RAS Techniques and Instruments

RASTAI **3,** 372–384 (2024) https://doi.org/10.1093/rasti/rzae025
Advance Access publication 2024 July 5

Enhanced Bayesian RFI mitigation and transient flagging using likelihood reweighting

Dominic Anstey [1,2]★ and Samuel A. K. Leeney [1,2]

[1]*Astrophysics Group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, UK*
[2]*Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA, UK*

Accepted 2024 July 3. Received 2024 May 21; in original form 2023 October 12

ABSTRACT

Contamination by radio frequency interference (RFI) is a ubiquitous challenge for radio astronomy. In particular, transient RFI is difficult to detect and avoid, especially in large data sets with many time bins. In this work, we present a Bayesian methodology for time-dependent, transient anomaly mitigation performed jointly with model fitting. The computation time for correcting transient anomalies in this manner in time-separated data sets grows proportionally with the number of time bins. We demonstrate that utilizing likelihood reweighting can allow our Bayesian anomaly mitigation method to be performed with a computation time close to independent of the number of time bins. In particular, we identify a factor of 44 improvement in computation time for a test case with 2000 time bins. We also demonstrate how this method enables the flagging threshold to be fit as a free parameter, fully automating the mitigation process. We find that this threshold fitting also prevents overcorrecting of the data in the case of wide priors. Finally, we investigate the potential of the methodology as a transient detector. We demonstrate that the method is able to reliably flag an individual anomalous data point out of 302 000 provided the Signal to Noise Ratio is ≥ 10.

Key words: Data Methods – Bayesian – Transients.

1. INTRODUCTION

The field of radio astronomy has been rapidly growing in terms of both reach and data complexity. The number of known radio sources has been exponentially increasing, and will increase further in the future with the development of the Square Kilometre Array (SKA) (Braun et al. 2015) and the next-generation Very Large Array (ngVLA) (McKinnon et al. 2019). As the volume of information on the sky increases, astronomers also seek increasingly faint signals requiring more sensitive instruments and advanced data analysis techniques.

With the development of modern telecommunications devices, data from radio telescopes are becoming increasingly contaminated with interfering, anomalous signals such as radio frequency interference (RFI; Arrubarrena et al. 2024), which has become very difficult to avoid entirely (Fridman & Baan 2001; Pritchard et al. 2024), except in observations from extremely remote locations (Monsalve et al. 2024). This problem is worsened by the fact that current signals of interest, such as the Global 21-cm signals (Bowman et al. 2018; de Lera Acedo et al. 2022; Singh et al. 2022; Razavi-Ghods et al. 2023; Monsalve et al. 2024), lie in the same unprotected frequency bands as said devices.

RFI can emanate from a range of human-made sources, such as communication devices, satellites, and radar systems. It poses a significant challenge for radio astronomy, as it can obscure or mimic genuine celestial signals. RFI can be constant in time or transient (Czech, Mishra & Inggs 2018a). Transient RFI is particularly problematic because it is hard to detect and thus much more difficult to avoid. The SKA will gather up to 1 TB of data per second (Scaife 2020). With the volume of data to be analysed so large and the level of complex contaminants so high, there is a serious need for new data analysis techniques that are highly efficient and sensitive to such transient events.

There are various proposed ways to mitigate RFI. Offringa et al. (2010) use post-correlation classification methods in AOFLAGGER, which is used by the LOFAR (Röttgering 2003). The *FAST* (Nan 2006) telescope uses spatial filtering techniques (Wang et al. 2022). More recently, deep learning methods have been utilized (Kerrigan et al. 2019), latent nearest neighbours used to distinguish RFI by learning uncontaminated data (Mesarcik et al. 2022), and Bayesian methods used to calibrate satellite RFI based on trajectories (Finlay et al. 2023). For a more in-depth review of the current literature, we recommend Ford & Buch (2014) or Baan (2019). There are few methods designed specifically for transient RFI detection, as noted by Czech, Mishra & Inggs (2018b), who propose a dictionary-based approach to transient RFI detection.

Transient RFI is exceptionally problematic when the signal of interest is itself transient. For example, a transient RFI burst could not only obscure a signal [such as a fast radio burst (FRB)] but also mimic it leading to a false detection (Cendes et al. 2018). A satellite passing over the telescope e.g. could lead to such a transient anomaly (Finlay et al. 2023). This problem can be partially addressed using spectral kurtosis (Nita, Keimpema & Paragi 2019). However, spectral kurtosis is inadequate in various cases as described in Smith, Lynch & Pisano (2022). Furthermore, with many modern projects utilizing

★ E-mail: da401@cam.ac.uk

Bayesian methods in their data analysis pipelines, there is an urgent need for Bayesian RFI mitigation techniques.

In this paper, we present a novel Bayesian anomaly detection methodology that is efficient and sensitive to transient anomalies as well as time constant anomalies. In Section 2, we define the method. In Section 3, we test our methods on a simple toy model. In Section 4, we evaluate these methods when used to locate transient signals themselves, as well as mitigate against them. Finally, in Section 5, we present our conclusions.

2. METHODS

2.1 Bayesian anomaly mitigation

Leeney, Handley & Acedo (2023) proposed a fully Bayesian methodology for simultaneous anomaly flagging and excision, which is performed jointly with the primary model fit to the data, and can be readily folded into Bayesian analysis pipelines via a simple modification to their likelihood. The term *likelihood*, in the context of a single data point, defines the probability of observing that data point $\mathcal{D}_i$ given some model and its constituent parameters $\mathcal{M}_i(\theta)$. This assumes a predefined probability distribution for the noise. For instance, assuming uncorrelated Gaussian noise results in a likelihood of the form

$$\log \mathcal{L}_i(\theta) = -\frac{1}{2}\log\left(2\pi\sigma_{\rm n}^2\right) - \frac{1}{2}\left(\frac{\mathcal{D}_i - \mathcal{M}_i(\theta)}{\sigma_{\rm n}}\right)^2, \tag{1}$$

where θ is the parameter(s) and $\sigma_{\rm n}$ is the noise amplitude. Alternatively, for a data set containing N_x data points, and a corresponding model with a single parameter set θ for all points, assuming no correlations, the overall likelihood is the product of the probabilities of each datum (or alternatively, the sum of log probabilities), giving

$$\log \mathcal{L}(\theta) = \sum_i^{N_x} \log \mathcal{L}_i(\theta), \tag{2}$$

where i indexes over the data points. In many cases, this will be an index over a variable x. For the examples in this work, this is taken to be observation frequency.

However, this likelihood cannot account for isolated data points that do not match the probability distribution of the rest of the data set. Throughout this work, we refer to such data points as 'anomalous'. Contamination by RFI is a common source of such anomalous data points. Typically, anomalies are flagged and excised prior to the Bayesian fitting procedure. This can be problematic, as it leads to potentially useful information being thrown away. Leeney et al. (2023) showed that anomalies can be modelled simultaneously in a Bayesian fashion by using a piece-wise likelihood capable of modelling both the probability of abnormality and the probability that each datum fits to the model of interest.

This was achieved in Leeney et al. (2023) by first defining a model in which, if a data point is contaminated, the probability of observing that data becomes independent of the parameters. It is instead assumed to be uniform over a defined scale $[0 - \Delta]$, where Δ describes the scale of the contamination, in the same units as those of the data, and can typically be set to $\Delta \approx \text{MAX}$ (data). The probability of observing a given data point given the model then depends on whether that point is contaminated or not, as

$$P(\mathcal{D}_i|\theta) = \begin{cases} \mathcal{L}_i(\theta), & \text{uncontaminated} \\ \Delta^{-1}, & \text{contaminated.} \end{cases} \tag{3}$$

This can be expressed with a Boolean mask to identify each data point as contaminated or not:

$$P(\mathcal{D}_i|\theta, \epsilon) = \prod_i \mathcal{L}_i(\theta)^{\epsilon_i} \left(\Delta^{-1}\right)^{1-\epsilon_i}. \tag{4}$$

Here, $\epsilon_i = 1$ indicates uncontaminated data and $\epsilon_i = 0$ indicates contaminated data.

As it is not necessarily known which data points are contaminated a priori, a probability can be assigned to each given epsilon mask. This gives the probability of observing the data point, given the model, as

$$P(\mathcal{D}_i, \epsilon|\theta) = \prod_i [\mathcal{L}_i(\theta) P(\epsilon_i = 1)]^{\epsilon_i} [\Delta^{-1} P(\epsilon_i = 0)]^{1-\epsilon_i}. \tag{5}$$

Leeney et al. (2023) then makes the assumption that the probability of data points being contaminated are uncorrelated, with a probability p_i that any individual point is contaminated, such that

$$P(\epsilon_i = 0) = p_i \tag{6}$$

and

$$P(\epsilon_i = 1) = 1 - p_i. \tag{7}$$

Substituting in these values into equation (5) gives

$$P(\mathcal{D}_i, \epsilon|\theta) = \prod_i [\mathcal{L}_i(\theta)(1 - p_i)]^{\epsilon_i} [p_i/\Delta]^{1-\epsilon_i}. \tag{8}$$

Ideally, ϵ_i should be marginalized over. However, this is computationally impractical in most cases. Therefore, Leeney et al. (2023) makes the approximation that the ϵ_i that returns the highest likelihood dominates in the marginalization. This highest likelihood mask is given by

$$\epsilon_i = \begin{cases} 1, & \mathcal{L}_i(1 - p_i) > p_i/\Delta \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Therefore, the final likelihood can be expressed as

$$\log \mathcal{L}(\theta) = \sum_i \begin{cases} \log \mathcal{L}_i + \log(1 - p_i), & \log \mathcal{L}_i + \log(1 - p_i) \\ & > \log p_i - \log \Delta \\ \log p_i - \log \Delta, & \text{otherwise.} \end{cases} \tag{10}$$

In practice, this likelihood operates by removing very low likelihood data points from the overall likelihood and replacing them with an Occam's penalty. This is because the Bayesian evidence favours the simplest solution that best describes the data, so without the penalty the 'most likely' solution would be to flag all of the data as anomalous.

Leeney et al. (2023) showed that this likelihood enabled anomalous data points to be efficiently flagged and corrected for automatically in a Bayesian model fit of a one-dimensional (1D) data set. The focus of this work is extending this methodology for application to 2D, time-varying data sets.

2.2 Time-binned modelling

Anstey, de Lera Acedo & Handley (2023) proposed a methodology for efficient inclusion of time-dependent data sets in Bayesian modelling. Typically, in the case of time-varying data, modelling each time bin separately is unfeasible as it requires a complete model with its own set of parameters for each bin, which can quickly result in the dimensionality of the fit becoming very large for more than a few time bins. As a result, time-varying data are typically modelled by fitting a model to the time averaged data set. For example, in the case

of the uncorrelated Gaussian likelihood described in equations (1) and (2), a time-averaged likelihood would take the form

$$\log \mathcal{L}(\theta) = \sum_i -\frac{1}{2}\log\left(2\pi\sigma_{\rm n}^2\right) - \frac{1}{2}\left(\frac{\frac{1}{N_t}\sum_j \mathcal{D}_{ij} - \mathcal{M}_i(\theta)}{\sigma_{\rm n}}\right)^2, \tag{11}$$

where j indexes time bins and N_t is the total number of time bins.

However, in Anstey et al. (2023), the case was considered where a time varying model can be defined as a product of a parameter-dependent component and a parameter-independent component, where only the parameter-independent component has time dependence, of the form

$$\mathcal{M}_{ij}(\theta) = \mathcal{F}_i(\theta)\,\mathcal{G}_{ij}. \tag{12}$$

If it is the case that the time-dependent component G_{ij}, despite being independent of the specific value of the parameters, is required to be different for the different parameters in θ, which will be the case if different parameters are describing different effects, this can be expressed more fully as

$$\mathcal{M}_{ij}(\theta_k) = \mathcal{F}_i(\theta_k)\,\mathcal{G}_{ijk}, \tag{13}$$

where k indexes over the parameters and θ_k expresses the individual elements of the vector θ.

An example of a case where a model of this form can be defined is for a set of parameters defining an astrophysical observable that is constant on the time-scales of the experiment, with time variance only being introduced by the rotation of the Earth. Another possibility would be a data set of just noise with transient signals, which would simply have a model of $\mathcal{M}_{ij}(\theta) = 0$, which still satisfies this condition.

It should be noted that time dependence is not the only way a model can be factorized in the manner shown in equation (13). For example, if the frequency component of the model were parameter independent, it could be factorized instead. Provided their effect on the model is not dependent on the parameters being fit for, it may also be possible to factorize polarization, or even different instruments in this fashion. The fast anomaly mitigation techniques described in Sections 2.3 and 2.4 could then be equivalently applied in such cases. However, for the purposes of this work, we will focus on time-dependence as a test case.

If such a model can be defined, it becomes possible to implement simultaneous fitting of separate time bins of data to corresponding models without the dimensionality of the parameters increasing, as each time bin fits for the same parameter set. Such a fitting process can be implemented by modifying the likelihood shown in equation (11) to

$$\log \mathcal{L}(\theta) = \sum_{ij} -\frac{1}{2}\log\left(2\pi\sigma_{\rm n}^2\right) - \frac{1}{2}\left(\frac{\mathcal{D}_{ij} - \mathcal{M}_{ij}(\theta)}{\sigma_{\rm n}}\right)^2. \tag{14}$$

It was demonstrated in Anstey et al. (2023), in the context of global 21-cm experiments, that using this full time-dependent likelihood in a Bayesian model fit enabled the time variance of the model to be exploited to constrain the model parameter more tightly than could be achieved for a time-averaged model.

2.3 Time-binned anomaly mitigation

Applying the Bayesian anomaly mitigation technique described in Section 2.1 to the time-dependent likelihood described in Section 2.2 is straightforward, requiring only that the flagged likelihood shown in equation (10) be extended into two dimensions as

$$\log \mathcal{L}(\theta) = \sum_{ij} \begin{cases} \log \mathcal{L}_{ij} + \log(1-p_{ij}), & \log \mathcal{L}_{ij} + \log(1-p_{ij}) \\ & > \log p_{ij} - \log\Delta \\ \log p_{ij} - \log\Delta, & \text{otherwise,} \end{cases} \tag{15}$$

where $\log \mathcal{L}_{ij}$ is the likelihood of a single data point in a single time bin

$$\log \mathcal{L}_{ij}(\theta) = -\frac{1}{2}\log\left(2\pi\sigma_{\rm n}^2\right) - \frac{1}{2}\left(\frac{\mathcal{D}_{ij} - \mathcal{M}_{ij}(\theta)}{\sigma_{\rm n}}\right)^2. \tag{16}$$

However, implementing this method in practice faces a challenge. The likelihood shown in equation (14) requires a summation over time bins. This means the computation time of the likelihood grows linearly with the number of time bins used in the data set. As a result, larger numbers of time bins can greatly slow the fitting procedure.

In Anstey et al. (2023), this issue was resolved by reformatting the likelihood such that all summations over time could be calculated once, outside of the likelihood, removing the dependence of the calculation time on the number of time bins. However, this solution is not possible to implement when the anomaly correcting procedure is also implemented. This is because the value of $\log \mathcal{L}_{ij}$ must be calculated for every time bin within the likelihood, in order to evaluate $\log \mathcal{L}(\theta)$, as shown in equation (15). Therefore, an alternative method of speeding the likelihood evaluation is needed to make this process viable in practice. This can be achieved using likelihood reweighting.

2.4 Likelihood reweighting

The process of likelihood reweighting is an extension to importance sampling, pioneered in the context of gravitational waves (Payne, Talbot & Thrane 2019; Romero-Shaw, Lasky & Thrane 2019). It is a method for speeding the evaluation of a posterior and evidence in a Bayesian fit for the case of a complex model that is otherwise slow to evaluate. This process relies on several key criteria.

Firstly, two models, $\mathcal{M}_{\rm F}(\theta)$ and $\mathcal{M}_{\rm S}(\theta)$, are required. One of these must be quick to evaluate, which will henceforth be assumed to be model F, and one is slower to evaluate, which we define as model S. These two models must be parametrized by the same parameter vector θ, with the same prior distribution $\pi(\theta)$. They must also have their posterior peak in approximately the same region of parameter space.

By definition, the posteriors of the two models can be expressed as

$$\mathcal{P}_{\rm F}(\theta|\mathcal{D}, \mathcal{M}_{\rm F}) = \frac{\mathcal{L}_{\rm F}(\mathcal{D}|\theta, \mathcal{M}_{\rm F})\,\pi(\theta)}{\mathcal{Z}_{\rm F}} \tag{17}$$

and

$$\mathcal{P}_{\rm S}(\theta|\mathcal{D}, \mathcal{M}_{\rm S}) = \frac{\mathcal{L}_{\rm S}(\mathcal{D}|\theta, \mathcal{M}_{\rm S})\,\pi(\theta)}{\mathcal{Z}_{\rm S}}, \tag{18}$$

where $\mathcal{L}_{\rm F}(\mathcal{D}|\theta, \mathcal{M}_{\rm F})$ and $\mathcal{L}_{\rm S}(\mathcal{D}|\theta, \mathcal{M}_{\rm S})$ are the likelihoods calculated from the two models and $\mathcal{Z}_{\rm F}$ and $\mathcal{Z}_{\rm S}$ are the respective evidences.

However, given the aforementioned criterion that the two models have the same priors, the prior can be expressed in terms of model F, as

$$\pi(\theta) = \frac{\mathcal{Z}_{\rm F}\mathcal{P}_{\rm F}(\theta|\mathcal{D}, \mathcal{M}_{\rm F})}{\mathcal{L}_{\rm F}(\mathcal{D}|\theta, \mathcal{M}_{\rm F})} \tag{19}$$

and substituted into equation (18) to give

$$\mathcal{P}_{\rm S}(\theta|\mathcal{D}, \mathcal{M}_{\rm S}) = \mathcal{P}_{\rm F}(\theta|\mathcal{D}, \mathcal{M}_{\rm F})\,\frac{\mathcal{L}_{\rm S}(\mathcal{D}|\theta, \mathcal{M}_{\rm S})}{\mathcal{L}_{\rm F}(\mathcal{D}|\theta, \mathcal{M}_{\rm F})}\,\frac{\mathcal{Z}_{\rm F}}{\mathcal{Z}_{\rm S}}. \tag{20}$$

Thus, the posterior of model S can be evaluated without having to perform a full Bayesian model fit with the slow-to-calculate likelihood. This is achieved by instead performing a model fit of the much faster evaluated model F. Given the criterion that the bulk of the two models' posteriors occupy similar regions of the parameter space, the samples of this fast evaluated posterior will cover the same parameter volume as a hypothetical posterior of the slower model. Therefore, reweighting the posterior samples of model F by a factor of the ratio of the likelihoods

$$w(\theta) = \frac{\mathcal{L}_{\rm S}(\mathcal{D}|\theta, \mathcal{M}_{\rm S})}{\mathcal{L}_{\rm F}(\mathcal{D}|\theta, \mathcal{M}_{\rm F})}, \tag{21}$$

will convert the samples to samples of the posterior of model S, to within a constant factor given by the ratio of the evidence. By this method, a set posterior for model S can be evaluated with the slow likelihood only needing to be computed for the relatively small number of posterior samples. This can be significantly faster than sampling the complex likelihood across the entire prior volume. For a particularly complex and thus slow to calculate likelihood, this can make the fitting procedure significantly faster. This effect will be demonstrated in Section 3.3.3.

Therefore, the process of likelihood reweighting is well suited to achieving fast and efficient time-dependent anomaly mitigation. By taking an anomaly correcting likelihood that acts on precomputed time-summed likelihoods as the quickly evaluated model F and the full time separated likelihood as the slowly evaluated model S, the full time-dependent likelihood posterior can be evaluated quickly, allowing the Bayesian anomaly mitigation procedure to be implemented practically on time-dependent data sets as follows. For the slow likelihood, the likelihood of each data point is calculated according to equations (15) and (16).

For the fast likelihood, however, in order to preserve as much of the time-dependent data as possible, the following process is implemented.

First, the methodology described in Anstey et al. (2023) to evaluate a time-dependent likelihood in a time-independent fashion by expanding the likelihood given in equation (14) and precomputing the summations over time is implemented. As previously noted, precomputing time summations in this fashion, while greatly accelerating the likelihood evaluation time, prevents the evaluation of $\log \mathcal{L}_{ij}$ directly and thus prevents the implementation of the full time-dependent flagging method defined in equation (15). However, from equation (7) of Anstey et al. (2023), it is possible to evaluate the product of likelihoods (or equivalently the sum of log likelihoods) over all time bins in each data channel

$$\log \widetilde{\mathcal{L}}_i = \sum_j \log \mathcal{L}_{ij}, \tag{22}$$

from precomputed time summations.

Therefore, it is possible to perform the anomaly mitigation methodology in a time-independent fashion by flagging out entire channels where contamination occurs, based on the summed log likelihoods, rather than only the contaminated time bins of those channels.

In order to determine the appropriate flagging thresholds for this process, we repeat the derivation of Leeney et al. (2023), described in Section 2.1, applied to the summed log likelihood.

We first make the assumption that if any time bin in a particular channel is contaminated, that contamination will dominate, and the resulting probability of the product of all time bins becomes independent on the model, and is instead uniform over a defined scale. In this case, the resulting probability of a contaminated channel depends on how many contaminated points it contains:

$$P(\mathcal{D}_i|\theta) = \begin{cases} \widetilde{\mathcal{L}_i(\theta)}, & \text{all uncontaminated} \\ \Delta^{-k_i}, & k_i \text{ contaminated,} \end{cases} \tag{23}$$

where Δ again describes the scale of the contamination, k_i is the number of contaminated points in channel i, and $\widetilde{\mathcal{L}_i(\theta)} = \prod_j \mathcal{L}_{ij}(\theta)$.

This can again be expressed using a Boolean mask, ϵ_i, as

$$P(\mathcal{D}_i|\theta, \epsilon) = \prod_i \widetilde{\mathcal{L}_i(\theta)}^{\epsilon_i} (\Delta^{-k_i})^{1-\epsilon_i}. \tag{24}$$

Here, $\epsilon_i = 1$ refers to the case where no time bins in channel i are contaminated, and $\epsilon_i = 0$ to the case where at least 1 time bin is contaminated.

As before, the probability of observing the data channel, given the model, can then be expressed as

$$P(\mathcal{D}_i, \epsilon|\theta) = \prod_i \left[\widetilde{\mathcal{L}_i(\theta)} P(\epsilon_i = 1)\right]^{\epsilon_i} \left[\sum_{k_i=1}^{N_t} \Delta^{-k_i} P(k_i) P(\epsilon_i = 0|k_i)\right]^{1-\epsilon_i}. \tag{25}$$

It must be noted here that, as the value of k_i in each case is not known, we sum over all possible values, weighted by their probabilities.

If, again, the probability that any given point is contaminated is p_i, assuming again that there are no correlations, the probability that a channel contains no contaminated points is

$$P(\epsilon_i = 1) = (1 - p_i)^{N_t}, \tag{26}$$

where N_t is the number of time bins.

In the case of $\epsilon = 0$, the probability of k_i points being contaminated is a simple binomial distribution:

$$P(k_i) = \binom{N_t}{k_i} (p_i)^{k_i} (1 - p_i)^{N_t - k_i}. \tag{27}$$

As the case of $\epsilon_i = 0$ is defined to be where at least one point in the channel is contaminated, so $k_i \geq 1$, it can be seen that

$$P(\epsilon_i = 0|k_i) = 1 \tag{28}$$

for all $k_i \neq 0$.

Substituting equations (26)–(28) into equation (25) and noting that

$$\sum_{k_i=0}^{N_t} \binom{N_t}{k_i} x^{k_i} y^{N_t - k_i} = (x + y)^{N_t} \tag{29}$$

gives

$$P(\mathcal{D}_i, \epsilon|\theta) = \prod_i \left[\widetilde{\mathcal{L}_i(\theta)}(1 - p_i)^{N_t}\right]^{\epsilon_i} \left[(1 + p_i(\Delta^{-1} - 1))^{N_t} - (1 - p_i)^{N_t}\right]^{1-\epsilon_i}. \tag{30}$$

Making, as before, the assumption that the most likely model dominates the marginalization of the ϵ_i masks, and given that the most likely mask is given by

$$\epsilon_i = \begin{cases} 1, & \widetilde{\mathcal{L}_i(\theta)}(1 - p_i)^{N_t} \\ & > (1 + p_i(\Delta^{-1} - 1))^{N_t} - (1 - p_i)^{N_t} \\ 0, & \text{otherwise,} \end{cases} \tag{31}$$

the final likelihood can be expressed as

$$\sum_i = \begin{cases} \log \widetilde{\mathcal{L}_i} + N_t \log(1-p_i), & \log \widetilde{\mathcal{L}_i} + N_t \log(1-p_i) \\ & > \log\left[(1+p_i(\Delta^{-1}-1))^{N_t} \right. \\ & \left. -(1-p_i)^{N_t}\right] \\ \log\left[(1+p_i(\Delta^{-1}-1))^{N_t}\right. & \\ \left. -(1-p_i)^{N_t}\right], & \text{otherwise.} \end{cases} \tag{32}$$

It can be seen that in the case of $N_t = 1$, this reduces to equation (10), as would be expected.

Overall, this likelihood uses the methodology described in Anstey et al. (2023) to evaluate a time-dependent likelihood in a time-independent fashion, while simultaneously implementing the anomaly mitigation methodology in a time-independent fashion by flagging out entire channels where contamination occurs, rather than only the contaminated time bins of those channels. This results in a likelihood that takes the same parameters as the full flagging likelihood shown in equation (15), with a posterior peak expected in a similar position, but is much faster to evaluate, albeit with a lower constraining power due to utilizing a smaller fraction of the available data. Therefore, equation (32) serves as an ideal fast likelihood for a likelihood reweighting process, as required.

In the following sections, this complete process will be tested on simulated data to evaluate its efficacy.

3. TIME-DEPENDENT RFI MITIGATION

3.1 Toy model

In order to evaluate the efficacy of this method for correcting transient RFI in time separated data, a toy-simulated time-dependent data set was generated of the form

$$\mathcal{D}_{ij} = \left[\alpha_j \sin\left(\omega_j x_i + \phi_j\right) + \gamma_j\right] x_i^{-2.55} + \hat{\sigma}, \tag{33}$$

where x describes the data channel value, indexed by i. For this toy model, this is assumed to be frequency in MHz and runs from 50 to 200. α_j, ω_j, ϕ_j, and γ_j are time-dependent variables that are chosen to vary gradually over the time bins. For each of the four values, a start point is randomly chosen uniformly from the range [0–5] for α_j, [0–1] for ω_j and ϕ_j, and [110–130] for γ_j. The four variables are then iteratively defined according to

$$\text{variable}_j = \text{variable}_{j-1} + \mathcal{N}(\mu_{\text{variable}}, \sigma_{\text{variable}}), \tag{34}$$

where $\mathcal{N}(\mu_{\text{variable}}, \sigma_{\text{variable}})$ indicates a value randomly drawn from a normal distribution of mean μ_{variable} and standard deviation σ_{variable}. μ_α and μ_γ were both set to 0, to ensure that the absolute scale of the toy model did not significantly vary from time bin to time bin, with standard deviations of $\sigma_\alpha = 0.1$ and $\sigma_\gamma = 1$. μ_ω and μ_ϕ were both randomly drawn from the range $[-0.05 - 0.05]$, with $\sigma_\omega = \sigma_\phi = 0.05$, such that the sinusoidal distortion would vary over time.

$\hat{\sigma}$ is a realization of random Gaussian white noise added to the data. Unless otherwise specified, this was set to have a standard deviation of 0.25. A different noise realization is added to each time bin.

A toy model of this form was chosen as it approximates the form of data from a global 21-cm experiment, which is dominated by foregrounds of primarily diffuse synchrotron emission (Shaver et al. 1999). The power law with a 2.55 spectral index approximates the spectral variation of the diffuse emission from the sky and the time varying sinusoids approximate the convolution of the diffuse emission with a chromatic antenna beam, as the Earth rotates. This is described, e.g. in equation (18) of Anstey et al. (2021) and seen in Bowman et al. (2018). Global 21-cm experiments are an anticipated use case of this process, so a toy model of this form enables these experiments to be used as a test case of the methodology.

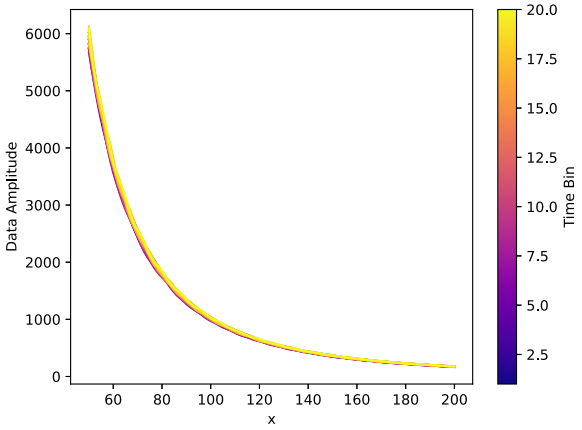

Figure 1. Example test data set with 20 time bins generated according to the toy model described in Section 3.1.

In addition, a simulated data set of this form has the required structure to apply the time-separated model fitting as specified in equation (12), with

$$\mathcal{G}_{ij} = \alpha_j \sin\left(\omega_j x_i + \phi_j\right) + \gamma_j \tag{35}$$

and

$$\mathcal{F}_i(\theta) = x_i^{-\theta} \tag{36}$$

with a 'true value' of $\theta = 2.55$.

Fig. 1 shows an example simulated data set generated for 20 time bins using this toy model. Once such a data set has been generated, any arrangement of anomalous points can then be added in order to test the proposed Bayesian anomaly mitigator.

3.2 Parametrizing the threshold

In Leeney et al. (2023), the likelihood threshold value for determining if a point should be flagged as an anomaly or not, p, was set to a fixed value. However, doing so produces a challenge when applied to time-dependent modelling.

As shown in equation (15), the general method of operation for the Bayesian anomaly correction procedure is to calculate the likelihood of each separate data point, compare it with a predefined threshold value and if it exceeds the threshold, include it with an appropriate weighting and if it does not, flag it as an anomaly and instead add a fixed penalty to the total likelihood. This is acceptable under the assumption that only anomalous points will have likelihoods lower than the threshold.

However, in cases where the model used has relatively large priors and significant variability, it becomes possible for a given parameter sample to produce a model sufficiently different from the data set that significant numbers of data points have likelihoods below the threshold, as demonstrated by Fig. 2.

This becomes of concern if the priors are wide enough compared with the posterior peak to result in sections of the prior space in which every data point has a likelihood below the defined threshold set by p. In this case, the overall likelihood is a sum of only fixed penalty terms, and thus is constant. As a result, all sections of the prior space that satisfy this condition have the same likelihood. This

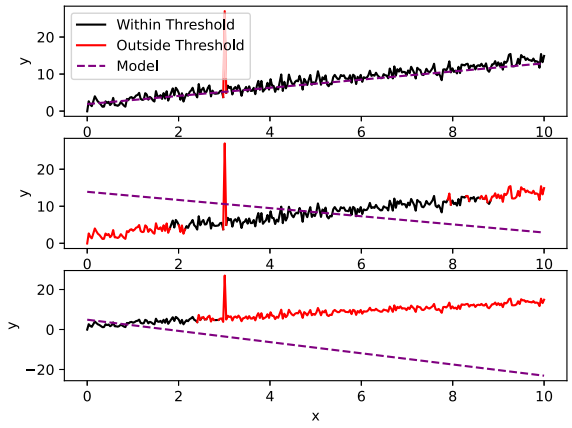

Figure 2. Example, for linear noisy data, of the range of data points that are further from the model than a defined threshold for a model that closely matches the data set (top panel), moderately differs from the data set (middle panel), and significantly differs from the data set (bottom panel).

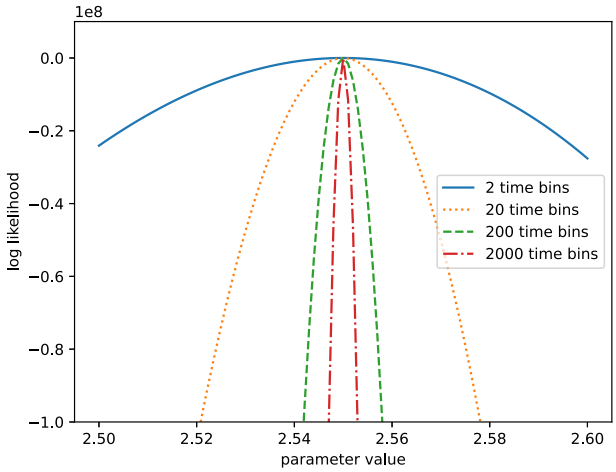

Figure 3. Plot of the total summed log likelihood shown in equation (14) for the toy model described in Section 3.1 as a function of parameter value for 2, 20, 200, and 2000 time bins.

manifests as the outer boundaries of the likelihood surface becoming flat.

If the majority of the prior space is not flattened in this manner, this has minimal impact on the process of performing a model fit and recovering the parameter posteriors. However, if the likelihood surface is particularly steep, it is possible for the majority of the space to become flat, except for a very narrow region around the posterior peak. In this case, performing a model fit becomes impossible, as there exists no variation to guide an algorithm towards the peak.

This is a challenge when performing time-dependent model fits, because the additional time information amplifies the effect of model disagreement. A parameter sample that gives a model that matches well to a data set with many time bins will give a high likelihood for each data point of each time bin, and thus give a larger overall likelihood than for an equally well matching data set with fewer time bins, and vice versa. As demonstrated in Fig. 3 for the proposed one-parameter toy model, defined by equations (36) and (35), with 2, 20, 200, and 2000 time bins, this has the effect of steepening the likelihood surface, the more time bins the data set has.

Fig. 4 shows the regions where the value of $\mathcal{L}_i$, averaged over time, falls above or below a threshold of $p =$1e−3, which is a typical value as determined from Leeney et al. (2023). It can be seen that as the number of time bins increases, increasingly large portions of the space fall below the threshold, even in absence of anomalies, and so are flattened. By 2000 time bins, only a very small region of the parameter space is not flattened, which makes performing a model fit almost impossible.

This demonstrates that, especially in the case of time separated data, having an externally fixed threshold value can impede model fitting. Therefore, ideally the value of p should be dynamic, allowing it to be low in suboptimal regions of the parameter space to avoid overflagging and flattening the likelihood surface, and higher around optimal regions to avoid missing genuine anomalous points.

Therefore, as was suggested in Leeney et al. (2023), this issue can be resolved by fitting the value of p as a free parameter, which is estimated as a parameter, simultaneously alongside other parameters of interest by the chosen Bayesian numerical solver, rather than assigning it a fixed value. It is assigned a wide prior from its theoretical maximum of one down to effectively zero. This results in there being sections of the parameter volume where the threshold is low and so the variations in the likelihood surface with the other parameters are visible. As a result, the model fit can progress, optimizing the parameter values towards their posteriors while simultaneously optimizing the threshold towards the optimal posterior value of the probability that a point is anomalous for that data set. This resolves the issue described above and allows the fit to proceed while still accurately flagging anomalous points. The effect this additional nuisance parameter may have on the likelihood reweighting process can then be mitigated by marginalizing over the parameter before reweighting.

In the next section, tests of the entire process will be performed to demonstrate its efficacy. Throughout the rest of this work, p is fit as a free parameter with a logarithmically uniform prior in the range [1e−30 – 1]. The next section will demonstrate the functionality of this method.

3.3 Results

In order to demonstrate the performance of this anomaly mitigation procedure, four test data sets were generated according to the model described in Section 3.1, with $N_t = 2$, 20, 200, and 2000 time bins, respectively. These data sets will henceforth be referred to as the *uncontaminated* data. To each of these data sets, a random selection of anomalous peaks is added. In each case, the number of spikes added was equal to $N_t \times 5$, such that every data set is contaminated in equal proportion. The amplitudes were uniformly sampled from the range 10–50. This scale was chosen to provide a trial case for the process, in which the anomalies are sufficiently above the noise to be large enough to impact the model fit, but still relatively small compared with the overall scale of the data, so they cannot be trivially removed. The time bins in which each anomaly was placed were uniformly randomly chosen and the x bin was uniformly set to one of 40 randomly chosen channels. This constrained the contamination to a subset of the data channels, in a manner more resembling RFI. This will be discussed in more detail in Section 3.3.4. Fig. 5 shows an example of the anomalous points added to the data for $N_t =$ 200. These data sets will henceforth be referred to as *contaminated*. As the primary objective of this work is to present a method of fast correction for transient RFI, the contamination we test here is transient. However, this method will also compensate for time-constant and wideband RFI, which was demonstrated in Leeney et al. (2023).

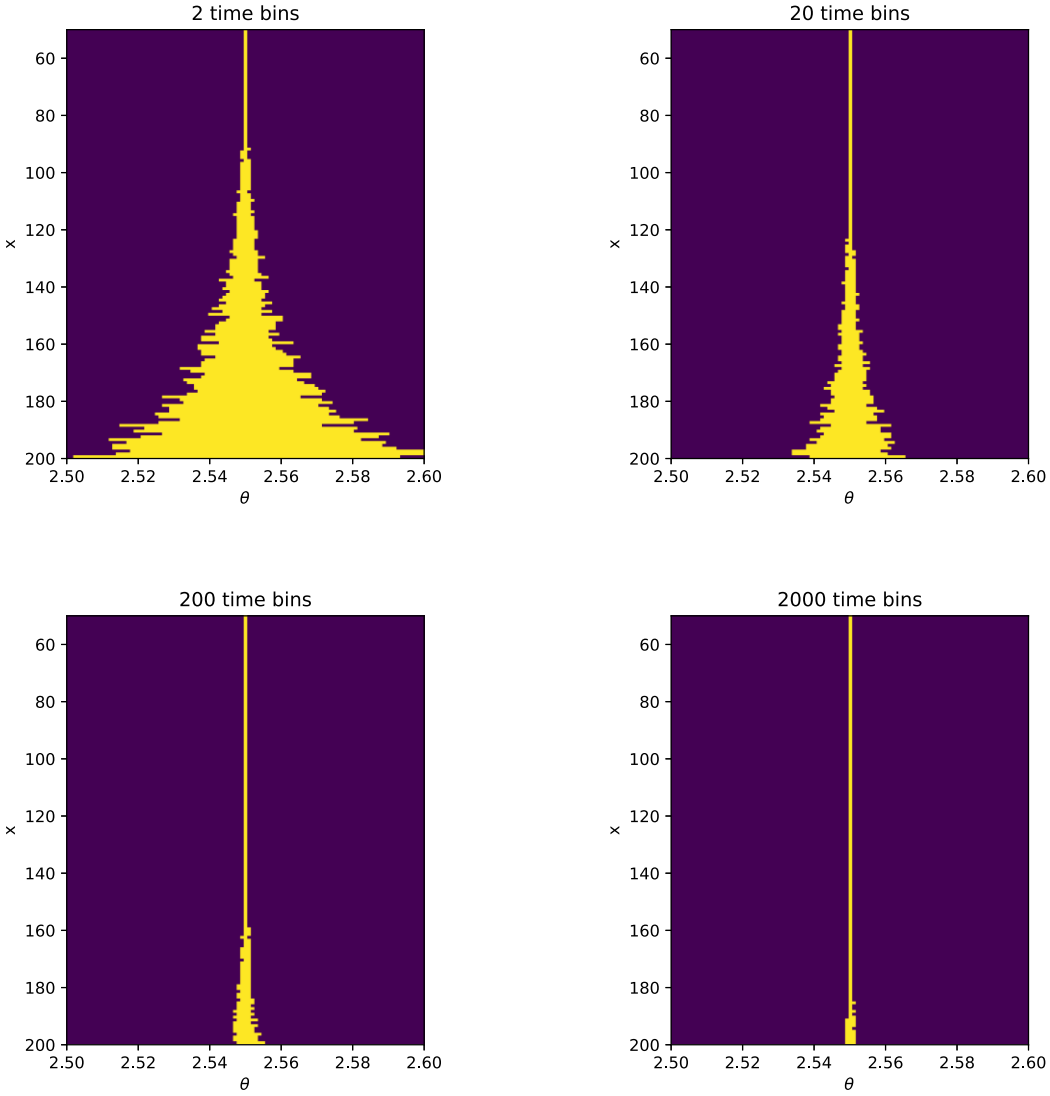

Figure 4. Plots of the time averaged likelihoods $\mathcal{L}_{ij}$ for simulated data sets generated according to equation (11), excluding the summation, with 2, 20, 200, and 2000 time bins. The parameter and x values that give likelihoods above a fixed threshold of p =1e−3 are highlighted, demonstrating that larger regions of parameter space fall below a fixed likelihood threshold as the number of time bins increases.

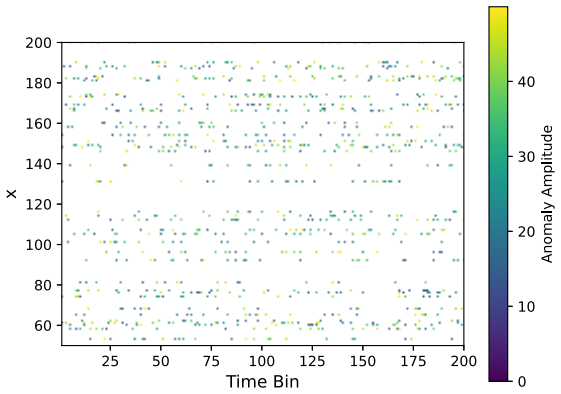

Figure 5. Plot of the anomalous points added to the $N_t = 200$ uncontaminated data set in order to produce the corresponding contaminated data set. The amplitudes were uniformly randomly chosen from the range 10–50. The time bin locations were uniformly randomly chosen and the x bin locations were uniformly set to one of 40 randomly chosen channels.

In all model fits performed in this paper, Δ is taken to be 7000, which is the maximum scale of the toy model data in all cases, as can be seen in Fig. 1. Furthermore, all fits were performed using the `PolyChord` nested sampling algorithm (Handley, Hobson & Lasenby 2015a, b), implemented in Python and run on a laptop on one CPU. As the core of this method is a modification to the likelihood and model of a Bayesian model fit; however, it can be implemented with any Bayesian fitting algorithm and will scale and parallelize as that algorithm does.

3.3.1 Anomaly correction

Two models were fit to each of the contaminated and uncontaminated data sets described in the previous section. The first was a direct fit of the model given in equation (35) and equation (36), with no attempt to correct for any anomalies. The second was a full fit of this model together with the time-dependent Bayesian anomaly flagging method described in the previous section, including the likelihood reweighting process and fitting for the threshold value as a parameter. Fig. 6 shows the posteriors on the parameter θ for each

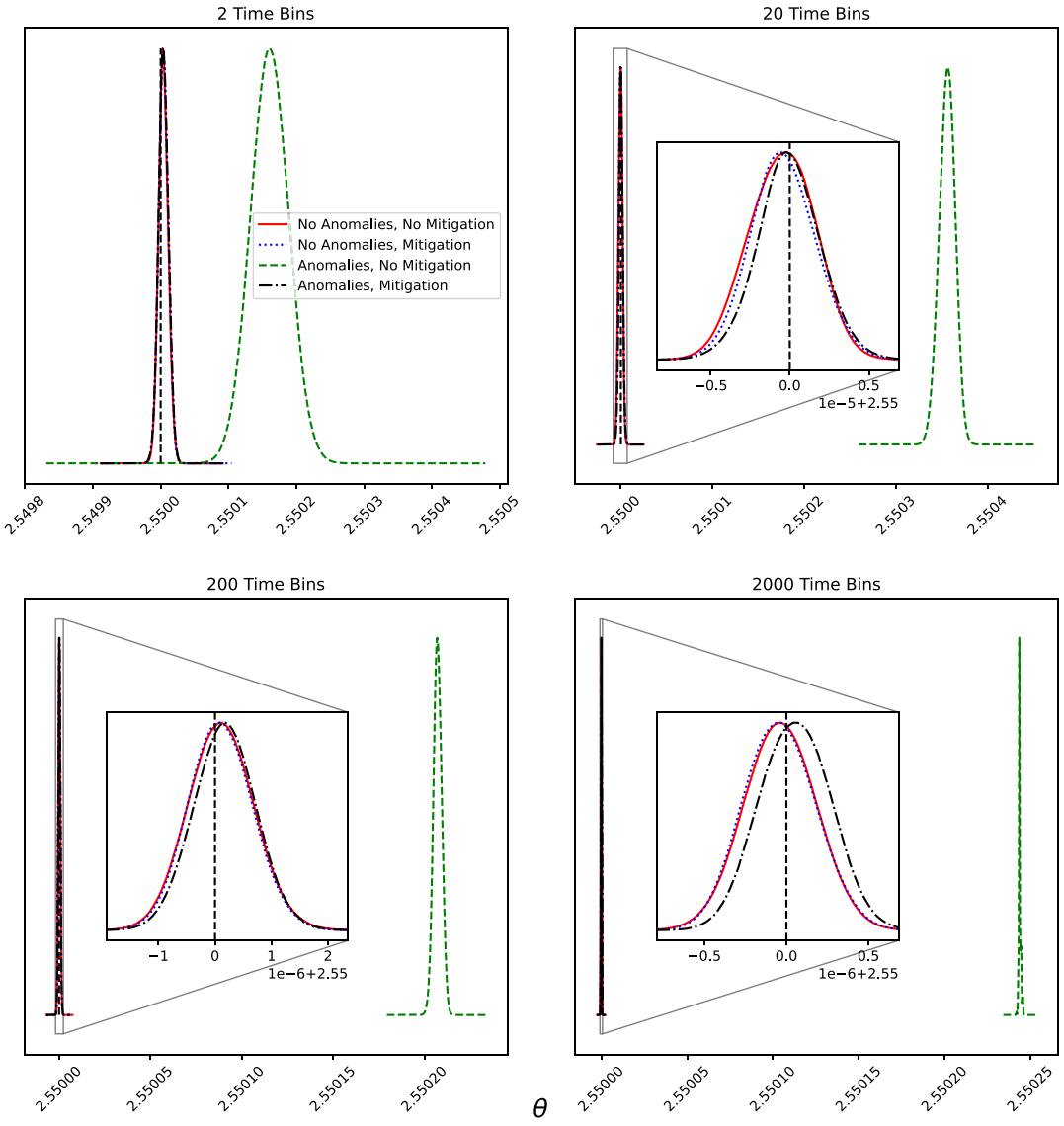

Figure 6. Plots of the posteriors on the toy model parameter θ, defined in equation (36), when fitting both uncontaminated and contaminated test data sets with models that include and do not include time-dependent Bayesian anomaly correction. The test data sets used a true value of $\theta = 2.55$, marked with a vertical dashed line. Each subfigure shows the results for a simulated data set with a different number of time bins, from 2 to 2000. This demonstrates that uncorrected anomalies lead to parameter biases which are corrected by the application of the anomaly mitigation process.

of these fits. Fig. 7 summarizes the standard deviations, σ_θ, of each of these posterior distributions and Fig. 8 shows the biases between the posterior means and the true parameter value of 2.55, expressed as a number of standard deviations.

For each number of time bins, the results of fitting the uncontaminated data sets without including any anomaly corrections, are shown in red in order to provide a benchmark. In all of these benchmark cases, the posterior correctly identifies the true parameter value of 2.55 to within $<1\sigma$, as can be seen in Fig. 8, with the standard deviation of those posteriors reducing approximately proportionally to the number of time bins, seen in Fig. 7, as could be expected.

The posteriors when the correction is applied but the data are uncontaminated are shown in blue. In these cases, it can be seen that the recovered posteriors are highly consistent with those of the uncorrected cases. This demonstrates that including the time-dependent Bayesian anomaly correction method does not bias the fit in absence of any anomalies. This is the expected result.

The posteriors generated from contaminated data sets but with no correction applied are shown in green. It can be seen that in all cases, the parameter posterior is biased from the true value. For $N_t = 2$, 20, 200, and 2000, Fig. 8 shows the true value falls at $5.8\sigma_\theta$, $42.7\sigma_\theta$, $94.1\sigma_\theta$, and $393.7\sigma_\theta$, respectively. This is again the expected result, demonstrating that the results will be biased if contamination is present but not accounted for. The offset increases with the number of time bins due to the high time bin cases having narrower posteriors, which makes the bias more apparent. It can also be seen in Fig. 7 that, although the posterior standard deviations in these cases still decreases proportionally to the number of time bins, the posteriors are consistently a factor of ~ 4 larger than the uncontaminated cases.

The posteriors when the correction is applied to contaminated data are shown in black. It can be seen that the proposed time-dependent Bayesian anomaly mitigation methodology has successfully countered the bias in the posterior seen in the contaminated but uncorrected cases. For $N_t = 2$, 20, 200, and 2000, Fig. 8 shows

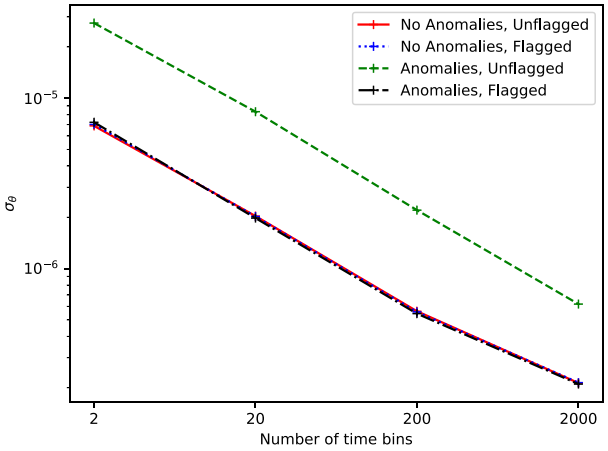

Figure 7. Plots of the standard deviations, σ_θ, of each of the posterior probability distributions shown in Fig. 6.

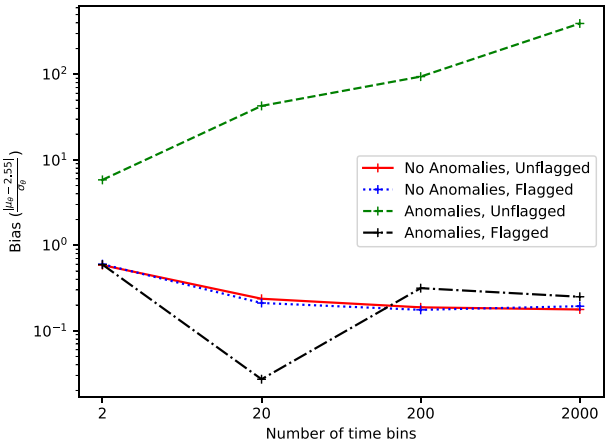

Figure 8. Plots of the biases between the posterior mean and the true fiducial parameter value, expressed as a number of standard deviations, for each of the posterior probability distributions shown in Fig. 6.

the true value now falls at $0.59\sigma_\theta$, $0.03\sigma_\theta$, $0.31\sigma_\theta$, and $0.25\sigma_\theta$, respectively. The true value of the parameter is therefore now recovered to within 1σ in all cases, with Fig. 6 showing that the posterior closely matches the uncontaminated benchmark in all cases.

This demonstrates that the proposed anomaly mitigation technique is successfully correcting the added anomalies in the data and enabling the model parameters to be recovered accurately. It is worth noting that in this test case the simulated data are heavily contaminated, with 3.3 per cent of data points featuring an anomaly and 26.5 per cent of x channels being contaminated to some extent. Despite this, anomalies are still correctly accounted for.

3.3.2 Anomaly recovery

Figs 6–8 demonstrate that our methods account for the presence of anomalous data points to a level sufficient that the underlying model can be accurately recovered. It is also worth assessing directly whether the points predicted to be anomalies correctly correspond to those added into the data.

This was investigated for the four corrected fits to contaminated data shown in black in Figs 6–8. In order to determine the accuracy with which the flagged anomalies correspond to the true anomalies, it is necessary to first determine which data points are being identified as anomalous by the algorithm. This was achieved, for each case, by evaluating the condition given in equation (15) for every posterior sample. Any data point where $\log \mathcal{L}_{ij} + \log(1 - p_{ij}) \leq \log p_{ij} - \log \Delta$ is considered anomalous for that sample. Evaluating the weighted average of all samples in that posterior then outputs a probability, for each data point, that that point is considered anomalous by the fit.

A given data point can then be considered flagged as anomalous by the fit if this probability exceeds a threshold. This threshold was set to 0.5, such that a point is considered flagged if it more probable that it is anomalous than not. The resulting flagged points for each number of time bins can be compared with the anomalies added into the simulated data to determine the accuracy of the flagging process.

Table 1 summarizes the number of points flagged as anomalous by the fitting that do or do not correspond to a real anomaly, true and false positives respectively, and the number of points not flagged that do or do not correspond to uncontaminated data, true and false negatives, respectively. It should be noted that the total number of anomalous data points does not exactly equal $5 \times N_t$ due to the random locations of the contaminated points occasionally resulting in overlap. It can be seen that for all four N_t values tested, the rate of both false positives and false negatives was zero. This results in both the precision, true positive/(true positive + false positive), and the recall, true positive/(true positive + false negative), being 1 for all cases. Therefore, the F_2 score, defined as the harmonic mean of these two values, giving twice as much weight to recall as precision,

$$F_2 = \frac{(1 + 2^2) \times \text{precision} \times \text{recall}}{2^2 \times \text{precision} + \text{recall}}, \tag{37}$$

is 1 for all cases.

It should also be noted that this level of precision and recall is maintained even if the probability at which a point is considered anomalous in the fit is dropped from 0.5 to 1e−10. This accuracy of anomaly recovery, even for these heavily contaminated toy models, raises the possibility that this anomaly correction methodology could also be used to detect anomalous points of interest, and thus functions as a transient flagger. This possibility will be explored further in Section 4.

3.3.3 Computation time

The motivation for implementing likelihood reweighting as described in Section 2.4 was to improve the computational efficiency of the proposed process and significantly reduce the otherwise strong dependency of the total runtime on the number of time bins.

In order to investigate the effects on runtime, the four contaminated data sets, with $N_t = 2$, 20, 200, and 2000, were fit to the corresponding model with the anomaly correction method implemented, but without using likelihood reweighting. Instead the full 'slow' likelihood defined in equation (15) was used for the entire fit.

Fig. 9 shows the means and standard deviations of the time, on the laptop CPU used, for a single evaluation of this slow likelihood for each of these fits, calculated from 1000 repeat evaluations of each likelihood. It can be seen that the evaluation time increases proportionally with the increasing number of time bins, as expected.

Fig. 9 also shows the means and standard deviations of the ratio of the total runtime of the fit with no likelihood reweighting to an equivalent fit on the same data set with likelihood reweighting implemented, evaluated from five repeats of each fit. Whilst using the full model fit directly is more efficient for cases with very few time bins, it can be seen that implementing likelihood reweighting reduces the runtime when $N_t > 20$, with the speed up following ap-

Table 1. Summary of the accuracy with which the anomalous data points inserted into each contaminated simulated data set were recovered by the flagging process, quantified by the number of true positives (points flagged as contaminated that were actually contaminated), false positives (points flagged that were not actually contaminated), false negatives (points not flagged as contaminated despite actually being contaminated), and true negatives (uncontaminated points that were correctly not flagged).

N_t	True positives	False positives	False negatives	True negatives
2	9	0	0	293
20	91	0	0	2929
200	946	0	0	29 254
2000	9433	0	0	292 567

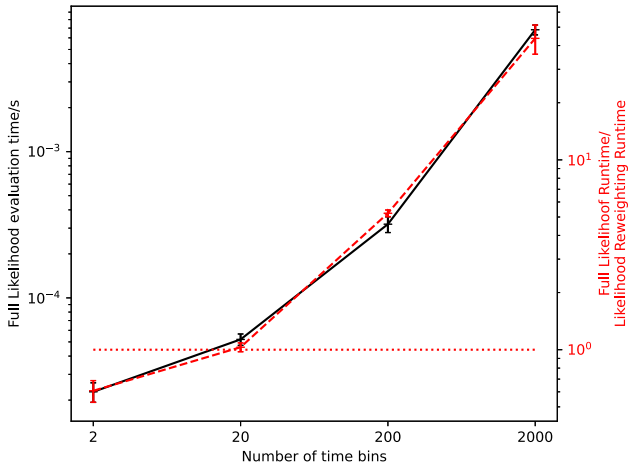

Figure 9. Plot of runtimes as a function of number of time bins in the data set. The time required, on the laptop CPU used, for a single likelihood evaluation of the full time-dependent anomaly mitigation method described in equation (15), calculated from 1000 likelihood evaluations in each case, is shown by the solid line. The ratio of the total runtime of a model fit implementing time-dependent Bayesian anomaly flagging using the full slow likelihood described in equation (15), to that of the equivalent fit with using likelihood reweighting, calculated from five repeats of each fit, is shown by the dashed line. The horizontal dotted line marks the ratio of 1, above which the likelihood reweighting methodology is faster than the full fit.

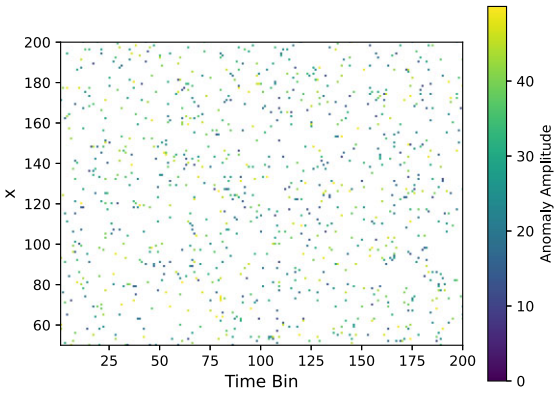

Figure 10. Plot of the anomalous points added to the $N_t = 200$ uncontaminated data set in order to produce the corresponding contaminated data set in which most x channels are contaminated to some degree. The amplitudes were uniformly randomly chosen from the range 10–50. The time bin and x bin locations were uniformly randomly chosen.

proximately the same trend as that of the likelihood evaluation time, increasing approximately proportionally to the number of time bins. For example, for 2000 time bins, the runtime improves by a factor of 44. This demonstrates that implementing likelihood reweighting successfully makes the total runtime close to independent of the number of time bins. This enables the underlying model fit to be performed in the efficient fashion described in Anstey et al. (2023) and Section 2.3 without being significantly slowed by the anomaly flagging, and thus enables the flagging process to be implemented efficiently on very large data sets, as was the objective.

Given that this runtime improvement is obtained from improving the likelihood evaluation time to be independent of the number of time bins, this trend is expected to continue for higher numbers of time bins. It should also be maintained with parallelization of the underlying model fitting algorithm.

3.3.4 Overcontamination

In Section 3.3.3, it was demonstrated that the process of likelihood reweighting described in Section 2.4 significantly improves the computation time for the proposed method. However, it also introduces a minor limitation that is not present if the full, slow likelihood is used.

When likelihood reweighting is implemented for this process, the initial model fit is performed using the 'fast' likelihood defined in equation (32). However, in this fast likelihood, in order for the evaluation time to not be dependent on the number of time bins as required, anomalies are identified in the product of the likelihoods of all data points in a given channel, flagging the entire data channel if the likelihood product is below the specified threshold. However, as a result of this, if the data are contaminated in such a way that all x channels are contaminated to some extent, it could result in every channel being flagged, thus giving no constraints on the parameters of interest.

In order to test this effect, a new set of contaminated data sets were generated, by adding anomalous data points to the four uncontaminated data sets described in Section 3.3, in exactly the same manner as previously, except the x bin of each contaminated point was chosen entirely randomly, rather than being confined to certain channels. Fig. 10 shows an example of the anomalous points added to the data for $N_t = 200$.

The tests described in Section 3.3.1 were repeated on these new randomly contaminated data sets. Fig. 11 shows the posteriors for the cases where the contaminated data sets were fit with the anomaly correcting likelihood, in comparison with the equivalent fits where the anomalies were confined to specific channels.

Given that the number of anomalous points injected was $5 \times N_t$ and the number of x channels in the simulated data sets are 151, for fully randomly distributed anomalies, the expected number of anomalous points per channel will be 0.07, 0.66, 6.62, and 66.23

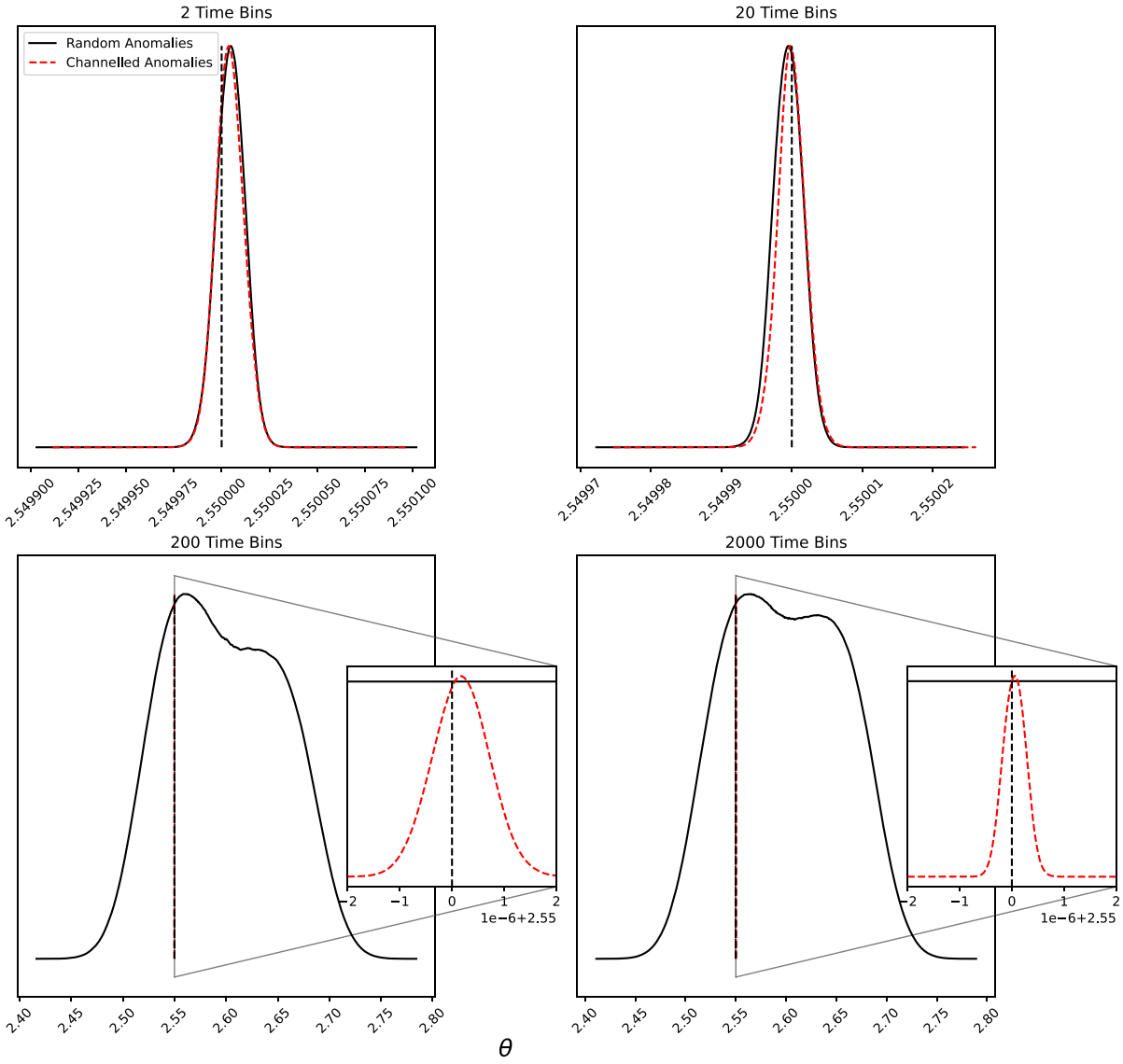

Figure 11. Plot of the parameter posteriors recovered when applying the Bayesian anomaly mitigation method with likelihood reweighting to simulated test data sets contaminated with $5N_t$ anomalous points located at random (solid line), in comparison to the recovered posteriors for performing the same fits on data instead contaminated with the same number of anomalous points but constrained to a maximum of 40 x channels (dashed line), previously shown in Fig. 6 (by a dash-dotted line). The true parameter value of 2.55 is indicated by the vertical dashed line.

for $N_t = 2$, 20, 200, and 2000, respectively. For all cases except $N_t = 2$, it is more likely that a channel be contaminated than not, with $N_t = 200$ and 2000 expecting one or more contaminated points in every channel. It can therefore be expected that for $N_t = 200$ and 2000, every channel will be flagged in the fast likelihood and thus no constraints on the parameter will be achieved.

It can be seen in Fig. 11 that, for the $N_t = 2$ and 20 cases, where the random anomalies are expected to have less than one contaminated point per channel, the parameter recovery for the random contamination is equivalent to the channel constrained contamination. However, for the $N_t = 200$ and 2000 cases, where the expected number of contaminated points per channel increases above 1, the expected flagging of every channel is seen in the random contamination case, which returns the prior on the parameter θ.

This over contamination failure can be overcome by utilizing the full, slow likelihood instead of implementing likelihood reweighting. However, doing so would prevent the improved calculation time the reweighting provides. The degree of contamination required for this limitation to become relevant is very high, particularly in the context of RFI as it requires most or all of the channels to show contamination. However, this does represent a limitation of the proposed methodology to improve flagging speed through likelihood reweighting, which may not be viable on extremely heavily contaminated data sets.

4. TRANSIENT DETECTION

In Section 3.3.2, it was demonstrated that along with enabling accurate identification of the underlying model from beneath contamination, the proposed method also gives accurate recovery of the anomalous points themselves. This raises the possibility that, in addition to applications in RFI excision, this method could also be applied to detect transient signals of interest. As the implementation of likelihood reweighting enables model fitting to have a compu-

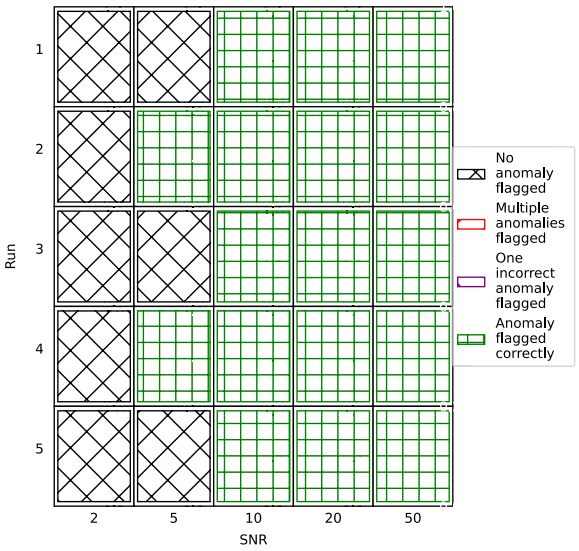

Figure 12. Plot of the attempts to identify a single added anomalous point into 2000 time bin data sets using the proposed method as a transient detector. Each panel shows the results for different SNR of the added anomalous point and for different repeats with the anomaly in a different random data bin. Cases where no anomalies were identified at all are shown in black with x hatching. Cases where the single anomaly was correctly flagged are shown in green with + hatching. Cases where multiple points were erroneously flagged and cases where only one point was flagged, but at the wrong location, would be shown in red with / hatching and purple with \ hatching, respectively. However, no such cases were seen.

tational time almost independent of the number of time bins, this process potentially provides a Bayesian methodology for efficiently searching large data sets for transients such as FRBs and pulsars.

4.1 Anomaly recovery

In order to test the ability of the method to correctly identify small transients in large data sets, a series of tests were run in which a single anomalous data point was added at random to the previously described 2000 time bin test data set, and the model fit with the flagging process implemented. This was repeated five times each for anomalous points with amplitudes of 0.5, 1.25, 2.5, 5, and 12.5, which correspond to SNRs of 2, 5, 10, 20, and 50 respectively.

The results of these fits are shown in Fig. 12. It can be seen that when acting as a flagger, this method correctly and reliably identifies the single anomalous point for the cases with SNR $\geq$ 10. Furthermore, it never erroneously identifies any data points as anomalous that were not anomalous in the data. However, the flagger does begin to fail to detect the anomalous point at lower SNRs, which can be expected given there is only one anomalous point out of 302 000 in this test case, meaning lower SNR anomalies can become indistinguishable from simple statistical fluctuations and so will not be flagged.

Although real transients such as FRBs and pulsars will often have additional structure and cover more data points than this simple, single contaminated data point test case, this minimal case demonstrates the potential for the proposed methodology to function as an efficient Bayesian transient detector. This will be explored in greater detail in a future work.

5. CONCLUSIONS

RFI is a significant challenge in radio astronomy. In this paper, we extend the Bayesian RFI mitigation methodology presented in Leeney et al. (2023) into the time domain. This enables transient anomalies to be flagged and properly accounted for in a Bayesian manner when fitting models to time-series data.

The process of likelihood reweighting was implemented in order to enable this process to be performed in a manner mostly independent of the number of time bins in the data. This was demonstrated to produce significant improvements in the computation time as the number of time bins increases, reaching a 44 times speed increase on a test case with 2000 time bins, by breaking the proportional relation between the number of time bins and the runtime.

Our methodology was demonstrated to be successful when correcting for contamination in a series of test data sets that approximate global 21-cm experiment data. It accurately corrected the bias in the model parameters that occurs if the contamination was not accounted for, while not affecting the results if no contamination is present.

Furthermore, it was demonstrated that our methods can correctly locate and extract anomalous points from data. Therefore, the efficacy of our methods as an efficient transient detector was explored. It was demonstrated that they were successfully able to identify a single anomalous point out of 302 000, provided the anomaly had an SNR of 10 or higher. The use of this process as a transient flagger will be explored in greater depth in a future work.

A potential limitation was also identified, in which the implementation of likelihood reweighting results in a failure to correctly account for anomalies in cases where every data channel is contaminated to some degree. Therefore, this method of improving computation time may not be viable on extremely heavily contaminated data sets, or in cases where most or all channels have a finite probability of showing some contamination and the data set has a very long time series. However, in such cases, the full slow likelihood can still be applied, albeit with a much longer computational time.

Overall, the methodology presented here represents an efficient and fully Bayesian technique for correcting time-dependent contamination or identifying transients in large data sets.

DATA AVAILABILITY

The data used and generated in this article will be shared on reasonable request to the corresponding author.

ACKNOWLEDGEMENTS

We would like to thank Will Handley for his work on the development of the original methodology. DA was supported by the Science and Technologies Facilities Council and SL was supported by the European Research Council and the UKRI.

REFERENCES

Anstey D., de Lera Acedo E., Handley W., 2021, MNRAS, 506, 2041
Anstey D., de Lera Acedo E., Handley W., 2023, MNRAS, 520, 850
Arrubarrena P., Lemercier M., Nikolic B., Lyons T., Cass T., 2024, preprint (arXiv:2402.14892)
Baan W. A., 2019, J. Astron. Instrum., 8, 1940010
Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67
Braun R., Bourke T., Green J. A., Keane E., Wagg J., 2015, Proc. Sci. Advancing Astrophysics with the Square Kilometre Array, Vol. 215. SISSA, Trieste, PoS#174

Cendes Y. et al., 2018, Astron. Comput., 23, 103
Czech D., Mishra A., Inggs M., 2018a, Astron. Comput., 25, 52
Czech D., Mishra A., Inggs M., 2018b, Radio Sci., 53, 656
de Lera Acedo E. et al., 2022, Nature Astron., 6, 984
Finlay C., Bassett B. A., Kunz M., Oozeer N., 2023, MNRAS, 524, 3231
Ford J. M., Buch K. D., 2014, Proc. 2014 IEEE Geosci. Remote Sens. Symp., RFI Mitigation Techniques in Radio Astronomy. IEEE, Quebec City, p. 231
Fridman P., Baan W., 2001, A&A, 378, 327
Handley W. J., Hobson M. P., Lasenby A. N., 2015a, MNRAS, 450, L61
Handley W. J., Hobson M. P., Lasenby A. N., 2015b, MNRAS, 453, 4384
Kerrigan J. et al., 2019, MNRAS, 488, 2605
Leeney S. A. K., Handley W. J., Acedo E. d. L., 2023, Phys. Rev. D, 108, 062006
McKinnon M., Beasley A., Murphy E., Selina R., Farnsworth R., Walter A., 2019, BAAS, 51, 81
Mesarcik M., Boonstra A.-J., Ranguelova E., van Nieuwpoort R. V., 2022, MNRAS, 516, 5367
Monsalve R. A. et al., 2024, MNRAS, 530, 4125
Nan R., 2006, Sci. China G, 49, 129
Nita G. M., Keimpema A., Paragi Z., 2019, J. Astron. Instrum., 8, 1940008
Offringa A. R., de Bruyn A. G., Biehl M., Zaroubi S., Bernardi G., Pandey V. N., 2010, MNRAS, 405, 155
Payne E., Talbot C., Thrane E., 2019, Phys. Rev. D, 100, 123017
Pritchard N. J., Wicenec A., Bennamoun M., Dodson R., 2024, Publ. Astron. Soc. Aust., 41, e028
Razavi-Ghods N. et al., 2023, preprint (arXiv:2307.00099)
Romero-Shaw I. M., Lasky P. D., Thrane E., 2019, MNRAS, 490, 5210
Röttgering H., 2003, New Astron. Rev., 47, 405
Scaife A., 2020, Phil. Trans. R. Soc. A, 378, 20190060
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Singh S. et al., 2022, Nature Astron., 6, 607
Smith E., Lynch R. S., Pisano D., 2022, AJ, 164, 123
Wang Y., Zhang Z., Zhang H., Zhu W., Li D., Wang P., 2022, Astron. Comput., 39, 100568

SUPPORTING INFORMATION

Supplementary data are available at *RASTAI* online.

ywhmxcfkcxvfmyjxjvfysqkyhsgmtbzx.zip

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TEX/LATEX file prepared by the author.