



This is the accepted manuscript made available via CHORUS. The article has been published as:

# Transdimensional Bayesian approach to pulsar timing noise analysis

J. A. Ellis and N. J. Cornish

Phys. Rev. D **93**, 084048 — Published 27 April 2016

DOI: [10.1103/PhysRevD.93.084048](https://doi.org/10.1103/PhysRevD.93.084048)

# A Trans-dimensional Bayesian Approach to Pulsar Timing Noise Analysis

J. A. Ellis\*

*Jet Propulsion Laboratory, California Institute of Technology,  
4800 Oak Grove Drive, Pasadena, CA 91109, USA<sup>†</sup>*

N. J. Cornish

*Department of Physics, Montana State University, Bozeman, MT 59717, USA*

The modeling of intrinsic noise in pulsar timing residual data is of crucial importance for Gravitational Wave (GW) detection and pulsar timing (astro)physics in general. The noise budget in pulsars is a collection of several well studied effects including radiometer noise, pulse-phase jitter noise, dispersion measure (DM) variations, and low frequency spin noise. However, as pulsar timing data continues to improve, non-stationary and non-powerlaw noise terms are beginning to manifest which are not well modeled by current noise analysis techniques. In this work we use a trans-dimensional approach to model these non-stationary and non-powerlaw effects through the use of a wavelet basis and an interpolation based adaptive spectral modeling. In both cases, the number of wavelets and the number of control points in the interpolated spectrum are free parameters that are constrained by the data and then marginalized over in the final inferences, thus fully incorporating our ignorance of the noise model. We show that these new methods outperform standard techniques when non-stationary and non-powerlaw noise is present. We also show that these methods return results consistent with the standard analyses when no such signals are present.

PACS numbers:

## I. INTRODUCTION

Pulsar timing has been used as a tool to probe important aspects of astrophysics and fundamental physics including the indirect detection of gravitational waves (GWs) [1], the precise determination of neutron star masses [2], precise tests of General Relativity and alternative theories of gravity [3], constraints on equation-of-state physics, and several other areas [4]. Furthermore, the goal of modern PTAs is the direct detection of low-frequency gravitational waves (GWs) which could usher in a new era of astronomy and astrophysics [5].

As pulsar timing becomes more precise, increasingly robust and sophisticated noise modeling will be needed in order to make (astro)physical inferences from the data. In the last several years noise modeling has become an active area of research within the pulsar community [6–13] and has proven crucial to current timing and GW detection efforts [14–20]. However, even with current sophisticated methods there are still several areas in which our current noise models are inadequate:

1. We currently have no model for non-white non-stationary noise features which could result from a transient instrumental or astrophysical source.
2. We have no robust way of determining the validity or the goodness of fit of our low-frequency red-noise modeling.

In this work we seek to address these two problems by using a trans-dimensional approach in which a variable number of wavelets model the non-white, non-stationary features and a variable number of interpolation control points determine the spectrum of the stationary red noise. Inferences on any parameters of interest (i.e. timing model, GW parameters, etc.) are then marginalized over all possible models (i.e. marginalized over the number of wavelets and/or control points) thereby incorporating our uncertainty in the noise *model* itself as opposed to the standard practice of choosing *one* model and incorporating the uncertainty from the *parameters* within that model. There is precedent for this type of analysis in other GW experiments, namely the **BayesWave** [21] and **BayesLine** [22] algorithms used in LIGO/Virgo analyses, both of which serve as a starting point for this work.

The paper is organized as follows: In Section II we review current Bayesian data analysis techniques used in pulsar timing. In Sections III and IV we introduce the non-stationary and adaptive spectral modeling techniques. In Section V we review the Bayesian algorithms used to implement our trans-dimensional models. In Section VI we test our analysis on simulated data and compare against traditional techniques. Finally we conclude in Section VII.

## II. STANDARD PULSAR TIMING DATA ANALYSIS

The basic data analysis techniques used in this work are very similar to that presented in [15, 18]. Here we will briefly review the formalism and set up the likelihood function for this work.

---

\*Electronic address: Justin.A.Ellis@jpl.nasa.gov

<sup>†</sup>Einstein Fellow

### A. Likelihood

Pulsar timing data consists of a set of times-of-arrival (TOAs) or the radio pulses and a timing model. Our  $N_{\text{TOA}}$  timing residuals (difference of measured and modeled TOAs) for a single pulsar  $\delta\mathbf{t}$  can be broken into individual components as follows:

$$\delta\mathbf{t} = M\boldsymbol{\epsilon} + F\mathbf{a} + U\mathbf{j} + \mathbf{s} + \mathbf{n}. \quad (1)$$

The term  $M\boldsymbol{\epsilon}$  describes inaccuracies in the subtraction of the timing model, where  $M$  is called the timing model design matrix, and  $\boldsymbol{\epsilon}$  is a vector of small timing model parameter offsets. The term  $F\mathbf{a}$  describes all low-frequency signals, including low-frequency (“red”) noise, with a limited number of Fourier coefficients  $\mathbf{a}$ . Our harmonics are chosen as integer multiples of the harmonic base frequency  $1/T$ , with  $T$  the length of our dataset. The matrix  $F$  then has alternating sine and cosine functions. The term  $U\mathbf{j}$  describes noise that is fully correlated for simultaneous observations at different observing frequencies, but fully uncorrelated in time. The matrix  $U$  is an  $N_{\text{TOA}} \times N_{\text{epoch}}$  matrix that maps the  $N_{\text{epoch}}$  observation sessions to the  $N_{\text{TOA}}$  TOAs. The vector  $\mathbf{j}$  describes the white noise per observation session that is fully correlated across all observing frequencies. The term  $\mathbf{s}$  is simply any other deterministic feature that we choose to include in our model. The last term,  $\mathbf{n}$ , describes Gaussian white noise that is assumed to remain in the data, after correcting for all known systematics. The white noise is assumed to be an uncorrelated and Gaussian as is described in [15, 18].

We define the noise-mitigated timing residuals as

$$\mathbf{r} = \delta\mathbf{t} - M\boldsymbol{\epsilon} - F\mathbf{a} - U\mathbf{j} - \mathbf{s}, \quad (2)$$

where  $\mathbf{r}$  is our best approximation of  $\mathbf{n}$ , given our knowledge of all the noise and signal parameters. The likelihood can now be written

$$p(\delta\mathbf{t}|\boldsymbol{\epsilon}, \mathbf{a}, \mathbf{j}, \phi) = \frac{\exp(-\frac{1}{2}\mathbf{r}^T N^{-1}\mathbf{r})}{\sqrt{\det(2\pi N)}}, \quad (3)$$

where  $N$  is the covariance matrix of the white noise. We have collectively denoted all parameters not directly represented by  $\boldsymbol{\epsilon}$ ,  $\mathbf{a}$ , and  $\mathbf{j}$  as  $\phi$ . We group the linear signals as follows:

$$T = \begin{bmatrix} M & F & U \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \boldsymbol{\epsilon} \\ \mathbf{a} \\ \mathbf{j} \end{bmatrix}, \quad (4)$$

which allows us to elegantly place a Gaussian prior on the coefficients of these random processes. The prior covariance is:

$$B = \begin{bmatrix} \infty & 0 & 0 \\ 0 & \varphi & 0 \\ 0 & 0 & \mathcal{J} \end{bmatrix}, \quad (5)$$

resulting in a prior:

$$p(\mathbf{b}|\phi) = \frac{\exp(-\frac{1}{2}\mathbf{b}^T B^{-1}\mathbf{b})}{\sqrt{\det(2\pi B)}}, \quad (6)$$

where  $\infty$  is a diagonal matrix of infinities, which effectively means we have a uniform unconstrained prior on the timing model parameters  $\boldsymbol{\epsilon}$ . As described in [15], this representation allows us to analytically marginalize Eq. (3) times Eq. (6) over the waveform coefficients  $\mathbf{b}$  resulting in the marginalized likelihood

$$p(\delta\mathbf{t}|\phi) = \frac{\exp(-\frac{1}{2}(\delta\mathbf{t} - \mathbf{s})^T C^{-1}(\delta\mathbf{t} - \mathbf{s}))}{\sqrt{\det(2\pi C)}}, \quad (7)$$

with  $C = N + TBT^T$ . The Woodbury matrix identity [23] can be used to evaluate Eq. (7) efficiently. Notice that this is nearly identical to the likelihood of [15, 18] except that now we are including the additional non-linear deterministic feature  $\mathbf{s}$ .

The parameters that describe  $B$  are the hyperparameters  $\phi$ . The hyperparameters of the diagonal matrix  $\mathcal{J}$  are the per-backend TEMPO2 ECORR parameters. The matrix  $\varphi$  represents the spectrum of the low-frequency noise Denoting frequency bin with  $(i, j)$ , we can write:

$$[\varphi]_{ij} = \frac{1}{T} S_n(f_i) \delta_{ij}, \quad (8)$$

where  $S_n(f_i)$  is the noise power spectral density (PSD) at frequency  $f_i$ .

### B. Bayesian analysis

Bayesian inference is a method of statistical inference in which Bayes rule of conditional probabilities is used to update one’s knowledge as observations are acquired. Given a model  $\mathcal{M}$ , model parameters  $\Theta$ , and observations  $d$ , we write Bayes rule as:

$$p(\Theta|d, \mathcal{M}) = \frac{p(d|\Theta, \mathcal{M})p(\Theta|\mathcal{M})}{p(d|\mathcal{M})} \quad (9)$$

where  $p(\Theta|d, \mathcal{M})$  is the posterior probability (probability of obtaining the model parameters  $\Theta$  given the data  $d$  and model  $\mathcal{M}$ ),  $p(d|\Theta, \mathcal{M})$  is the likelihood function (probability of obtaining the data  $d$  given the model parameters  $\Theta$  and model  $\mathcal{M}$ ),  $p(\Theta|\mathcal{M})$  is the prior probability of the model parameters  $\Theta$  given a model  $\mathcal{M}$ , and  $p(d)$  is the marginal likelihood or the Bayesian evidence, usually denoted by  $\mathcal{Z}$ .

The left-hand side of Eq. (9) can be regarded as the “output” of the Bayesian analysis, and the right-hand side is the “input” modulo the evidence term. Indeed, provided we have a generative model of our observations (meaning we can simulate data, given the model parameters), we know the likelihood and prior. However, for

parameter estimation we would like to know the posterior, and for model selection we need the evidence.

For parameter estimation, the evidence is usually ignored, and one can use  $p(d|\Theta, \mathcal{M})p(\Theta|\mathcal{M})$  directly to map out the posterior distribution (up to a normalizing constant) and to estimate confidence intervals. When  $\Theta$  is higher-dimensional, Monte-Carlo sampling methods are typically used to perform this multi-dimensional integral. We use Markov Chain Monte Carlo methods in this work to sample the posterior distribution.

Model selection between two models  $\mathcal{M}_0$  and  $\mathcal{M}_1$  can be carried by calculating the ‘‘Bayes factor’’: the ratio between the evidence for the two models. Assuming we have a prior degree of belief of how likely the two model are ( $p(\mathcal{M}_0)$  and  $p(\mathcal{M}_1)$ ), we can write:

$$O = \frac{p(\mathcal{M}_1|d)}{p(\mathcal{M}_0|d)} = \mathcal{B}_{10}(d) \frac{p(\mathcal{M}_1)}{p(\mathcal{M}_0)}, \quad (10)$$

where  $\mathcal{B}_{10}(d) \equiv \mathcal{Z}_1/\mathcal{Z}_0$  is the Bayes factor, and  $O$  is the odds ratio. The odds ratio can be obtained by calculating the evidence  $\mathcal{Z}$  for each model separately (e.g. with Nested Sampling or thermodynamic integration), or by calculating the Bayes factor  $\mathcal{B}$  between two models directly (e.g. with trans-dimensional markov chain Monte Carlo methods).

### III. NON-STATIONARY NOISE MODEL

To model non-stationary noise events we use a non-orthogonal representation of Morlet-Gabor wavelets with a functional form

$$\Psi(t; A, f_0, Q, t_0, \phi_0) = A e^{-(t-t_0)^2/\tau^2} \cos(2\pi f_0(t-t_0) + \phi_0), \quad (11)$$

where  $\tau = Q/(2\pi f_0)$  is the width of the gaussian envelope,  $A$  is the amplitude,  $t_0$  and  $f_0$  are the central time and frequency of the wavelet and  $\phi_0$  is a phase offset. This representation is simply a matter of choice and could be replaced by any other basis, such as shapelets [24]. Morlet-Gabor wavelets were chosen as they are simple and widely used in gravitational wave data analysis. They also have the smallest possible time frequency volume. We expect that a wide range of continuous wavelet representations would yield similar performance. Thus, to model non-stationary noise events we use a collection of independent wavelets where the number of wavelets,  $N_{\text{wave}}$ , is a free parameter that must be constrained by the data. In principle, this analysis is similar to that used in [25, 26] where shapelets are used to model the pulse profile; however, in this case the number of wavelets is free to vary and is *marginalized* over in the final inference, where in the shapelet analysis separate runs must be done in order to find the number of shapelets with the highest Bayes factor.

As can be seen from Eq. (11), the Morlet-Gabor wavelet is a function of five parameters, thus this analysis includes an additional  $5N_{\text{wave}}$  over the standard gaussian

noise model. Our RJMCMC implementation will therefore balance the goodness-of-fit for the likelihood with the model complexity (i.e., prior volume) of the additional wavelets. Getting the correct balance of the two aforementioned effects is a conceptually and computationally difficult problem that depends heavily on the choice of priors on the wavelet parameters and on the number of wavelets.

#### A. Wavelet Priors

As we will see, the priors on the wavelet amplitude are quite important in balancing goodness-of-fit with the model complexity. While the priors on the other parameters will likely have some effect on our trans-dimensional model, for this work we use uniform priors on  $\log_{10}(f_0)$ ,  $Q$ ,  $t_0$ , and  $\phi_0$  with ranges  $\log_{10}(f_0) \in [\log_{10}(3/T), \log_{10}(1/4\Delta t)]$ , with  $\Delta t = 2 \text{ weeks}^{-1}$ ,  $Q \in [0.5, 40]$ ,  $t_0 \in [t_{\text{start}}, t_{\text{stop}}]$ , where  $t_{\text{start}}$ , and  $t_{\text{stop}}$  are the start and end times of the residuals time-series, and  $\phi_0 \in [0, 2\pi]$ . Here we will only focus on choosing a good amplitude prior that in effect will limit the number of wavelets used in the RJMCMC. When doing standard fixed dimension noise analysis it is standard to use a log-uniform prior on the amplitude of deterministic sources unless we have some a-priori information; however, in this trans-dimensional model a log-uniform prior will lead to many wavelets being used that essentially just explore the prior and contribute little to the signal being modeled. As is used in the **BayesWave** algorithm [21], a more reasonable prior is one that is based on the SNR of the wavelet

$$\rho^2 = (\Psi|\Psi) = \Psi^T C^{-1} \Psi, \quad (12)$$

where  $\Psi$  is the wavelet waveform and  $C$  is the full covariance matrix from Eq. (7). In practice, computing the SNR as defined above carries nearly the same computational cost as that of computing the likelihood. Furthermore, to use the SNR in a prior we will need to compute it for *every* wavelet used in the model. Therefore, it is computationally undesirable to use the full SNR but to instead use an SNR proxy that is computationally efficient. For this work we choose a simple white noise inner product

$$\bar{\rho}^2 = \Psi^T N^{-1} \Psi = A^2 \bar{\Psi}^T N^{-1} \bar{\Psi}, \quad (13)$$

where  $N$  is the diagonal white noise matrix and  $\bar{\Psi}$  is the amplitude-free waveform. In the simulations presented later in this paper we have found that this overestimates the true SNR by a factor of two in many cases; however this proxy is very fast to compute and scales similarly with the other wavelet parameters as the true SNR. Using this SNR proxy, we parameterized the signal in terms of  $\bar{\rho}$  as

$$\Psi(t) = \frac{\bar{\rho}}{(\bar{\Psi}^T N^{-1} \bar{\Psi})^{1/2}} \bar{\Psi}(t), \quad (14)$$

where  $\bar{\rho}$  is now the free parameter. As was done in [21] we would like to use a parameterized prior on  $\bar{\rho}$  that peaks at some characteristically detectable value and slowly decreases as  $\bar{\rho}$  gets large; however, since we are using an SNR-proxy the choice of the peak in the distribution is problem specific and is very difficult to determine without some a-priori knowledge. For this reason we have chosen a simple uniform prior of  $\bar{\rho} \in [0, 100]$  for this work. Finally we choose a uniform prior on the number of active wavelets,  $N_{\text{wave}} \in [0, 30]$ . We have experimented with various exponential priors on  $N_{\text{wave}}$  but we have found that these have little effect and the choice of wavelet amplitude prior is the dominant factor in constructing a parsimonious model. Further consideration of these prior choices will be addressed in a future work.

#### IV. ADAPTIVE SPECTRAL MODELING

As mentioned in Section I, current noise analysis methods rely on specifying a model a-priori and then performing Bayesian model selection to determine the best model. However, in many cases several different spectral models seem to describe the data equally well. This uncertainty in the model and the computational complexity of performing several different runs to determine the Bayesian evidence has led us to develop a modified version of the **BayesLine** algorithm [22] to model the red noise in our residual data. Here we model the power spectral density (PSD)<sup>1</sup> with linear interpolation in the log-log space of the the PSD and frequency. The parameters of the model are the number,  $N_c$ , and the location in the frequency-PSD space  $\{S_n(f_i), f_i\}$  of each control point in the linear interpolation. Thus the logarithm of the PSD at frequency  $f_i$  is modeled by

$$\log(S_n(f_i)) = \log(S_n(f_{i-1})) + \frac{\log(S_n(f_{i+1})/S_n(f_{i-1}))}{\log(f_{i+1}/f_{i-1})} \log(f_i/f_{i-1}) \quad (15)$$

for  $i \in [0, N_c]$ . We have chosen to carry this interpolation out in the log frequency-PSD space in order to facilitate numerical stability and in order to fit the fewest parameters. In the course of this analysis we have tried other interpolation methods, including cubic splines as are used in **BayesLine** [22] in both log and linear frequency-PSD space. It was found that our pulsar timing residual data does not have enough frequency evolution to warrant a more sophisticated interpolation technique. We note that even though we may not have a free control point at every frequency we do in fact include all frequencies in the analysis differentiating this method from those of [10] in which various combinations of frequencies were used in order to determine the so-called “optimal” model.

In principle, one could allow the frequencies at which we place the control points to be free parameters but this could lead to many ordering and double counting problems. Instead we place the frequencies on a regularly spaced grid with  $\Delta f = 1/T$  in the range  $[1/T, n_f/T]$ , where  $n_f$  is the number of frequencies. In our case we use  $n_f = 50$  unless stated otherwise. We assume uniform priors on both the logarithm of the PSD amplitude,  $S_n(f_i)$ , and on the number of active control points,  $N_c$ . The control points at the lowest and highest frequencies are always active, thus the simplest spectrum is parameterized by two parameters in this setup. Therefore, this analysis encapsulates both the standard power-law and free spectrum analyses which are parameterized by 2 and  $n_f$  parameters, respectively. In practice, during the RJMCMC, control points can be subtracted and added with varying amplitudes. This means that the data decide the shape and complexity of the spectrum instead of prescribing a model a-priori.

#### V. BAYESIAN TECHNIQUES

##### A. Reverse Jump MCMC

The main driver of this trans-dimensional noise modeling technique is the RJMCMC algorithm, a variant of traditional MCMC where the model itself is included the explored parameter space. Traditionally RJMCMC has been used as a model selection tool by computing the Bayes factor of competing models as the ratio of the iterations spent in each model. In our analysis it is used more as a tool to perform a type of Bayesian model averaging. In other words, we are more interested in the ability of the RJMCMC to map out the trans-dimensional distribution and provide marginalized distributions over the full model-space than we are interested in choosing a *specific* model via the resulting Bayes factors of the RJMCMC.

In the RJMCMC framework we must include a separate Metropolis Hasting step that proposes the trans-dimensional move. The parameters  $\Theta_i$  of model  $\mathcal{M}_i$  are drawn from the proposal distribution  $q(\Theta_i|\mathcal{M}_i)$ . Since our model is nested (i.e. there are shared parameters between models), we hold all other parameters fixed while the new parameters are drawn from the proposal distribution. The criteria for accepting this trans-dimensional jump is the Hastings ratio

$$\mathcal{H}_{\mathcal{M}_i \rightarrow \mathcal{M}_j} = \frac{p(d|\Theta_j, \mathcal{M}_j)p(\Theta_j|\mathcal{M}_j)q(\Theta_i|\mathcal{M}_i)}{p(d|\Theta_i, \mathcal{M}_i)p(\Theta_i|\mathcal{M}_i)q(\Theta_j|\mathcal{M}_j)} |J_{ij}|, \quad (16)$$

where the Jacobian  $|J_{ij}|$  accounts for any change in dimension across models. However, if our jump proposals yield parameter values directly instead of a set of random numbers that are then used to determine the new model parameters, then the Jacobian can be neglected. Choosing a good proposal distribution is notoriously difficult and is the main drawback in using RJMCMCs. A simple

<sup>1</sup> When we refer to the PSD here we are referring only to the red-noise component and not the white component of the spectrum.

choice for a RJ proposal distribution is a random draw from prior. If the models are nested and the difference in dimension between models is relatively low then this prior draw usually provided adequate mixing.

As such, we only use prior draws for our RJ proposals in this work, for both the wavelet parameters and interpolation control points. Furthermore, we always propose to either add or subtract one component (i.e., one wavelet or one control point) and never more. In principle we could choose to add several wavelets or control points at one time but it is unlikely to be accepted. To add a new wavelet, we simply draw all 5 wavelet parameters from the prior (using either the SNR or the log-uniform prior for the amplitude). To subtract a wavelet, we randomly choose between the active wavelets at that iteration. The technique is very similar for control points. To add a new control point, we first randomly choose a frequency at which to place the control point and then draw from the log-uniform prior to choose the amplitude. To subtract a control point we randomly choose one from the active control points. When modeling both wavelets and control points simultaneously, we always hold one fixed and jump in the other, that is, we never propose to add or subtract wavelets and control points simultaneously. Of course, more complicated and more efficient proposals could be made for both the wavelet and spectral models and this will be addressed in the future; however, for this work we find that these random uniform draws provide quite adequate mixing.

## B. Parallel Tempering

Although uniform trans-dimensional proposals provide adequate mixing, the trans-dimensional acceptance rates are quite low (typically  $< 6\%$  for the wavelet model). This means that we have to run for many iterations in order to adequately explore the full trans-dimensional model space. In order to ameliorate this problem we use a parallel tempering method introduced in [21] that allows different chains to be in different models, thus jumps between different temperature chains are also trans-dimensional jumps. These parallel tempering moves *greatly* enhance the inter- and intra-model mixing. Currently we use a geometric temperature spacing that is tuned to give  $\sim 40\%$  acceptance for temperature swaps to the coldest chain. In future versions of the code we will incorporate a more sophisticated adaptive temperature spacing as was introduced in [27].

## VI. RESULTS

Here we test our non-stationary and adaptive noise modeling techniques. In all cases we have used the time stamps and timing model for PSR B1855+09 from the NANOGrav 9-year data release [15] and inject white gaussian noise consistent with the TOA uncertainties us-

ing LIBSTEMPO<sup>2</sup>. We use this as the base data set because it has all of the features of real pulsar timing data including a large gap, heteroscedastic TOA uncertainties, a full relativistic binary timing model, and time varying DM variations. We then take this base data set and inject a white noise burst, a realization of a gaussian process described by some power spectral density, or both, in order to test our modeling techniques. We then compare the results with those of standard noise analyses. Note that in all cases we have used a full white noise model including one EFAC, EQUAD, and ECORR per backend as described in [15].

### A. BayesWavePTA

Here we test our non-stationary noise modeling techniques described in Section III. As mentioned above we test this technique on a simulated dataset of pulsar B1855+09 that contains an injected white noise burst with amplitude of 447ns and a width of 250 days. A white noise burst is simply a realization of noise from a white PSD with a gaussian window that is localized in time. We have analyzed these data with the non-stationary model using both a uniform SNR and a uniform log-amplitude prior. In both cases we allow up to 30 wavelets with a uniform prior on the number of wavelets. The results of this analysis are shown in Figures 1 and 2.

In the top panels of Figures 1 and 2 we have the daily-averaged residuals plotted in blue, the injected waveform in green, the MAP waveform (chosen from the model with the highest Bayes factor) in red, and the 90% uncertainty on the waveform (marginalized over all models) in gray. The bottom left plot we show the utility as a function of number of wavelets. Here utility is simply the ratio of the number of iterations spent in a model with a given number of wavelets to the total number of iterations in the RJMCMC run. In the bottom right, we plot the histogram of the normalized residuals (normalized by the TOA uncertainties) when including (blue) and not including (red) the MAP wavelet model. The green curve is a zero-mean unit variance Gaussian distribution.

First, we point out that although this “event” seems quite large, it is quite comparable to a real event in the published data of PSR B1855+09 [see e.g. Figure 27 of 15]. Next, we see that for both choices of amplitude prior we recover the injected waveform very well and recover the Gaussian nature of the underlying white noise shown in the bottom right panels. Most important, however, is the number of wavelets that the data prefer under the two different amplitude priors and how this effects the uncertainty in the recovered waveform. In the case where we have used a uniform prior on the SNR of the wavelet

<sup>2</sup> <https://github.com/vallis/libstempo>

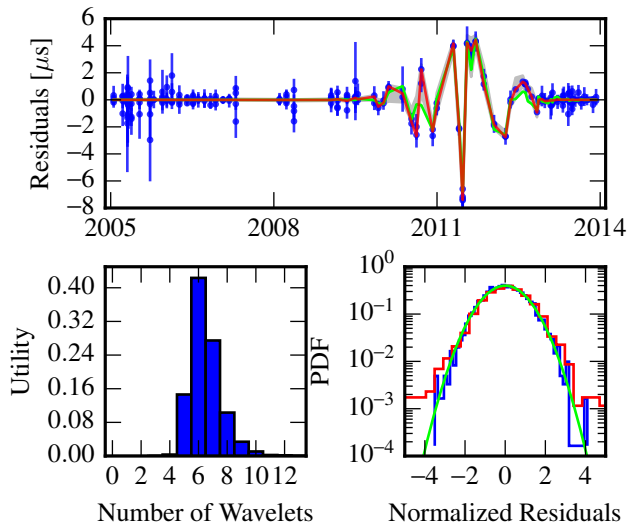


FIG. 1: Summary of RJMCMC results from the uniform SNR prior run. In the top panels the daily-averaged residuals are plotted in blue, the injected waveform in green, the MAP waveform (chosen from the model with the highest Bayes factor) in red, and the 90% uncertainty on the waveform (marginalized over all models) in gray. The bottom left plot we show the utility as a function of number of wavelets. In the bottom right, we plot the histogram of the normalized residuals (normalized by the TOA uncertainties) when including (blue) and not including (red) the MAP wavelet model. The green curve is a zero-mean unit variance Gaussian distribution.

we see from the bottom left panel of Figure 1 that the data prefer 6 wavelets and can support up to 11 wavelets, albeit with a much lower utility (utility is proportional to the Bayesian evidence). Furthermore we see from the wavelet model uncertainty (marginalized over all wavelet models) in gray in the top panel of the figure that the wavelet model only contributes to modeling the white noise burst and not any other features in the data.

Alternatively, for the log-uniform amplitude prior we see that the data prefer  $\sim 18$  wavelets but can support  $> 30$  resulting in very broad spread in the number of wavelets that are allowed by the data. Furthermore, we see from the waveform uncertainty at early times in the reconstructed waveform (gray shaded area of the top panel of Figure 2) that many wavelets are being placed far away from the white noise burst. The fact that this prior allows for many more wavelets that are effectively sitting below the white noise level can be understood by the fact that there is quite a large prior volume at low amplitudes since a uniform prior in  $\log_{10} A$  is proportional to a prior on the amplitude of  $p(A) \propto A^{-1}$ . In contrast, the uniform SNR prior is much more similar to a uniform amplitude prior. Thus, a-priori the log-uniform SNR prior prefers low amplitude wavelets whereas the uniform SNR prior prefers high amplitude wavelets.

As mentioned above we want to choose a prior on the

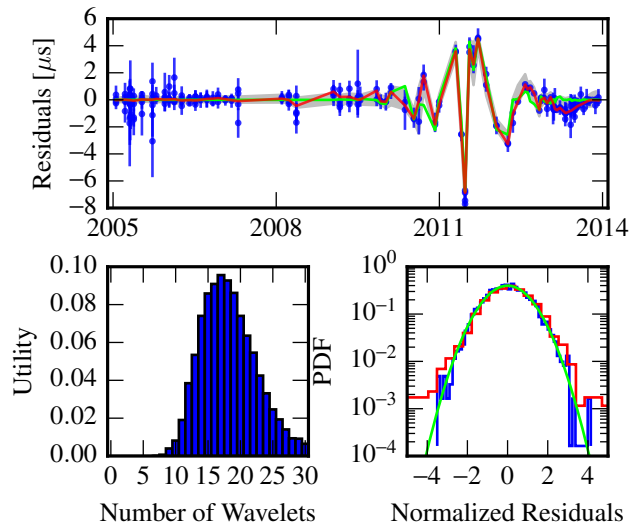


FIG. 2: Same as Figure 1 but for the log-uniform wavelet amplitude RJMCMC run.

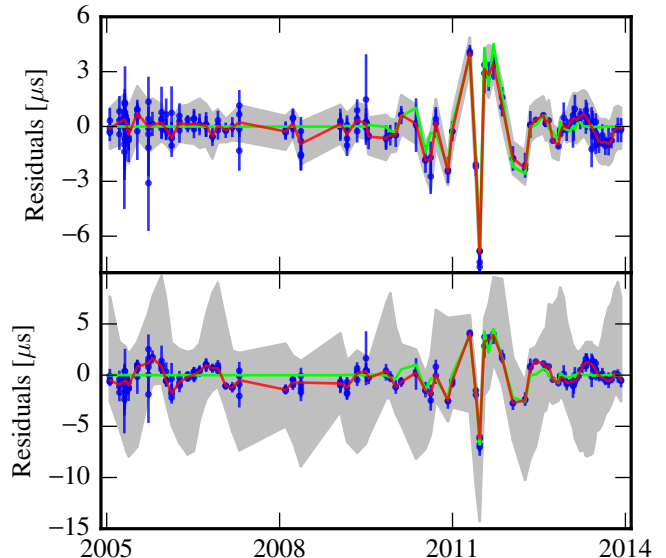


FIG. 3: Waveform reconstruction using a power-law (top) and free-spectrum (bottom) parameterization for the red noise with no additional non-stationary terms. The green is the injected waveform, the red is the MAP recovered waveform and the gray is the 90% uncertainty on the waveform.

wavelet amplitude so that we use the minimum number of wavelets to model the signal. As we have seen from this example, a uniform SNR prior performs admirably in this respect whereas a log-uniform amplitude prior fails this test.

In Figure 3, we perform the analysis without the wavelet model using standard noise models with a power-law (top panel) and free-spectrum (bottom panel) red noise parameterization. In comparison with Figure 1 we see that while the waveform is accurately recovered in

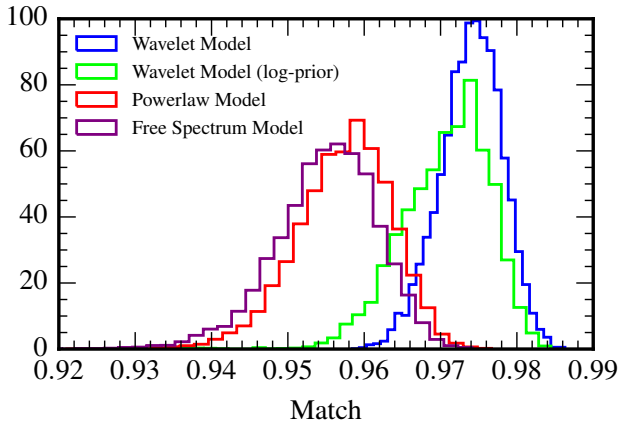


FIG. 4: Match between the injected white noise burst waveform and waveform reconstructions for the wavelet and standard noise models. We see that the wavelet models have a higher match overall, thus indicating better fitting, and the power-law and free spectral models have similar matches that are both lower than the wavelet model.

both cases, the uncertainty in this inferred waveform is much larger than in our trans-dimensional wavelet model. Most importantly we note that the uncertainty is very large in the regions where the injected waveform is zero. This is mostly due to the fact that both the power-law and free-spectral models are constrained to be a realization of a stationary gaussian process. Furthermore we point out that the large uncertainty and periodic structure seen in the uncertainty region for the free-spectrum model is due to covariances between the power spectral amplitudes and the sky location and parallax terms in the timing model. This feature is unavoidable in such an unconstrained model. **Note that the gray areas are the uncertainty regions on the red noise waveform itself and not the residuals, which are produced from the MAP red and white noise parameters. However, investigating the waveform reconstruction does not fully describe the validity of this model.**

The match is defined as

$$M = \frac{(s|\bar{s})}{\sqrt{(\bar{s}|\bar{s})(s|s)}}, \quad (17)$$

where  $s$  is the injected waveform and  $\bar{s}$  is the reconstructed waveform. Figure 4 shows the match distribution of the injected waveform with the recovered waveform for all four models. Note that both wavelet models have a higher match than the standard noise models, indicating better fitting. However, it is important to note that even though the free-spectral model is unable to decouple timing model and red noise, the combination does indeed match with the injected waveform in a nearly identical manner as the power-law model. Finally, in these simulations we have only included white

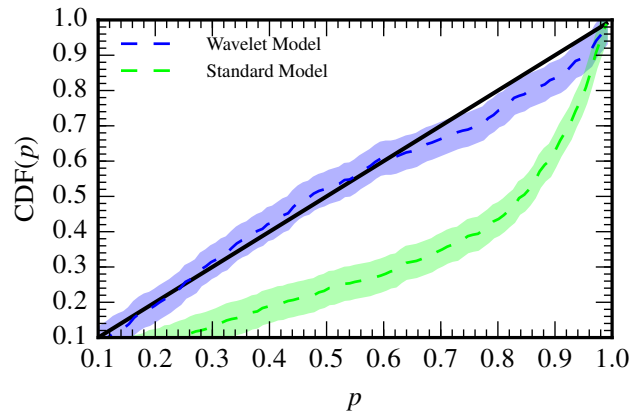
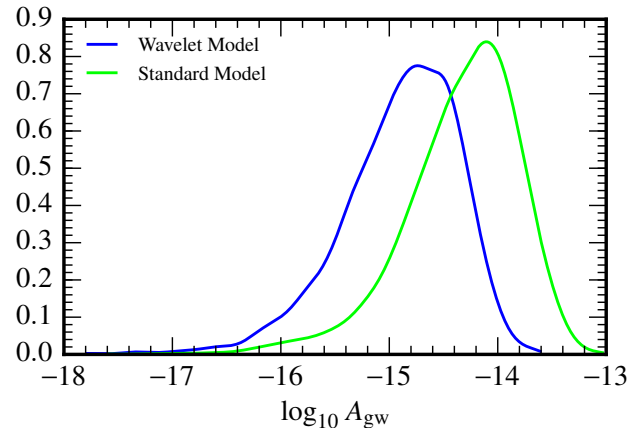


FIG. 5: Top: Marginalized posterior distribution of the dimensionless strain amplitude of a stochastic GW background using the trans-dimensional wavelet model (blue) and standard power-law red noise model (green). Bottom: probability - probability plot for the GW amplitude parameter. On the  $x$ -axis is the probability  $p$  contained in a credible interval, and on the  $y$ -axis is the fraction of true values that lie within that interval. The black diagonal line is the ideal distribution where the credible interval is perfectly consistent with the fraction of recovered injections. Again, the blue and green curves represent the wavelet and standard noise models, respectively, with the shaded area accounting for the uncertainty due to the limited number (500) of injections.

noise and the white noise burst, in such cases where there is steep red noise *and* transient behavior, it is likely that standard methods would perform much worse.

Lastly, we show that this kind of noise analysis could be crucial for placing tight constraints or making a detection of a stochastic GW background. In Figure 5 we plot the marginalized posterior distribution of the dimensionless GW strain amplitude using the wavelet model above and the standard power-law red noise model, plotted in blue and green, respectively. For this simulation we use the same data that was analyzed above. In both cases we model the red noise via a power-law and include an additional noise term with a fixed spectral index (13/3) cor-



responding to a stochastic GW background of SMBHBs. For the wavelet model we allow up to 30 wavelets. As we see from Figure 5, we can constrain the amplitude of a potential stochastic GW background significantly better when we include the additional wavelets in the noise model. Specifically, the upper 95% upper limits on the GWB amplitude for the wavelet and non-wavelet model are  $6.7 \times 10^{-15}$  and  $2.3 \times 10^{-14}$ , – a factor of 3.5 improvement. As our PTA data becomes more precise over time, this additional noise term could prove critical in either detecting or constraining the stochastic GW background.

One may be concerned that the wavelet model might absorb some of the GW signal and lead to artificially low bounds. An analysis of a set of 500 injections with an injected GW background and different white noise and white noise burst realizations shows this is not the case. The wavelet model is unbiased, while the standard model without wavelets is biased toward high values. For each injection we construct  $p\%$  credible intervals and compare this to the fraction of realizations in which the injected value lies within the  $p\%$  credible interval. If the posterior samples are an unbiased estimator of the true probability then  $p\%$  of the realizations should find the injected value within the  $p\%$  credible interval. We then perform a two-sample KS-test to test whether the results match the expected one-to-one relation. We find that the wavelet model passes the KS-test (with  $p$ -value 0.9) while the standard power-law noise model fails (with  $p$ -value  $\sim 10^{-8}$ ). These results show that the wavelet model is indeed statistically unbiased when estimating the GW background amplitude and that using a less flexible power-law noise parameterization to model transient signals results in biased GW background amplitude estimates.

We have performed one last test in which we have simulated a GWB with amplitude  $5 \times 10^{-14}$  and power spectral index  $13/3$  into a pure white noise dataset with no white noise burst. In Figure 6 we plot the marginalized 2-d posterior of GWB amplitude and spectral index for a standard model with no wavelets (green), wavelet model marginalized over number of wavelets (blue), and wavelet model using only samples from the RJMCMC that use 0 wavelets. The black lines and cross are the injected values. We see that the posteriors are nearly identical in all three cases, showing that there is no evidence of GW absorption by the wavelet model. In the bottom plot we see that the data clearly favor 0 wavelets for this injection further indicating that this model is fully robust both in the presence and absence of transient noise signals.

Finally, though we have tested the transient noise modeling using a white noise burst, this method will recover the waveform of any unmodelled achromatic (with respect to radio frequency) signal. However, in such cases where we know the expected waveform of an event, such as a pulsar glitch which manifests as a ramp function in

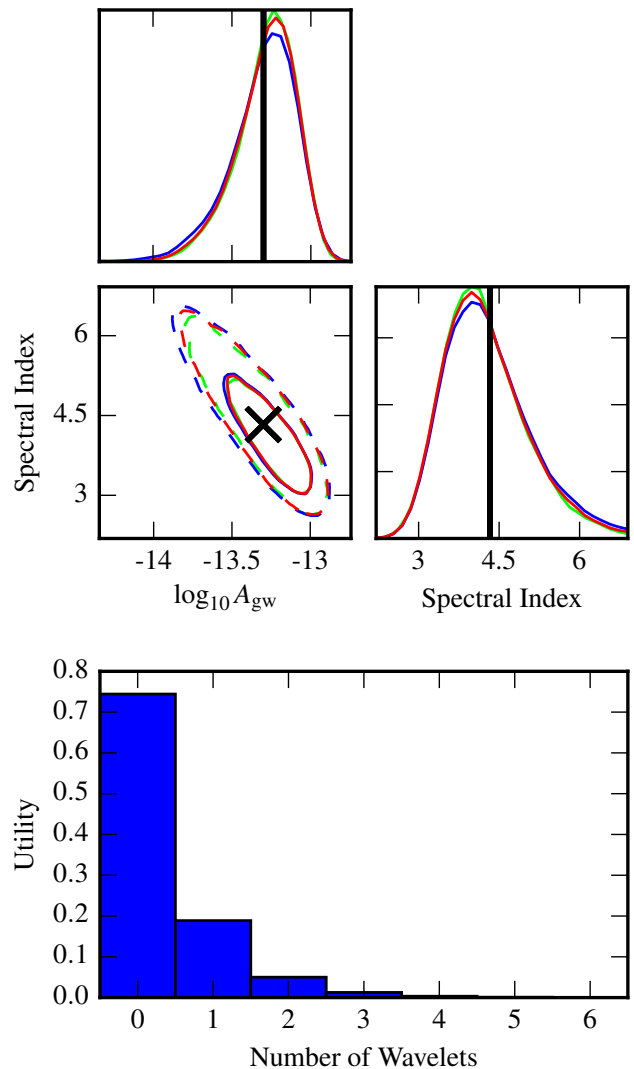


FIG. 6: Results from simulation with injected GWB and no white noise burst. Top: Marginalized 2-d posterior of GWB amplitude and spectral index for a standard model with no wavelets (green), wavelet model marginalized over number of wavelets (blue), and wavelet model using only samples from the RJMCMC that use 0 wavelets. The black lines and cross are the injected values. We see that the posteriors are nearly identical in all three cases, showing that there is no evidence of GW absorption by the wavelet model. Bottom: Utility vs. number of wavelets for this injection. The data clearly favor 0 wavelets.

the residuals, this trans-dimensional method will prove sub-optimal compared to a modeled waveform search.

## B. BayesSpecPTA

Here we will test the adaptive spectral modeling techniques of Section IV on three cases that we call the null, intermediate, and extreme cases. For each simulation we

will recover the spectrum using the adaptive technique, a standard power-law, and free spectral components.

### 1. Null Case

For the null case, we inject a realization of a gaussian process that follows a pure power-law distribution with a power spectral index of  $13/3$  as one would expect from a gravitationally wave driven isotropic stochastic background from a population of SMBHBs in circular orbits. We have dubbed this the null case since the spectrum is fully described by two parameters and this is the simplest model that the adaptive spectral modeling technique can achieve. In Figure 7 we show the results of our analysis on this null case. The middle panel shows the posterior probability density of the recovered spectrum from our adaptive method. The solid black lines are the 90% credible intervals and the median value. The dashed line is the PSD of the injected noise process. The top panel shows the utility of the various control points as a function of frequency. Here we define the utility as the ratio of the number of iterations of the RJMCMC that a given control point was active to the total number of iterations. In this case we see that none of the additional control points was active for a significant number of iterations. The bottom panel again shows the posterior probability density of the recovered spectrum using the power-law model.

The gray points and error bars are the median and the 90% confidence interval on each spectral component using the free spectral technique. From Figure 8 we can see that in this case the adaptive technique does indeed favor no additional control points (i.e. the spectrum is parameterized by two parameters). Furthermore we see from Figure 7 that the recovered spectrum is very similar for the adaptive and power-law models. The posterior is slightly broader in the adaptive case due to the fact that we allow for more than two spectral components and although the data favors only two, it is clear from Figure 8 that the data does not heavily disfavor more than two components. Furthermore, the free spectral model is less constraining in the lowest frequency component and does not constrain other frequencies significantly.

### 2. Intermediate Case

For the intermediate case we have injected a realization of a gaussian process that follows a broken power-law distribution. In this case we see in Figure 9 that the spectrum is well recovered with the adaptive technique and that the data strongly prefer an additional control point at the turnover frequency in the spectrum. Furthermore, we see from Figure 10 that, overall, the data prefers to describe the spectrum by three parameters. This is exactly the behavior that we wish to see in that the most parsimonious model to describe a broken power-law contains three parameters. Meanwhile, the bottom panels of Fig-

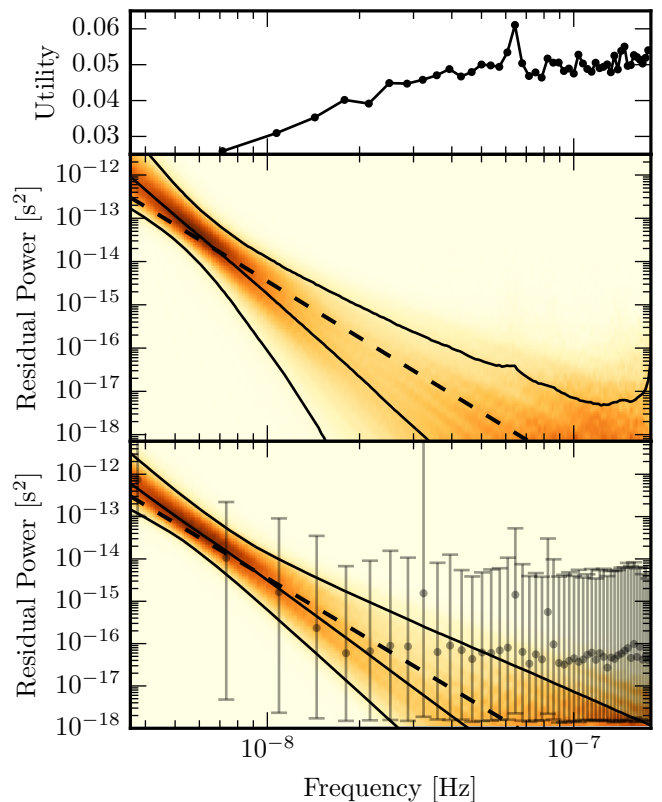


FIG. 7: Results of BayesSpecPTA and the two standard analyses (powerlaw and free spectral coefficients) on the null case. The top two panels show the recovered spectrum (middle panel) and the favored control points for the spectral interpolation (top panel), respectively. The bottom panel is the recovered spectrum from the powerlaw analysis (density) and the free spectral coefficients (error bars). In both density plots, the dashed line is the injected PSD and the solid black lines represent the median and 90% confidence region. In the bottom panel, the gray error bars represent the median and 90% confidence interval on each spectral coefficient.

ure 9 shows that the power-law noise model significantly overestimates the low frequency noise and the free spectral model loosely models the power around the turnover in the spectrum. In fact, the powerlaw model does not even recover the spectrum within its 90% credible interval, thus failing basic Bayesian consistency. This case really demonstrates the power of the adaptive technique in which it correctly identifies the most parsimonious model containing three parameters while still incorporating the uncertainty associated with more or less components, all while producing an accurate reconstruction of the power spectral density.

### 3. Extreme Case

For the so-called extreme case we have injected a realization of a gaussian process that follows a distribution with two distinct peaks. In this case we see in Figure

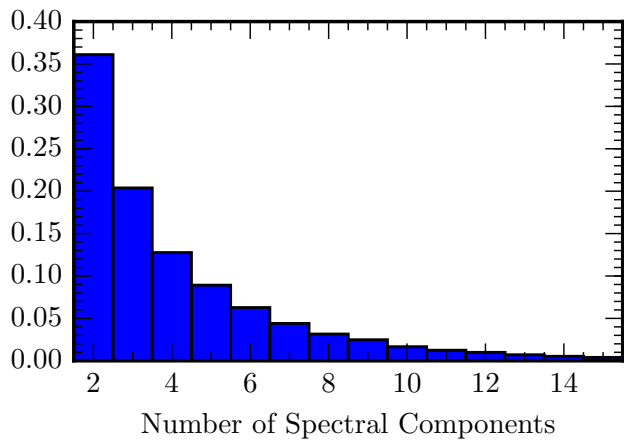


FIG. 8: Favored number of spectral components from the BayesSpecPTA analysis of the null case. In this case, where the injected spectrum is a simple powerlaw, the spectrum is best described by two parameters as one would expect.

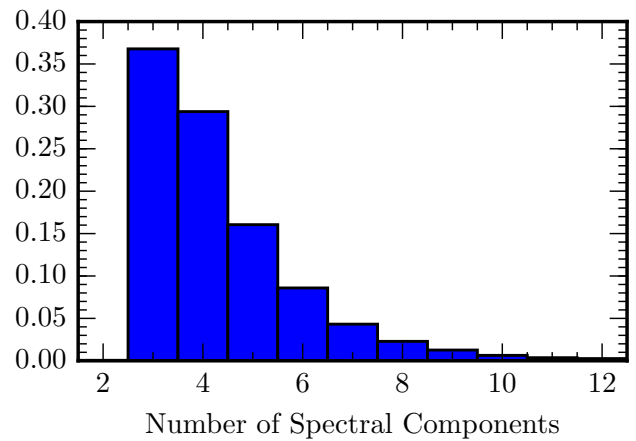


FIG. 10: Favored number of spectral components from the BayesSpecPTA analysis of the intermediate case. In this case, where the injected spectrum has a single turnover the spectrum is best described by three parameters as one would expect, and two parameters (powerlaw) is strongly disfavored while more than 3 components is allowed.

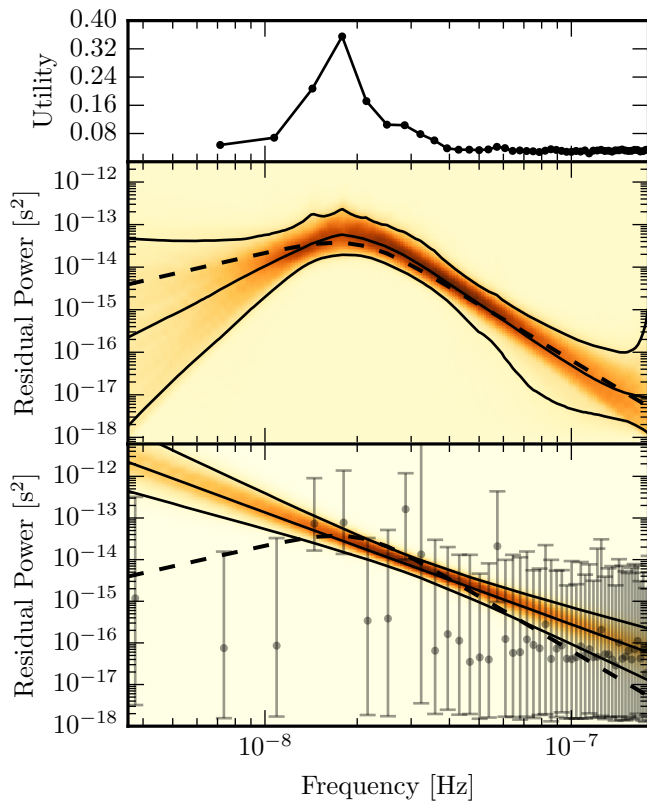


FIG. 9: Same as Fig. 7 but for the intermediate case.

11 that the spectrum is well recovered with the adaptive technique and that the high frequency peak is clearly distinguishable while the low frequency peak is not as constrained. In this case, as can be seen in both Figures 11 and 12, the model complexity is in strong competition with the goodness-of-fit in that data nearly equally pre-

fer either a shallow power-law (2 parameters) that models the power at the peaks of the PSD but also the frequencies in between, or the more complex spectrum that models both the peaks *and* valleys of the PSD. These features show the utility of this kind of analysis in that one does not have to a-priori choose a model for the spectrum and try to find the *best* model from the data but instead we allow the data to constrain the model while *marginalizing* over our uncertainty in that model. For comparison, from the bottom panel of Figure 11 we see that, because of the rigidity of the power-law model, it overestimates the power at low frequencies and underestimates the power at the high frequency peak. Furthermore, we see that even the free spectral method does not significantly constrain the PSD at either peak.

### C. Combination of Both Methods

In this section we combine both the wavelet and adaptive signal modeling in one large RJMCMC where the number of wavelets and the number of control points are both free to vary. In principle, this method should be near optimal in that we allow the data to decide the complexity of the non-stationary and stationary noise processes simultaneously; however, in practice it can be difficult to distinguish between PSDs with significant high frequency power and transient signals. Therefore, in many cases, although we are able to recover the actual waveform in the data, it is difficult to separate out the stationary and non-stationary components. Nonetheless, because neither the wavelet modeling nor adaptive spectral modeling can optimally account for all noise sources alone, we recommend this combined approach.

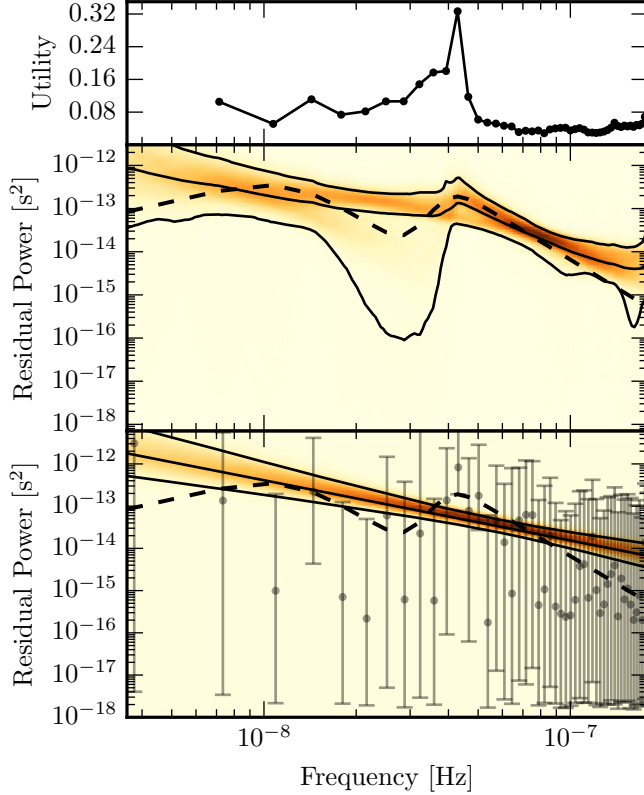


FIG. 11: Same content as Fig. 7 for the extreme case.

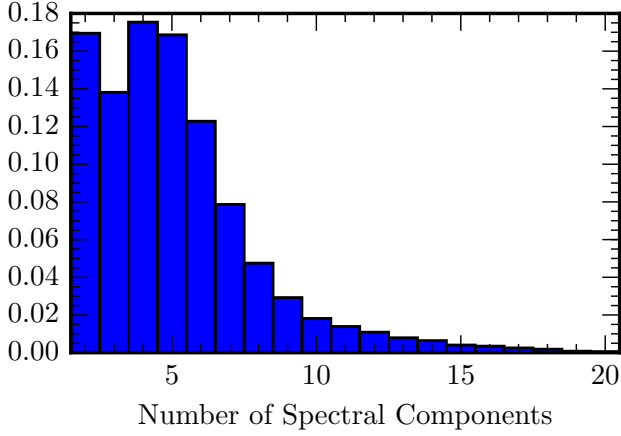


FIG. 12: Favored number of spectral components from the BayesSpecPTA analysis of the extreme case. In this case, the injected spectrum has two peaks. The data can either be described by a power law or a more complicated spectra that requires  $\geq 4$  control points.

To test this combined method we once again simulate a data set with low frequency red noise with a broken power spectrum (black dashed line in Figure 13) and a white noise burst. We have used a uniform SNR prior for the wavelet amplitudes and log-uniform prior for the amplitude of the interpolation control points. The results of

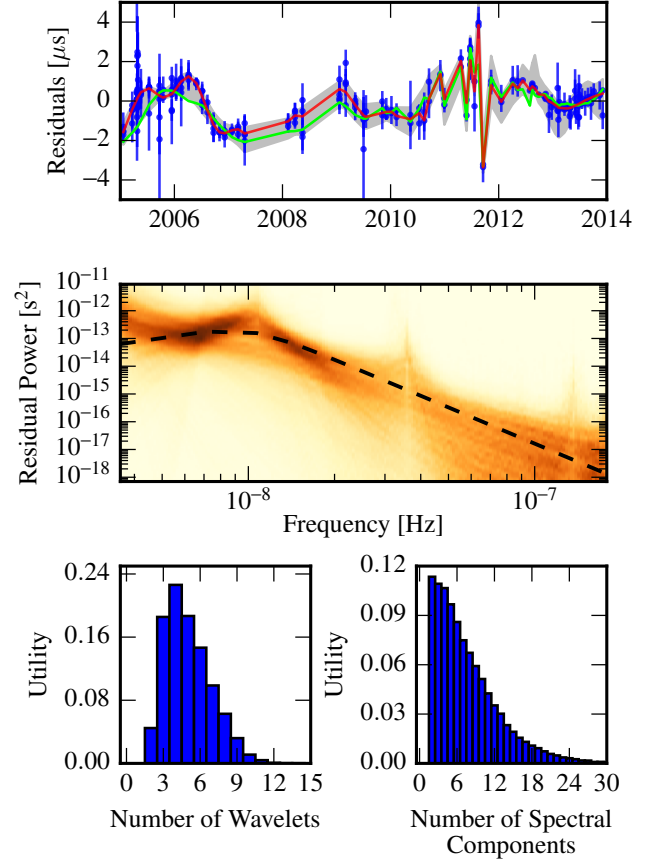


FIG. 13: Summary plots for combined run. In the top panel we plot the daily averaged residuals in blue, the MAP red noise and wavelet realization in red and the injected value in green. The gray region is the 90% confidence region of the signal. In the middle panel we plot the probability density of the recovered PSD and the injected PSD in black. The bottom left panel shows the utility for the number of wavelets and number of spectral components.

this analysis are shown in Figure 13. In the top panel we plot the daily averaged residuals (blue) along with the injected (green) and MAP (red) red noise and transient noise signal realization and the corresponding 90% credible region in gray. The middle panel shows the posterior probability of the recovered PSD along the injected PSD (dashed black line). The bottom panel shows the utility (same definition as above) of the number of wavelets and number of spectral components used in the model.

First, we see that the model recovers the injected waveform very well through a combination of  $\sim 5$  wavelets and a broken power spectrum. Although this is a different realization of the same white noise burst used above the complexity of the wavelet model is directly comparable to that of Figure 1. Furthermore, we see the spectrum is recovered fairly well but there is still a large spread in which many spectral components could be used. This is due to the covariances between the wavelet model and the spectral model and is likely unavoidable unless we

have some a-prior knowledge of the transient and/or the stationary noise spectrum.

## VII. DISCUSSION AND CONCLUSIONS

As the precision of PTAs increases, sophisticated and robust noise modeling techniques are quickly becoming one of the most important aspects in the pulsar timing process. In this paper we have presented a trans-dimensional noise modeling approach that lets the data decide the complexity of the model as opposed to traditional methods of choosing a specific model prior to analysis. Furthermore, this analysis does not seek to determine the “best” model but instead captures our inherent uncertainty in the noise model by marginalizing over a large range of models. In this first analysis we have modified the transient signal modeling and adaptive spectral techniques first developed in the LIGO context [21, 22] for use in the PTA regime. To model non-stationary transient noise events we use a sum of Morlet-Gabor wavelets and to model our stationary red noise PSD we use a set of control points for linear interpolation on a fixed frequency grid. In both cases, the number of wavelets and control points are free to vary and final inferences are made by marginalizing over the full trans-dimensional model.

Through simulations, we have shown that these methods perform better than standard methods both when

applied separately and together. Furthermore we have shown that in the presence of strong non-stationary noise, the implementation of these trans-dimensional techniques could result in significantly more constraining upper limits on a stochastic GW background.

While the application in this work has been on achromatic (with respect to radio frequency) single pulsar analysis, these methods could also be applied to dispersion measure (DM) modeling techniques and other radio-frequency dependent noise sources. Furthermore, this type of analysis could also be applied to multi-pulsar GW analysis both for noise modeling and for GW detection of unmodeled signals or as a cross check to other modeled GW searches.

## Acknowledgments

JAE acknowledges support by NASA through Einstein Fellowship grant PF4-150120. NJC was supported by NSF Physics Frontiers Center Award PFC-1430284. Portions of this research were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. This work was supported in part by National Science Foundation Grant No. PHYS-1066293 and by the hospitality of the Aspen Center for Physics where this work was initiated.

- 
- [1] R. A. Hulse and J. H. Taylor, **195**, L51 (1975).
  - [2] P. B. Demorest, T. Pennucci, S. M. Ransom, M. S. E. Roberts, and J. W. T. Hessels, *Nature (London)* **467**, 1081 (2010), 1010.5788.
  - [3] I. H. Stairs, *Living Reviews in Relativity* **6**, 5 (2003), astro-ph/0307536.
  - [4] D. R. Lorimer, *Living Reviews in Relativity* **11**, 8 (2008), 0811.0762.
  - [5] A. N. Lommen, *Reports on Progress in Physics* **78**, 124901 (2015), URL <http://stacks.iop.org/0034-4885/78/i=12/a=124901>.
  - [6] J. M. Cordes and R. M. Shannon (2010), 1010.3785.
  - [7] R. M. Shannon and J. M. Cordes, *Astrophys. J.* **725**, 1607 (2010), 1010.4794.
  - [8] R. van Haasteren and Y. Levin, **428**, 1147 (2013), 1202.5932.
  - [9] M. J. Keith, W. Coles, R. M. Shannon, G. B. Hobbs, R. N. Manchester, M. Bailes, N. D. R. Bhat, S. Burke-Spolaor, D. J. Champion, A. Chaudhary, et al., **429**, 2161 (2013), 1211.5887.
  - [10] L. Lentati, P. Alexander, M. P. Hobson, F. Feroz, R. van Haasteren, K. J. Lee, and R. M. Shannon, *MNRAS* **437**, 3004 (2014), 1310.2120.
  - [11] J. M. Cordes, R. M. Shannon, and D. R. Stinebring, *ArXiv e-prints* (2015), 1503.08491.
  - [12] M. T. Lam, J. M. Cordes, S. Chatterjee, and T. Dolch, *Astrophys. J.* **801**, 130 (2015), 1411.1764.
  - [13] M. T. Lam, J. M. Cordes, S. Chatterjee, M. L. Jones, M. A. McLaughlin, and J. W. Armstrong, *ArXiv e-prints* (2015), 1512.02203.
  - [14] Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, S. Chatterjee, J. M. Cordes, P. B. Demorest, X. Deng, T. Dolch, J. A. Ellis, et al., *Astrophys. J.* **794**, 141 (2014), 1404.1267.
  - [15] The NANOGrav Collaboration, Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. Chamberlin, S. Chatterjee, B. Christy, J. M. Cordes, N. Cornish, K. Crowter, et al., *Astrophys. J.* **813**, 65 (2015), 1505.07540.
  - [16] W. W. Zhu, I. H. Stairs, P. B. Demorest, D. J. Nice, J. A. Ellis, S. M. Ransom, Z. Arzoumanian, K. Crowter, T. Dolch, R. D. Ferdman, et al., *Astrophys. J.* **809**, 41 (2015), 1504.00662.
  - [17] L. Lentati, S. R. Taylor, C. M. F. Mingarelli, A. Sesana, S. A. Sanidas, A. Vecchio, R. N. Caballero, K. J. Lee, R. van Haasteren, S. Babak, et al., *ArXiv e-prints* (2015), 1504.03692.
  - [18] Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. Chamberlin, S. Chatterjee, B. Christy, J. Cordes, N. Cornish, P. Demorest, X. Deng, et al., *ArXiv e-prints* (2015), 1508.03024.
  - [19] S. Babak, A. Petiteau, A. Sesana, P. Brem, P. A. Rosado, S. R. Taylor, A. Lassus, J. W. T. Hessels, C. G. Bassa, M. Burgay, et al., *MNRAS* **455**, 1665 (2016), 1509.02165.
  - [20] D. J. Reardon, G. Hobbs, W. Coles, Y. Levin, M. J. Keith, M. Bailes, N. D. R. Bhat, S. Burke-Spolaor, S. Dai, M. Kerr, et al., *MNRAS* **455**, 1751 (2016), 1510.04434.
  - [21] N. J. Cornish and T. B. Littenberg, *Classical and Quan-*

- tum Gravity **32**, 135012 (2015), 1410.3835.
- [22] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **91**, 084034 (2015), 1410.3852.
  - [23] M. A. Woodbury, Memorandum report **42**, 106 (1950).
  - [24] A. Refregier, MNRAS **338**, 35 (2003), astro-ph/0105178.
  - [25] L. Lentati, P. Alexander, and M. P. Hobson, MNRAS **447**, 2159 (2015), 1412.1427.
  - [26] L. Lentati and R. M. Shannon, MNRAS **454**, 1058 (2015), 1509.07505.
  - [27] W. Voudsen, W. M. Farr, and I. Mandel, ArXiv e-prints (2015), 1501.05823.