Time-sensitive Anomaly Flagging through Likelihood Reweighting

KICC Focus Meeting on Astrostatistics and Astro-Machine Learning - 2nd June 2023

Dominic Anstey

Bayesian Anomaly Flagging

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

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

Bayesian Anomaly Flagging

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

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

Toy Model

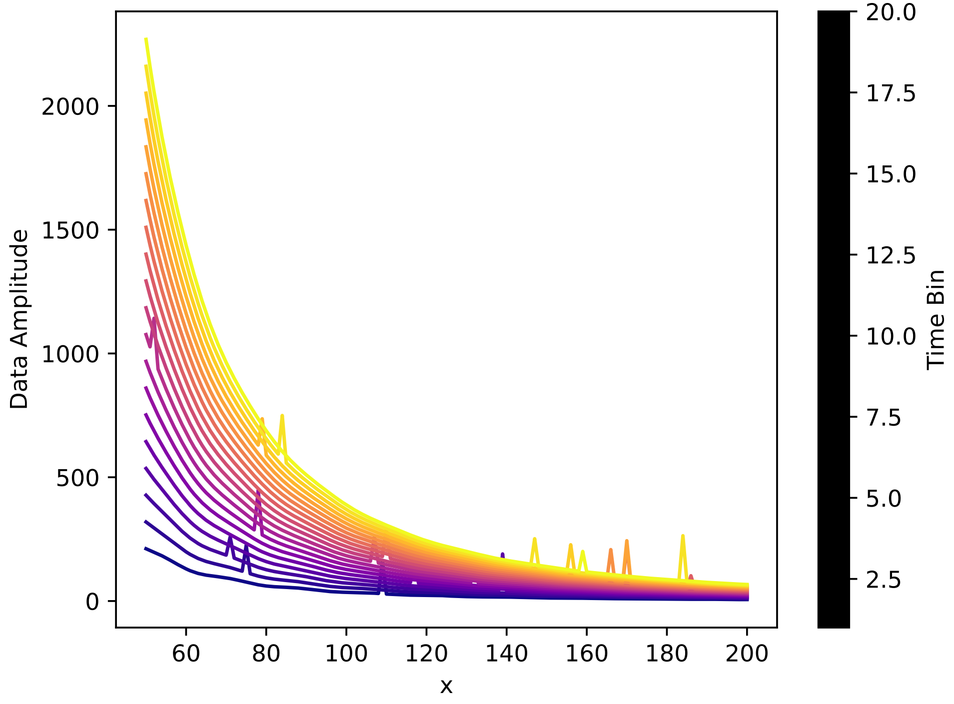
Sinusoid with time-varying parameters, multiplied by a power law

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

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

With randomly generated values for the sinusoid parameters

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

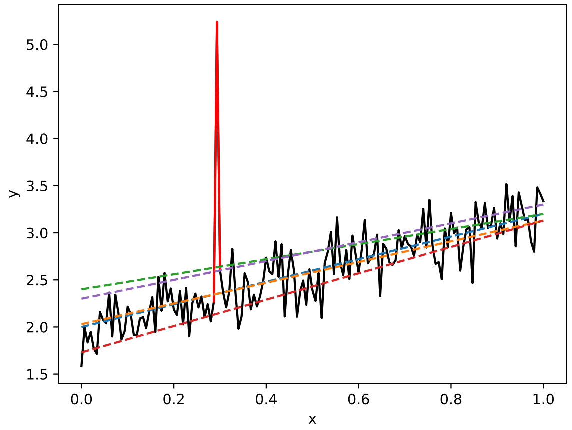


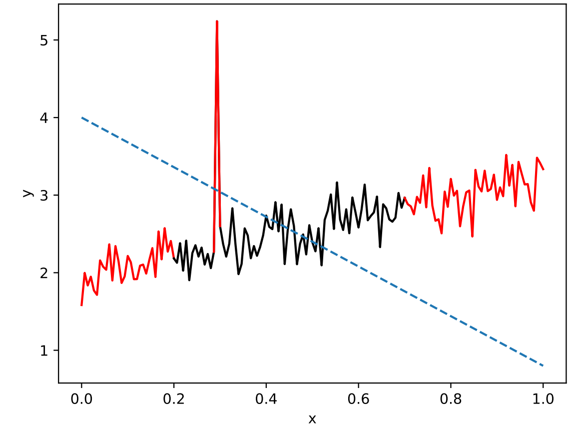
Challenges

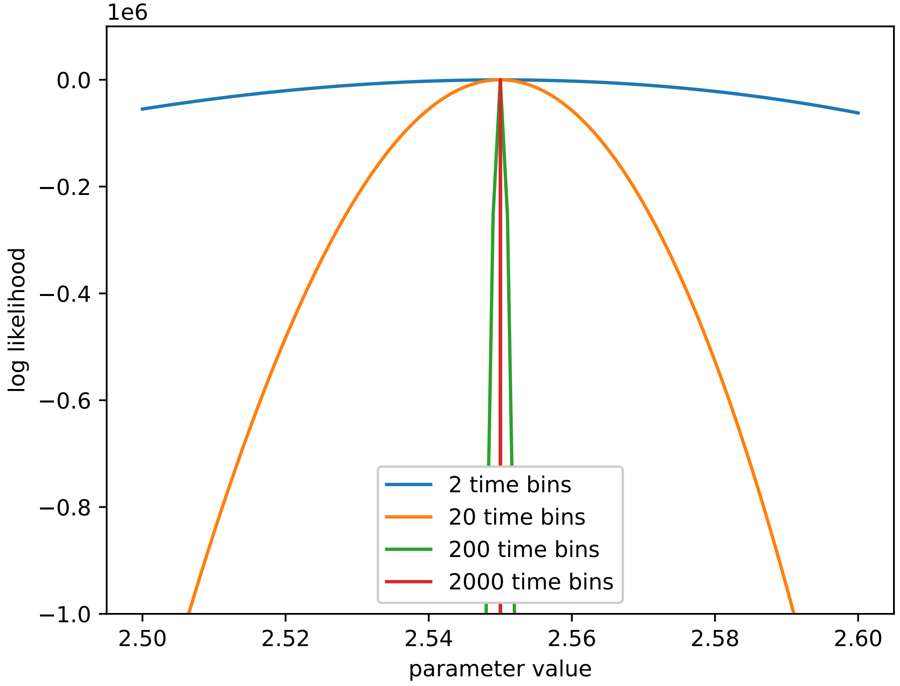
- Impact of model disagreements
  The anomaly flagger overflags points in regions of the parameter space where the model differs significantly from the data.

- Computation Time
  The likelihood evaluation time grows with the number of time bins, and the time summation cannot be factorised out and recalculated.

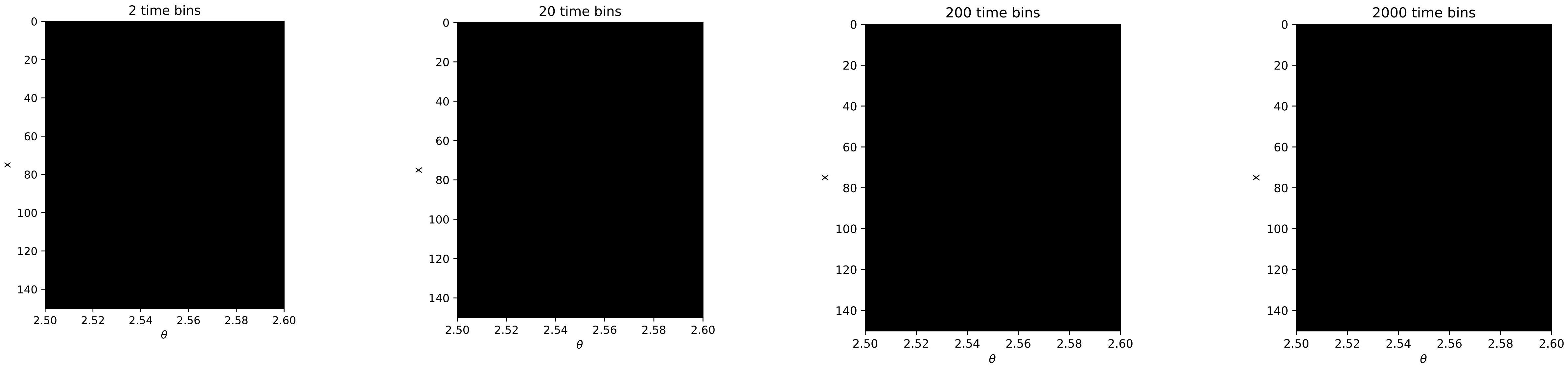
Model Disagreements







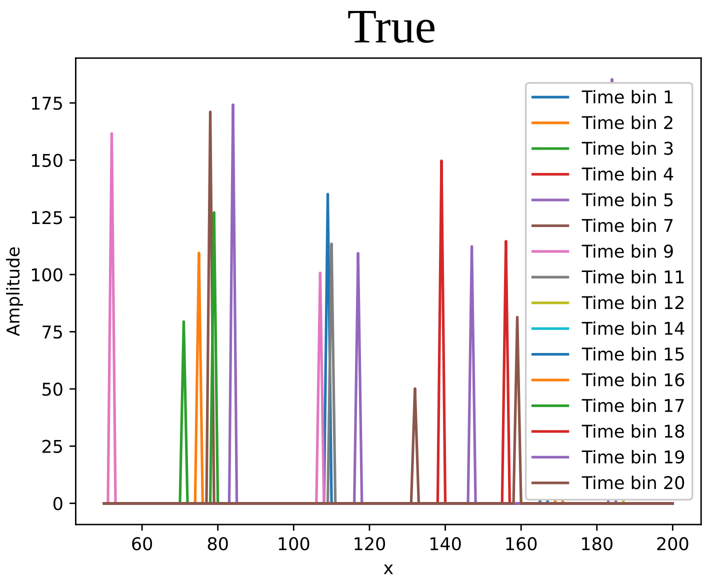
Model Disagreements

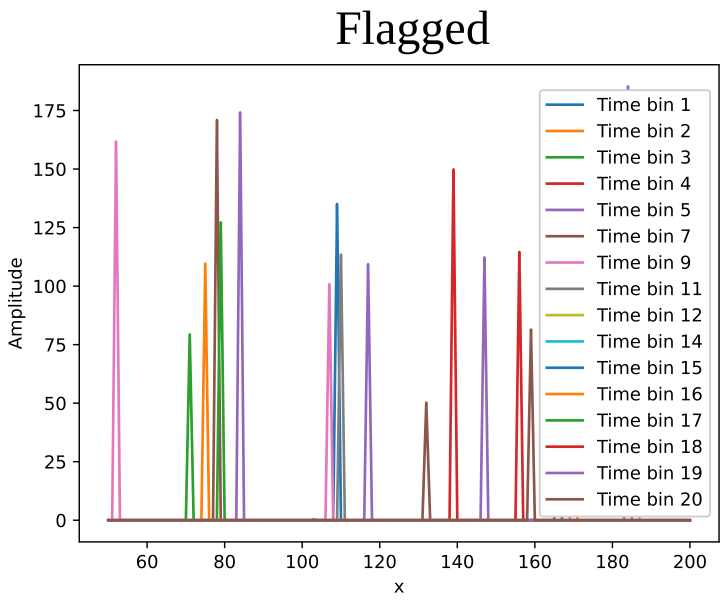


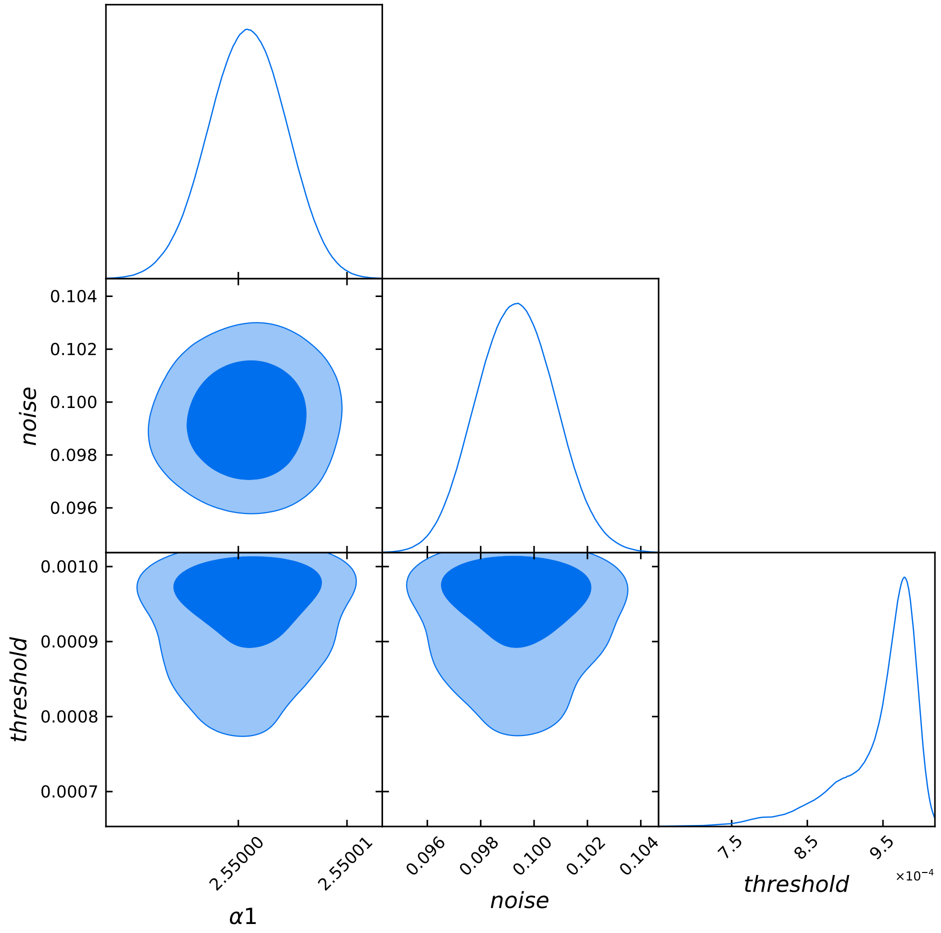
Model Disagreements

A fixed threshold p flattens likelihood surface.

Make p a variable parameter to be fit for.





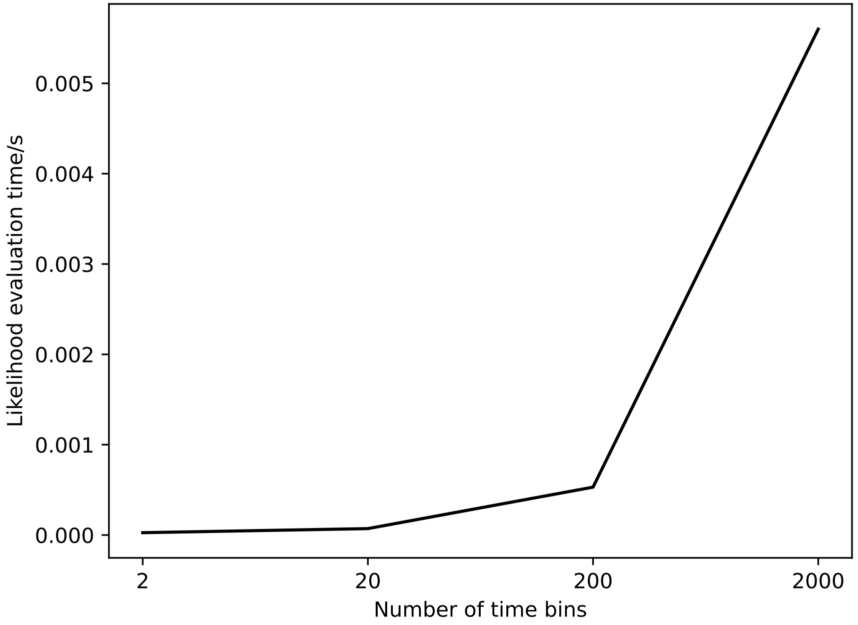


Computation Time

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.



Computation Time

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

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

Full likelihood - Slow to evaluate

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

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

Time averaged likelihood - Fast to evaluate

Both likelihoods use the same parameters and will have comparable posteriors

Likelihood Reweighting

Gravitational waves

- Payne et al. 2019
- Romero-Shaw et al. 2019

Likelihood Reweighting

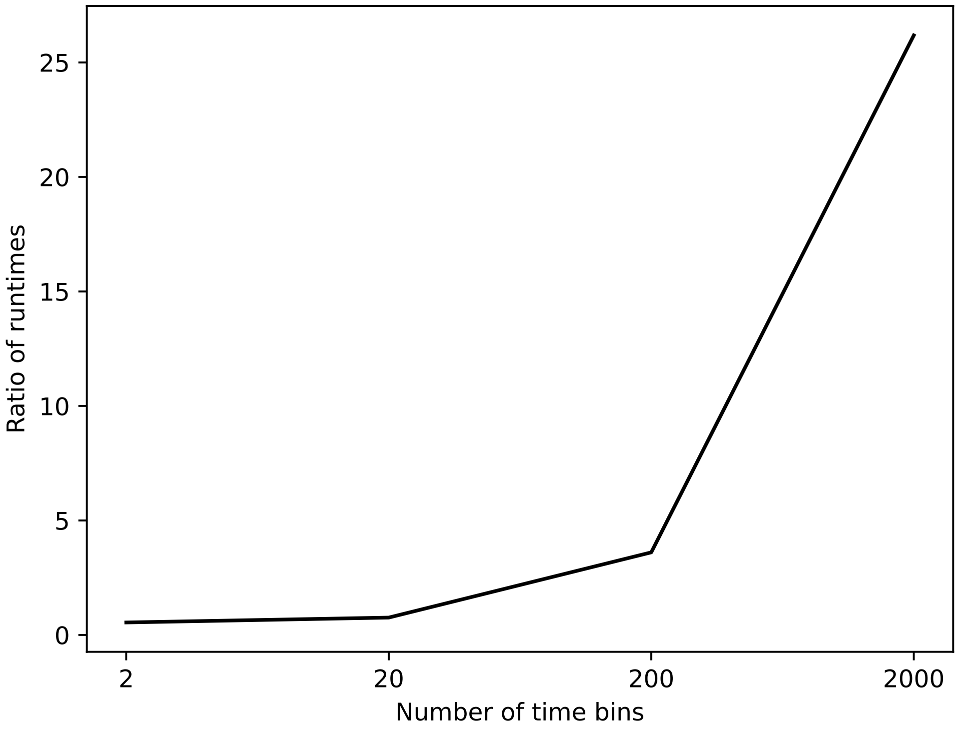
$$\mathcal{P}_{\rm S}\left(\theta|\mathcal{D},\mathcal{M}_{\rm S}\right) = \frac{\mathcal{L}_{\rm S}\left(\mathcal{D}|\theta,\mathcal{M}_{\rm S}\right)\pi\left(\theta\right)}{\mathcal{Z}_{\rm S}}$$

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

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

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

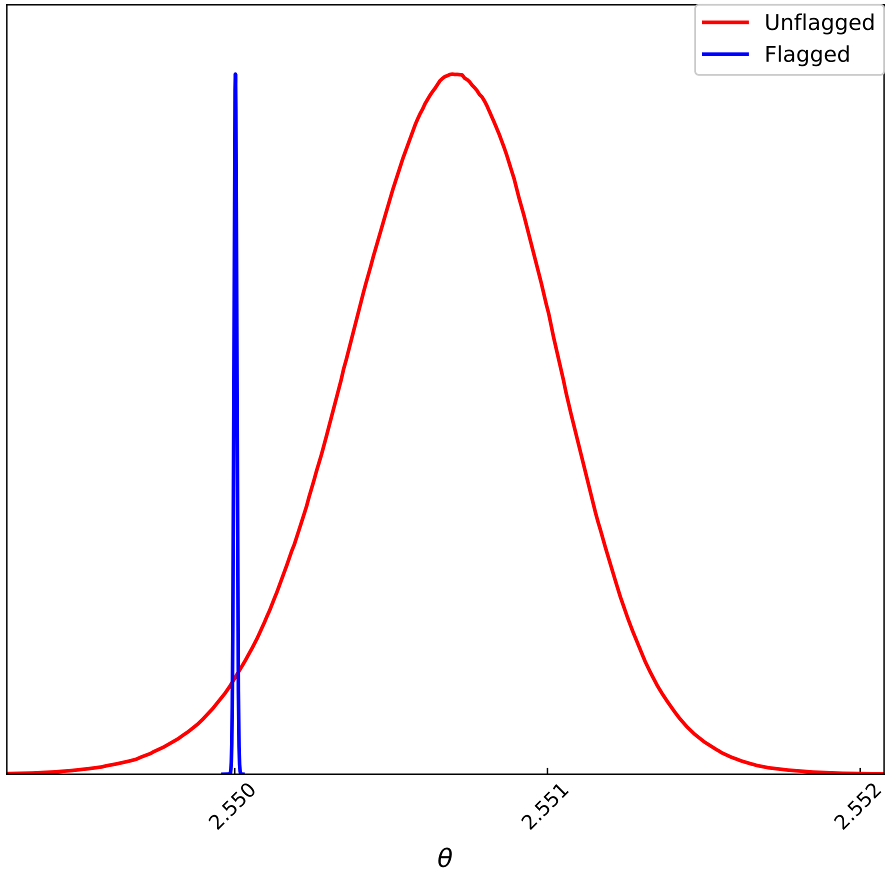
Results



Ratio of runtimes
0
5
10
15
20
25
2
20
200
2000
Number of time bins

Results

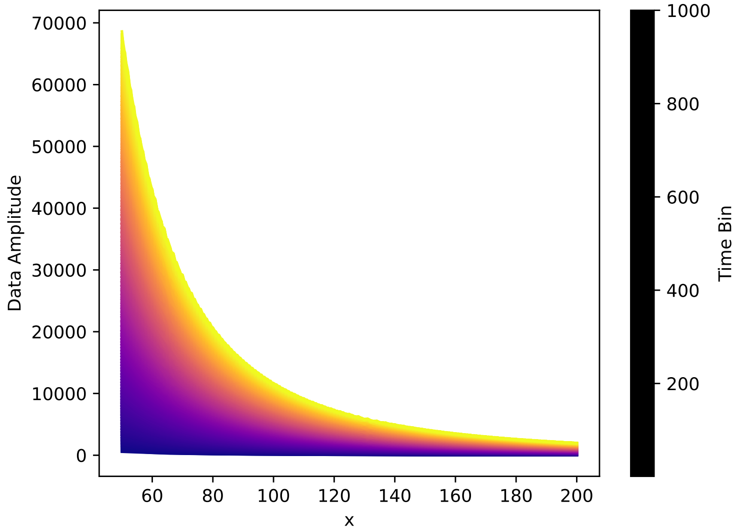
20 time bins, 20 contaminated points



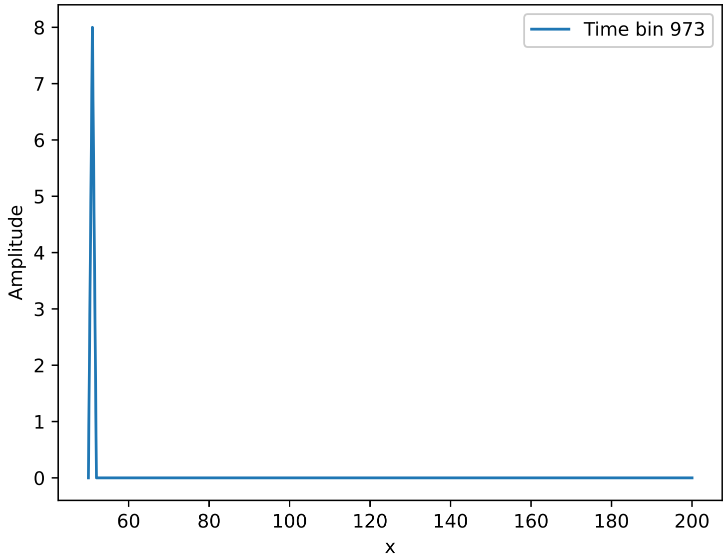
Results

1000 time bins, 1 anomalous point

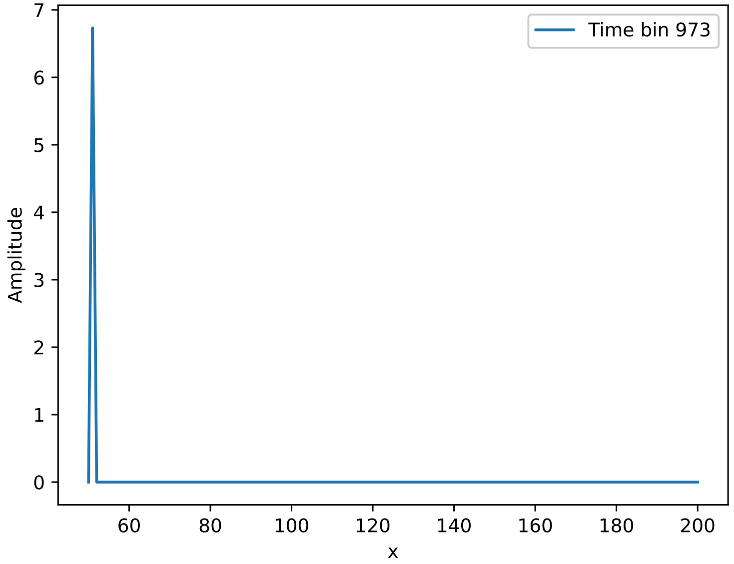
Data



Inserted

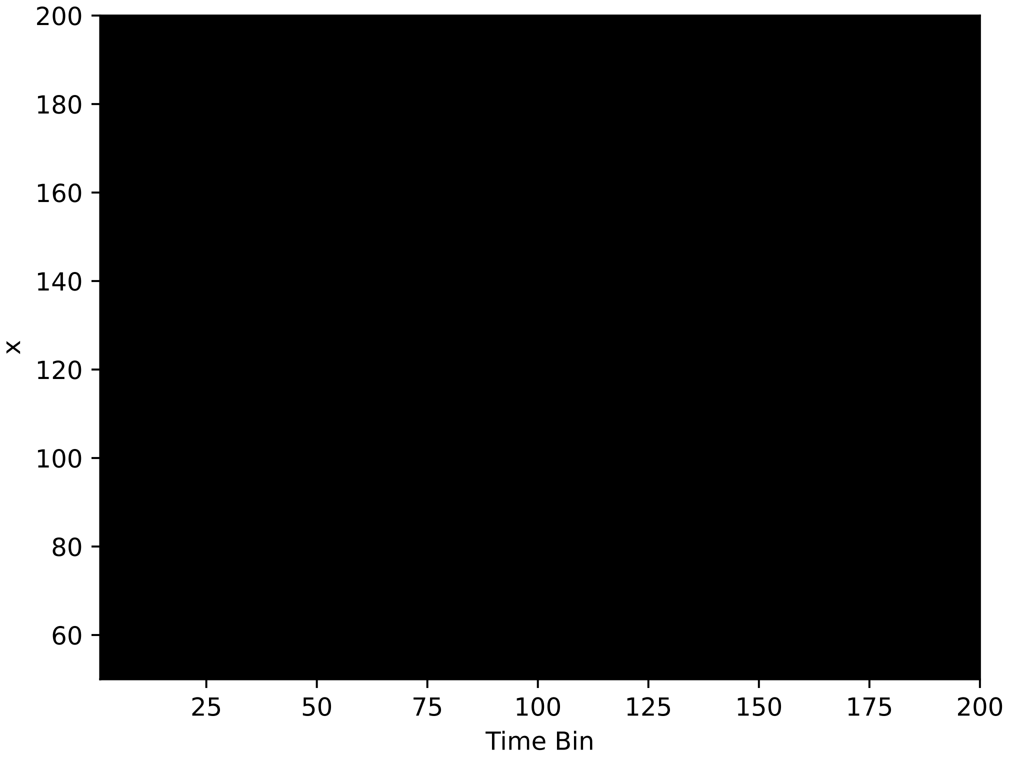


Recovered

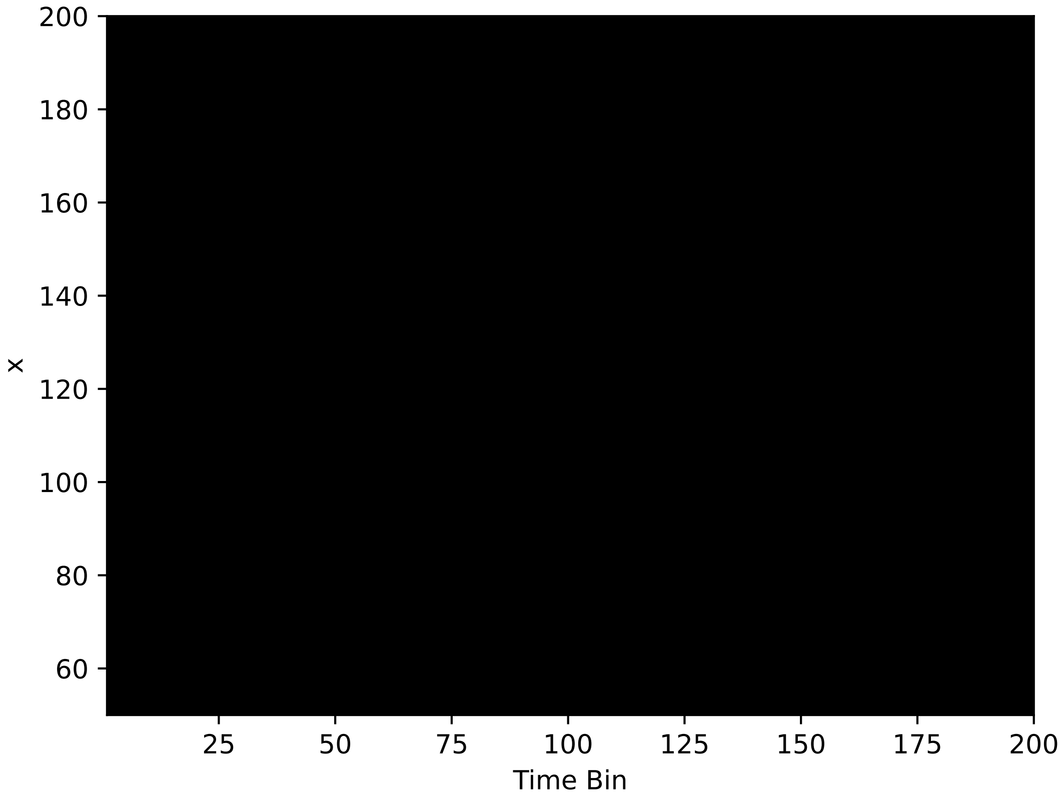


Results

Inserted



Recovered



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 flat likelihood surfaces can be resolved by fitting the threshold as a parameter
- 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