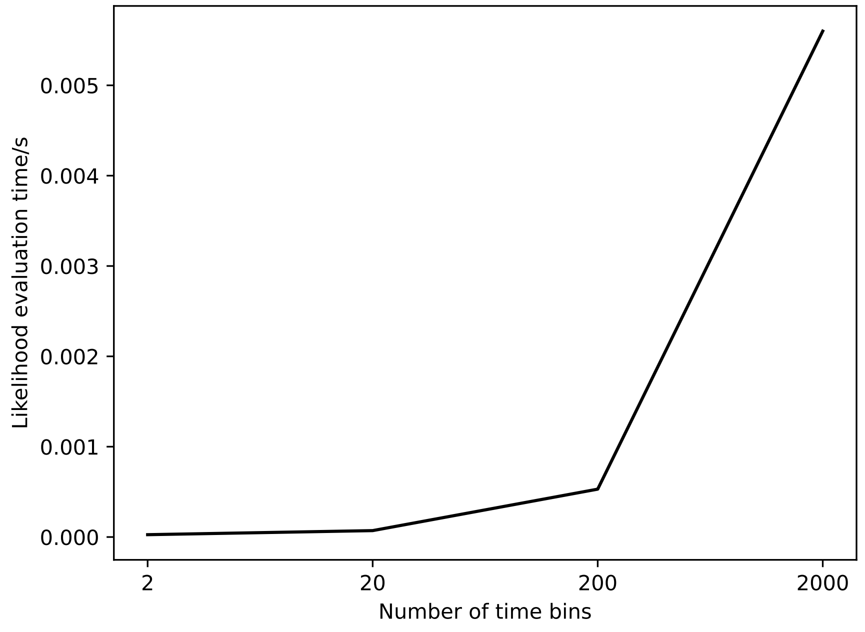



# Computation Time

The term

$$\log \mathcal{L}_{ij} = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\left(\frac{\mathcal{D}_{ij} - \mathcal{M}_{ij}\left(\boldsymbol{\theta}\right)}{\sigma}\right)^2$$

must be evaluated and compared to the threshold for every data bin and time bin. There is no way to factorise the time summation out to speed this calculation.



# Likelihood Reweighting

Gravitational waves

- Payne et al. 2019 (arxiv:1905.05477)
- Romero-Shaw et al. 2019 (arxiv:1905.05477)

# Computation Time

$$\log \mathcal{L} = \sum_{ij} \begin{cases} \log \mathcal{L}_{ij} + \log(1-p) & \text{if } \log \mathcal{L}_{ij} + \log(1-p) > \log p - \log \Delta \\ \log(p) & \text{otherwise} \end{cases}$$

$$\log \mathcal{L}_{ij} = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\left(\frac{\mathcal{D}_{ij} - \mathcal{M}_{ij}(\boldsymbol{\theta})}{\sigma}\right)^2$$

Full likelihood - Slow to evaluate

$$\log \mathcal{L} = \sum_{i} \begin{cases} \log \mathcal{L}_{i} + \log(1-p) & \text{if } \log \mathcal{L}_{i} + \log(1-p) > \log p - \log \Delta \\ \log(p) & \text{otherwise} \end{cases}$$

$$\log \mathcal{L}_{i} = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\left(\frac{\frac{1}{N_t}\sum_j \mathcal{D}_{ij} - \frac{1}{N_t}\sum_j \mathcal{M}_{ij}(\boldsymbol{\theta})}{\sigma}\right)^2$$

Time averaged likelihood - Fast to evaluate

Both likelihoods use the same parameters and will have comparable posteriors

# Likelihood Reweighting

$$\mathcal{P}_{\mathrm{S}}\left(\theta|\mathcal{D},\mathcal{M}_{\mathrm{S}}\right)=\frac{\mathcal{L}_{\mathrm{S}}\left(\mathcal{D}|\theta,\mathcal{M}_{\mathrm{S}}\right)\pi\left(\theta\right)}{\mathcal{Z}_{\mathrm{S}}}$$

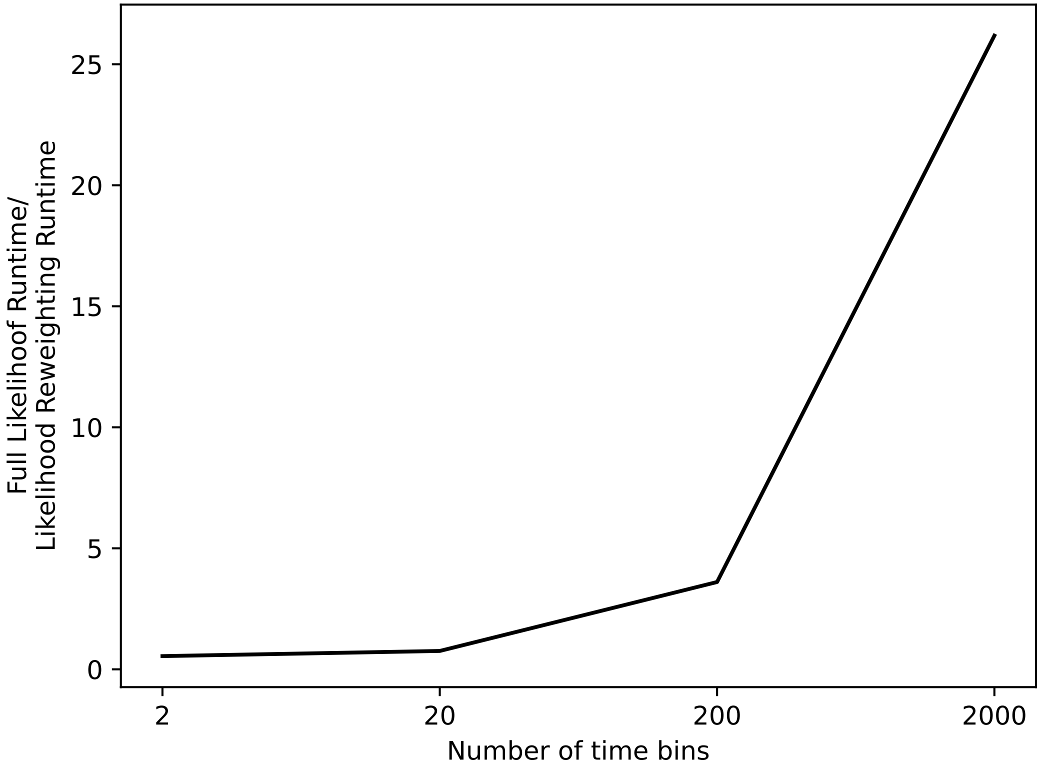
$$\mathcal{P}_{\mathrm{F}}\left(\theta|\mathcal{D},\mathcal{M}_{\mathrm{F}}\right)=\frac{\mathcal{L}_{\mathrm{F}}\left(\mathcal{D}|\theta,\mathcal{M}_{\mathrm{F}}\right)\pi\left(\theta\right)}{\mathcal{Z}_{\mathrm{F}}}$$

$$\pi\left(\theta\right)=\frac{\mathcal{Z}_{\mathrm{F}}\mathcal{P}_{\mathrm{F}}\left(\theta|\mathcal{D},\mathcal{M}_{\mathrm{F}}\right)}{\mathcal{L}_{\mathrm{F}}\left(\mathcal{D}|\theta,\mathcal{M}_{\mathrm{F}}\right)}$$

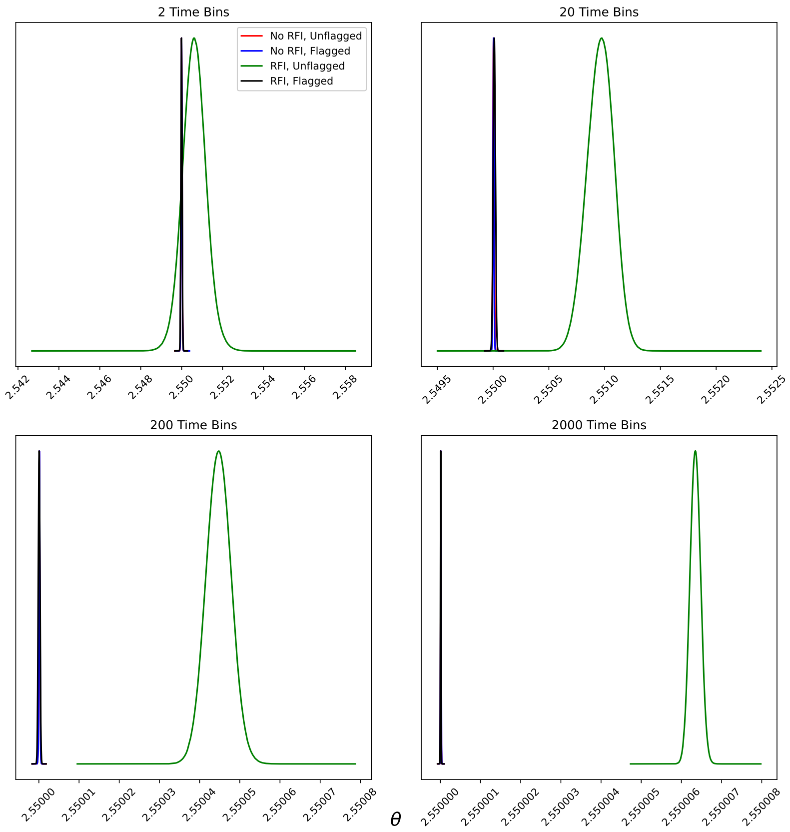
$$\mathcal{P}_{\mathrm{S}}\left(\theta|\mathcal{D},\mathcal{M}_{\mathrm{S}}\right)=\mathcal{P}_{\mathrm{F}}\left(\theta|\mathcal{D},\mathcal{M}_{\mathrm{F}}\right)\frac{\mathcal{L}_{\mathrm{S}}\left(\mathcal{D}|\theta,\mathcal{M}_{\mathrm{S}}\right)}{\mathcal{L}_{\mathrm{F}}\left(\mathcal{D}|\theta,\mathcal{M}_{\mathrm{F}}\right)}\frac{\mathcal{Z}_{\mathrm{F}}}{\mathcal{Z}_{\mathrm{S}}}$$

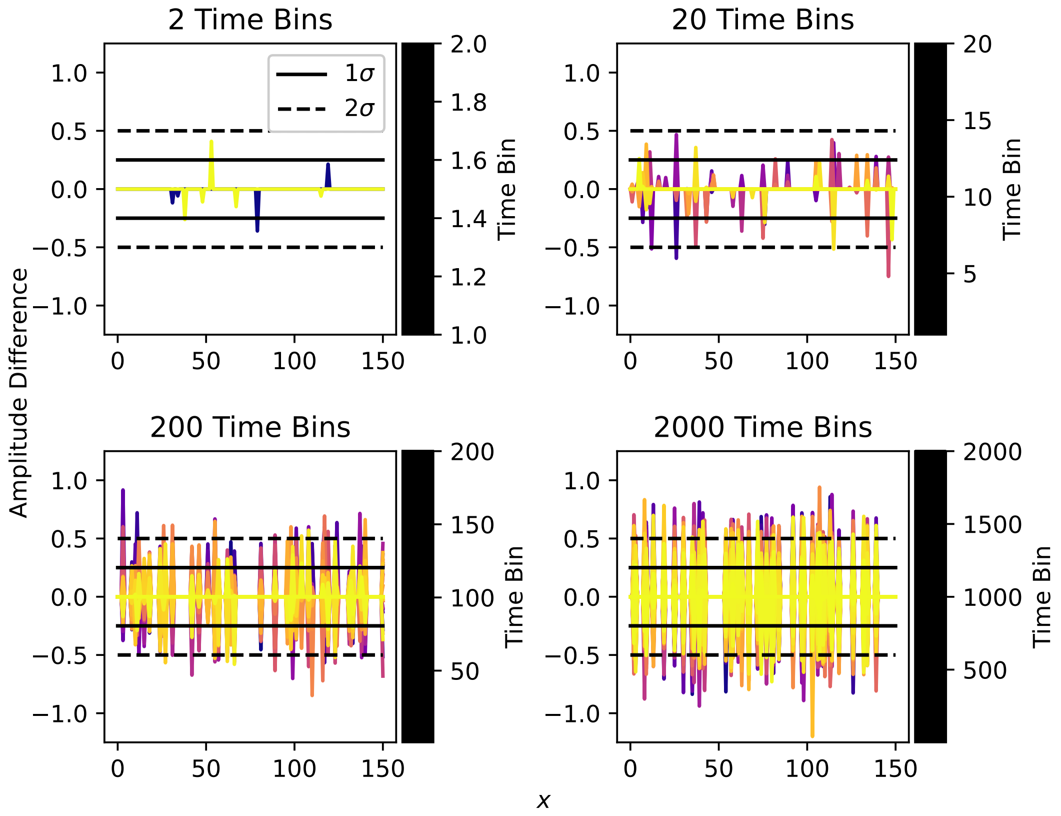
# Results

Ratio of fitting data set using the flagger for Full slow likelihood/likelihood reweighting



# Results


2 Time Bins
No RFI, Unflagged
No RFI, Flagged
RFI, Unflagged
RFI, Flagged
20 Time Bins
200 Time Bins
2000 Time Bins
θ


2 Time Bins
20 Time Bins
200 Time Bins
2000 Time Bins
1σ
2σ
Amplitude Difference
Time Bin
x

# Conclusions

- In data and models with predictable time variation, that variation can be leveraged to constrain the model more accurately.

- Bayesian anomaly flagging can be incorporated in this process

- Issue of runtimes can be resolved with likelihood reweighting

- Works effectively as both a time-sensitive Bayesian RFI compensator and an efficient Bayesian transient flagger