The Astrophysical Journal, 890:108 (15pp), 2020 February 20 https://doi.org/10.3847/1538-4357/ab68db
© 2020. The American Astronomical Society. All rights reserved.

The NANOGrav 11 yr Data Set: Evolution of Gravitational-wave Background Statistics

J. S. Hazboun[1,40], J. Simon[2,3], S. R. Taylor[2,3,4], M. T. Lam[5,6,7], S. J. Vigeland[8], K. Islo[8], J. S. Key[1], Z. Arzoumanian[9], P. T. Baker[6,7,10], A. Brazier[11,12], P. R. Brook[6,7], S. Burke-Spolaor[6,7], S. Chatterjee[11], J. M. Cordes[11], N. J. Cornish[13], F. Crawford[14], K. Crowter[15], H. T. Cromartie[16], M. DeCesar[17,40], P. B. Demorest[18], T. Dolch[19], J. A. Ellis[6,7,20], R. D. Ferdman[21], E. Ferrara[22], E. Fonseca[23], N. Garver-Daniels[6,7], P. Gentile[6,7], D. Good[15], A. M. Holgado[24], E. A. Huerta[24], R. Jennings[11], G. Jones[25], M. L. Jones[6,7], A. R. Kaiser[6,7], D. L. Kaplan[8], L. Z. Kelley[26], T. J. W. Lazio[2], L. Levin[27], A. N. Lommen[28], D. R. Lorimer[6,7], J. Luo[29,30], R. S. Lynch[31], D. R. Madison[6,7], M. A. McLaughlin[6,7], S. T. McWilliams[6,7], C. M. F. Mingarelli[32], C. Ng[33], D. J. Nice[17], T. T. Pennucci[34], N. S. Pol[6,7], S. M. Ransom[16,35], P. S. Ray[36], X. Siemens[8], R. Spiewak[37], I. H. Stairs[15], D. R. Stinebring[38], K. Stovall[18], J. Swiggum[8,40], J. E. Turner[6,7,8], M. Vallisneri[2,3], R. van Haasteren[2,41], C. A. Witt[6,7], and W. W. Zhu[39]
(The NANOGrav Collaboration)

[1] University of Washington Bothell, 18115 Campus Way NE, Bothell, WA 98011, USA; hazboun@uw.edu
[2] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[3] Theoretical AstroPhysics Including Relativity (TAPIR), MC 350-17, California Institute of Technology, Pasadena, CA 91125, USA
[4] Department of Physics and Astronomy, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN 37235, USA
[5] School of Physics and Astronomy, Rochester Institute of Technology, Rochester, NY 14623, USA
[6] Department of Physics and Astronomy, West Virginia University, P.O. Box 6315, Morgantown, WV 26506, USA
[7] Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, WV 26505, USA
[8] Center for Gravitation, Cosmology and Astrophysics, Department of Physics, University of Wisconsin–Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA
[9] X-Ray Astrophysics Laboratory, NASA Goddard Space Flight Center, Code 662, Greenbelt, MD 20771, USA
[10] Department of Physics and Astronomy, Widener University, One University Place, Chester, PA 19013, USA
[11] Department of Astronomy, Cornell University, Ithaca, NY 14853, USA
[12] Cornell Center for Advanced Computing, Ithaca, NY 14853, USA
[13] Department of Physics, Montana State University, Bozeman, MT 59717, USA
[14] Department of Physics and Astronomy, Franklin & Marshall College, P.O. Box 3003, Lancaster, PA 17604, USA
[15] Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
[16] University of Virginia, Department of Astronomy, P.O. Box 400325, Charlottesville, VA 22904, USA
[17] Department of Physics, Lafayette College, Easton, PA 18042, USA
[18] National Radio Astronomy Observatory, 1003 Lopezville Rd., Socorro, NM 87801, USA
[19] Department of Physics, Hillsdale College, 33 E. College Street, Hillsdale, MI 49242, USA
[20] Infinia ML, 202 Rigsbee Avenue, Durham, NC 27701, USA
[21] Department of Physics, University of East Anglia, Norwich, UK
[22] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[23] Department of Physics, McGill University, 3600 University St., Montreal, QC H3A 2T8, Canada
[24] NCSA and Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[25] Department of Physics, Columbia University, New York, NY 10027, USA
[26] Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Northwestern University, Evanston, IL 60208, USA
[27] Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester, M13 9PL, UK
[28] Haverford College, 370 Lancaster Ave., Haverford, PA 19041, USA
[29] University of Texas at San Antonio, San Antonio, TX, USA
[30] Center for Advanced Radio Astronomy, University of Texas Rio Grande Valley, Brownsville, TX 78520, USA
[31] Green Bank Observatory, P.O. Box 2, Green Bank, WV 24944, USA
[32] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA
[33] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
[34] Hungarian Academy of Sciences MTA-ELTE "Extragalactic Astrophysics Research Group," Institute of Physics, Eötvös Loránd University, Pázmány P.s. 1/A, 1117 Budapest, Hungary
[35] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[36] Naval Research Laboratory, Washington, DC 20375, USA
[37] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia
[38] Department of Physics and Astronomy, Oberlin College, Oberlin, OH 44074, USA
[39] CAS Key Laboratory of FAST, Chinese Academy of Science, Beijing 100101, People's Republic of China
Received 2019 September 18; revised 2019 December 24; accepted 2020 January 6; published 2020 February 18

Abstract

An ensemble of inspiraling supermassive black hole binaries should produce a stochastic background of very low frequency gravitational waves. This stochastic background is predicted to be a power law, with a gravitational-wave strain spectral index of $-2/3$, and it should be detectable by a network of precisely timed millisecond pulsars, widely distributed on the sky. This paper reports a new "time slicing" analysis of the 11 yr data release from the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) using 34 millisecond pulsars. Methods to flag potential "false-positive" signatures are developed, including techniques to identify

[40] NANOGrav Physics Frontiers Center Postdoctoral Fellow.
[41] Currently employed at Microsoft Corporation.

responsible pulsars. Mitigation strategies are then presented. We demonstrate how an incorrect noise model can lead to spurious signals, and we show how independently modeling noise across 30 Fourier components, spanning NANOGrav's frequency range, effectively diagnoses and absorbs the excess power in gravitational-wave searches. This results in a nominal, and expected, progression of our gravitational-wave statistics. Additionally, we show that the first interstellar medium event in PSR J1713+0747 pollutes the common red-noise process with low spectral index noise, and we use a tailored noise model to remove these effects.

Unified Astronomy Thesaurus concepts: Millisecond pulsars (1062); Pulsar timing method (1305); Radio pulsars (1353); Red noise (1956); Gravitational waves (678); Gravitational wave detectors (676)

1. Introduction

Pulsar timing arrays (PTAs; Sazhin 1978; Detweiler 1979; Foster & Backer 1990) are poised to detect the stochastic background of gravitational waves (GWs) from a population of supermassive binary black holes (SMBBHs) within approximately the next 5 yr (Siemens et al. 2013; Rosado et al. 2015; Taylor et al. 2016; Kelley et al. 2017). There are three PTA collaborations that have been in operation for over a decade: the North American Observatory for Gravitational Waves (NANOGrav; McLaughlin 2013), the European Pulsar Timing Array (Desvignes et al. 2016), and the Parkes Pulsar Timing Array (Hobbs 2013). A number of emerging collaborations, including the Chinese Pulsar Timing Array (Lee 2016), the Indian Pulsar Timing Array (Joshi et al. 2018), and telescope-centered timing groups such as MeerTime (Bailes et al. 2018) and CHIME/Pulsar (Ng 2018), all have a component of their programs directed toward nanohertz GW detection and characterization. Together with the more established PTAs, these groups form the International Pulsar Timing Array Collaboration (Verbiest et al. 2016).

The NANOGrav Collaboration has so far released four data sets based on, respectively, 5 yr of precision pulsar timing observations (Demorest et al. 2013, hereafter NG5a), 9 yr of observations (Arzoumanian et al. 2015, hereafter NG9a), 11 yr of observations (Arzoumanian et al. 2018b, hereafter NG11a), and 12.5 yr of observations (M. Alam et al. 2020, in preparation).[42] The present analysis was carried out on NG11a, since the newest data release has only recently been available.

The dominant signal expected at nanohertz GW frequencies (where the regime of sensitivity is set by the cadence Δt and total baseline $T_{\rm total}$ of the pulsar time-series sampling: $1/T_{\rm total} < f < 1/2\Delta t$) is the stochastic background of GWs from an SMBBH population (see, e.g., Burke-Spolaor et al. 2019). There are several models in the literature that predict the amplitude and spectral shape of this GW background (GWB; e.g., Sesana 2013; McWilliams et al. 2014; Simon & Burke-Spolaor 2016; Kelley et al. 2017; Chen et al. 2019). These models employ a range of galaxy surveys, galaxy evolution scenarios, and simulations to identify the most likely demographics of SMBBHs detectable by PTAs. Recent results from NANOGrav (Arzoumanian et al. 2018a, hereafter NG11b) and other PTAs (Lentati et al. 2015; Shannon et al. 2015) have reported constraints on the GWB characteristic strain amplitude that intersect astrophysically interesting regions of SMBBH parameter space. Using techniques developed in Sampson et al. (2015), Simon & Burke-Spolaor (2016), and Taylor et al. (2017b), the 11 yr data set was used to constrain the relationship of supermassive black hole masses to that of their host galaxies, as well as galactic center environments that may influence the final parsec of binary dynamical evolution.

Most searches for the GWB rightly focus on the most recent data set, first searching for and then, in the absence of a signal, setting upper limits (ULs) on the GWB. These results are often juxtaposed to earlier work from shorter data sets to illustrate the gains in sensitivity of these galactic-scale GW detectors. Here we analyze the past evolution of our statistics by slicing the NG11a data set in time and performing the analyses from NG11b on each slice. This allows us to characterize the growth of NANOGrav's GW sensitivity as a function of time, as well as diagnose previously unmodeled noise processes. With regard to the latter, in this article we discuss how a noise transient produced a high-significance false positive during the time span between the release of NG5a and NG9a. Understanding this spurious signal and tracking down the pulsars from which it originates will be the subject of most of this paper.

The paper is organized as follows. In Section 2 we discuss our methods for obtaining the pulsar timing data used in this study, and in Section 3 we review our data analysis methods, including a detailed introduction to our noise models. In Section 4 we discuss the motivation for understanding the evolution of our GW statistics and introduce theoretical models for the evolution of that signal. In Section 5 we present the results of the initial time-slice analysis, including anomalous evidence for GWs in out data set. We then turn, in Section 6, to identifying which pulsars are responsible for this behavior and elucidate the various data analysis methods, noise models, and other mitigation strategies used to understand and remove it. In Section 7 we connect the shallow spectral index of the common process in early slice analyses to the first interstellar medium (ISM) event in PSR J1713+0747 and show how models extant in the literature can mitigate this behavior. And finally, in Section 8 we conclude with a summary of the issues encountered in this analysis and possible paths forward for future PTA noise mitigation.

2. The 11 yr Data Set

The NANOGrav 11 yr data set contains observations of 45 pulsars made between 2004 and 2015. Details of the observations and pulsars can be found in NG11a. We briefly describe the data set here.

We made observations using two radio telescopes: the 100 m Robert C. Byrd Green Bank Telescope (GBT) of the Green Bank Observatory in Green Bank, West Virginia, and the 305 m William E. Gordon Telescope (Arecibo) of Arecibo Observatory in Arecibo, Puerto Rico. Since Arecibo is more sensitive than the GBT, all pulsars that can be observed from Arecibo ($0° < \delta < 39°$) were observed with it, while those outside Arecibo's decl. range were observed with the GBT. Two pulsars were observed with both telescopes: PSR J1713

[42] The release is available at https://data.nanograv.org.

+0747 and PSR B1937+21. We observed most pulsars once a month. In addition, we started a high-cadence observing campaign in 2013, in which we made weekly observations of two pulsars with the GBT (PSR J1713+0747 and PSR J1909 −3744) and five pulsars with Arecibo (PSR J0030+0451, PSR J1640+2224, PSR J1713+0747, PSR J2043+1711, and PSR J2317+1439).

At the GBT, the monthly observations used the 820 MHz and 1.4 GHz receivers, while weekly observations used only the 1.4 GHz receiver. At Arecibo, pulsars were observed with two of four possible receivers (327 MHz, 430 MHz, 1.4 GHz, and 2.3 GHz), though always including the 1.4 GHz receiver. Back-end instrumentation was upgraded about midway through our project from the ASP and GASP systems, which had bandwidths of 64 MHz, to the wideband systems PUPPI and GUPPI, processing up to 800 MHz for certain receivers (DuPlain et al. 2008).

For each pulsar, the observed times of arrival (TOAs) were fit to a timing model that described the pulsar's spin period and spin period derivative, sky location, proper motion, and distance. To this model were added a number of parameters that describe the radio-frequency-dependent behavior of the pulse arrival times. Additionally, for those pulsars in binaries the timing model also includes five Keplerian parameters that described the binary orbit, as well as additional post-Keplerian parameters that described relativistic binary effects if they statistically improved the timing fit.

3. Data Analysis Methods

The analysis techniques in this work largely follow the stochastic signal procedures in NG11b. The JPL solar system ephemeris DE436 (Folkner & Park 2016) was used along with the TT(BIPM2016) timescale. Our model likelihood is based on pulsar timing residuals, constructed for each pulsar $\delta\boldsymbol{t}$ as

$$\delta\boldsymbol{t} = \boldsymbol{T}\boldsymbol{b} + \boldsymbol{n} + \boldsymbol{s}. \tag{1}$$

$\boldsymbol{T}\boldsymbol{b}$ describes noise contributions modeled with Gaussian processes, including uncertainties in the pulsar timing model and low-frequency time-correlated (red) noise, $\boldsymbol{t}$ describes white noise (WN), and $\boldsymbol{s}$ describes residuals induced by a GWB, also modeled with a Gaussian process.

The WN is modeled using the rms template-fitting errors for the TOAs. These are inflated using additional pieces, one added in quadrature (EQUAD), and a multiplicative factor (EFAC),

$$\sigma^2_{\rm total} = {\rm EFAC}^2\sigma^2_{\rm TOA} + {\rm EQUAD}^2. \tag{2}$$

In practice, we build the WN correlation matrix by adding these diagonal contributions to the off-diagonal pieces, which model the correlated WN between TOAs observed in different subbands during the same observation (ECORR; Arzoumanian et al. 2014).

The standard likelihood for GW analysis with PTAs is well documented in the literature (Demorest et al. 2013; Lentati et al. 2013, 2014, 2016, 2015; van Haasteren & Levin 2013; Shannon et al. 2015; van Haasteren & Vallisneri 2015; Arzoumanian et al. 2016; NG11b; Taylor et al. 2017a). Here we focus on a more detailed introduction to the types of time-correlated (red) noise models we use for the analyses.

3.1. Red-noise Models and the GWB

The precision TOAs of radio pulses from millisecond pulsars have been used to measure a myriad of astrophysical interactions. Perhaps most famous is the observation of a negative binary period derivative accurately explained by the emission of GWs in the context of general relativity (Taylor & Weisberg 1982). Pulsar timing measurements are also responsible for the first detection of an extrasolar planet (Wolszczan & Frail 1992) and are used to monitor the content and movement of the galactic ionized ISM (e.g., Keith et al. 2013; Jones et al. 2017). In fact, observations from many pulsars have been put together to map the ISM content of the galaxy (Cordes & Lazio 2002; Yao et al. 2017). Lensing events from the ISM can also be monitored using pulsars, and a recent reoccurrence of an apparent lensing event in PSR J1713+0747 has been studied extensively in Lam et al. (2018b).

From the perspective of a search for GWs in pulsar timing data, these astrophysical interactions are considered sources of noise, i.e., they must be removed or mitigated in order to detect a GW. The deterministic processes are modeled by a pulsar timing model; however, in order to account for the stochastic astrophysical signals, various types of models are used in our GW search analyses (Lentati et al. 2016; Lam et al. 2018b; Madison et al. 2019).

Since the GWB manifests as a low-frequency, time-correlated stochastic process (a "red" spectrum), it is especially important to model astrophysical noise sources that leave a similar signature in our timing residuals. In our analyses both the GWB and the red noise intrinsic to a pulsar's line of sight are built with the same types of models. Most commonly they are built using a normal-kernel Gaussian process in a Fourier basis with a power-law prior (Williams & Rasmussen 2006; van Haasteren & Vallisneri 2014; Lentati et al. 2016),

$$P = \frac{A^2_{\rm GWB}}{12\pi^2}\left(\frac{f}{f_{\rm yr}}\right)^{-\gamma} {\rm yr}^3, \tag{3}$$

as a power spectral density. Here we work in terms of the timing residual spectral index, rather than GW strain, for convenience. The spectral index in strain (α) is related to γ by $\gamma = 3 - 2\alpha$. The spectral index parameter γ prior is restricted from 0 to 7, meaning that the model must be either "red" (higher power at lower frequencies) or "flat"("white"), i.e., $\gamma = 0$. The signal spectrum is then built using a Fourier basis from N frequencies (30 frequencies in most of the analyses presented here). Usually these frequencies are linearly spaced from $1/T$ to $30/T$, T being the time span of the full data set, in this case 11.4 yr. The prior used in the Gaussian process is an *Ansatz* for the type of time-correlated process that one expects to find in the residuals and describes the power spectral density of that stochastic process modeled in the frequency domain. The same frequency domain describes the GWB and non-GW red-noise sources and spans the nanohertz regime.

Using a power law is the simplest model; however, a few more complex models have been used in the literature, including a turnover model, a free spectral model, and trained Gaussian process models (Lentati et al. 2013; Taylor et al. 2013). The free spectral model is the most generic model for a time-correlated stochastic process. This allows for a different coefficient for the Fourier basis at each frequency and is not restricted by any model for the power spectral density. While

this model is very flexible, it incurs a large Occam penalty since it involves a large parameter volume. As in Arzoumanian et al. (2016) and NG11b, the models used to search for the GWB and mitigate noise in individual pulsar data sets are not dependent on the radio frequency of the TOAs; hence, we refer to these as "achromatic" red-noise models.[43]

The flagship Bayesian analysis for a PTA GW search includes the Hellings–Downs (HD) spatial correlations (Hellings & Downs 1983). In practice, these analyses are often the final analysis completed since the nondiagonal correlation matrix inversion is computationally expensive in the Bayesian framework. The HD correlation *Bayesian search* takes advantage of the spatial correlations between pulsars and the time correlations due to the GWB, modeled as an achromatic red-noise Gaussian process. In the weak signal regime, the autocorrelations within pulsars are a reasonable first estimate for a correlated stochastic background and have the same spectral content. In much of this manuscript we discuss the latter type of search for a common red-noise process, since we will need to run numerous iterations of search types over the 18 time slices we have made.

As mentioned above, the models used for the GWB and other time-correlated processes particular to the pulsar lines of sight are very similar. Historically, the *usual* model for red noise intrinsic to a pulsar, as well as its line of sight, is modeled with a power law with varying amplitude and spectral index (γ). In the most common GW analyses, the common noise process caused by the GWB is also modeled as a power law; however, the spectral index is set to that expected for a GWB, $\gamma = 13/3$. However, as we will see, it is also informative to allow the spectral index to vary when searching for the GWB. Effectively, these models are identical, but for each pulsar the model for intrinsic red noise has its own set of parameters, and no spatial correlations between other pulsars are considered. Unlike the power-law model, the free spectral model allows one to analyze noise independently at multiple frequencies, and as we will see, this can help to disentangle degeneracy between the noise process unique to the pulsar and the GWB. Additionally, the other functional forms of the spectral models mentioned above can be used for both pulsar red noise and the GWB.

"Chromatic" (radio-frequency-dependent) versions of the noise models, most often modeling dispersion with a $1/\nu^2$-dependence on radio frequency, can be found in the literature (Lee et al. 2014; Caballero et al. 2016; Lentati et al. 2016; Reardon et al. 2016) and were recently used in Lam et al. (2018b) regarding the previously mentioned ISM events in the timing of PSR J1713+0747. The standard NANOGrav analysis has not included these types of noise models, instead using a piecewise binning of dispersion measure (DM; the integrated line-of-sight electron density causing the $1/\nu^2$ dependence in the arrival times) fluctuations, called DMX, implemented as part of the timing model using the pulsar timing software `TEMPO/TEMPO2` (NG11a). This method works well at describing broadband DM fluctuations that have a timescale longer than individual DMX bins ($\leqslant$1 week; Jones et al. 2017). There are a significant number of possible chromatic effects (e.g., Cordes & Shannon 2010; Lam et al. 2018a), primarily due to radio propagation through the ISM, which need to be modeled appropriately (Shannon & Cordes 2017). The total chromatic noise assuming misestimation of DM was performed by Lam et al. (2017) on NG9a.

[43] This nomenclature is sometimes confusing, as there are two frequency domains discussed in pulsar timing, the frequencies of the GWB and red noise and the *radio* frequencies of the pulsar observations.

3.2. Slicing the Data Set

Here we use the methods of NG11b to analyze various slices of the NANOGrav 11 yr data set presented in NG11a. The data set was partitioned by setting Modified Julian Date (MJD) cutoffs in 6-month increments after an initial 3 yr span. Three years is the nominal length of individual pulsar data sets used in NG11b, and hence it is adopted here as the minimal time span of data needed to do a worthwhile analysis. The slices were cumulative, adding 6 months of data at a time. In order to understand the noise evolution in the pulsars, we performed single-pulsar noise analyses at every slice, where all of the WN and red-noise parameters are allowed to vary in a Bayesian analysis. This follows from the general philosophy throughout this investigation—use the information known at the time of each slice to do the analysis. The WN maximum likelihood values from these analyses were then used to set the WN parameters for the full PTA analyses, analogous to NG11b.

The Bayesian analysis was done using the `enterprise` software suite (Ellis et al. 2017) and the Markov Chain Monte Carlo (MCMC) sampling software `PTMCMCSampler` (Ellis & van Haasteren 2017). Detection statistics were acquired by using log-uniform priors on the red-noise amplitudes for both the individual pulsar red noise and the common red-noise process. ULs were acquired by running analyses using linear exponential priors (meant to emulate a uniform prior but sampled in log space) for the GWB amplitude and the individual red-noise amplitudes. In the detection analyses log-uniform priors were instead used for $A_{\rm GWB}$ and $A_{\rm RN}$.

A frequentist analysis was also undertaken using the same software as above and the optimal statistic submodule in the PTA model software package `enterprise_extensions` (Taylor et al. 2018). A noise-marginalized analysis (Vigeland et al. 2018) was done at each slice using the MCMC chains from the Bayesian runs to sample over the red-noise parameters. The maximum likelihood values were then used to calculate the noise-maximized values. In both cases the optimal statistic and signal-to-noise ratio (S/N) were calculated.

In most cases the spectral index for the common red-noise process was set to $\gamma = 13/3$, the theoretical spectral index (in terms of timing residuals) for a stochastic GWB originating from circular binary inspirals, where the loss of energy in the binary is driven by the radiation of GWs (Phinney 2001). In addition, an analysis was done where the spectral index of the common process was also allowed to vary.

4. Evolution of GWB Statistics

The first signal detected by PTAs is expected to be the stochastic sum of SMBBHs from the cosmological neighborhood (Rosado et al. 2015) and should grow very steeply as our data sets become sensitive farther into the nanohertz regime. Unlike the first detections of GWs from compact binary coalescences (Abbott et al. 2016, 2017), a detection of the GWB will not appear as a single event, but rather as a steady growth in significance over the course of a number of data releases. The evolution of detection statistics has been studied in the literature extensively (Siemens et al. 2013; Vigeland &

Siemens 2016), including theoretical studies of the scaling of the frequentist optimal statistic and numerical simulations using realistic data to predict when PTAs will reach specified sensitivities. Work of this kind is important for understanding the context of current data releases and the near-future ability to characterize nanohertz GW astrophysics. These types of studies also have obvious applications to the strategic planning of future PTA facilities.

4.1. Theory

The scaling laws presented in Siemens et al. (2013) provide a straightforward framework for the comparison of NANOGrav's evolving sensitivity to the GWB. In the simple case of a PTA where all pulsars only have (identical) WN, the expectation value of the S/N, $\rho = A^2_{\rm GWB}/\sigma_0$, where σ_0 is the standard deviation of $A^2_{\rm GWB}$, is shown to evolve in the weak signal regime as

$$\langle \rho_{\rm weak} \rangle = \left(\sum_I \sum_{J>I} \chi^2_{IJ} \right)^{\frac{1}{2}} A^2_{\rm GWB} \frac{bcT^\gamma}{\sigma^2 \sqrt{4\gamma - 2}}, \tag{4}$$

where χ_{IJ} is the HD spatial correlation between pulsars I and J, c is the cadence of observations, σ is the measurement error of TOAs, γ is the spectral index of the power-law background, T is the total time of observations, $A_{\rm GWB}$ is the amplitude of the GWB at $1/\mathrm{yr}$ ($f_{\rm yr}$), and b contains the frequency dependence of the GWB signal,

$$b = \frac{1}{24\pi^2} \left(\frac{1}{f_{\rm ref}} \right)^{\gamma+3}. \tag{5}$$

Similarly, in the intermediate signal regime the S/N scales as

$$\langle \rho_{\rm int} \rangle = \left(\sum_I \sum_{J>I} \chi^2_{IJ} \right)^{\frac{1}{2}} \left[2\theta \left(A^2_{\rm GWB} \frac{bc}{2\sigma^2} \right)^{\frac{1}{\gamma}} T \right]^{\frac{1}{2}}, \tag{6}$$

where θ is a function of the spectral index that includes the Γ function,

$$\theta(\gamma) = \frac{\gamma - 1}{\gamma} \Gamma(1 - \gamma^{-1}) \Gamma(1 + \gamma^{-1}). \tag{7}$$

The S/N can be related to the Bayes factor, $\mathcal{B}_{10}$, from a GWB model versus noise-only model comparison using the Laplace approximation (MacKay 2002; Romano & Cornish 2017),

$$2 \ln \mathcal{B}_{10} \approx \rho^2 + 2 \ln \left(\frac{\Delta V_1 / V_1}{\Delta V_0 / V_0} \right), \tag{8}$$

where $\Delta V_{\mathcal{M}}$ is the characteristic spread of the likelihood around the maximum and $V_{\mathcal{M}}$ is the total parameter space volume of the model. The second term on the right-hand side is negative and imposes an Occam penalty, favoring models with fewer parameters. While this expression is simple, in practice such a calculation requires detailed knowledge about the likelihood function for both the signal model and noise-only model. Current Bayesian PTA analyses use a nested model approach and a Savage–Dickey approximation to the Bayes factor, which does not furnish the noise-only likelihood function. Nonetheless, it is obvious from Equations (4) and (6) that one expects a monotonically increasing Bayes factor as the observation time for a PTA increases.

One can relate the aforementioned scaling laws to a UL by using the complimentary error function,

$$A^2_{\rm UL} = \hat{A}^2_{\rm GWB} + \sqrt{2}\, \sigma_0\, \mathrm{erfc}^{-1}[2(1 - \epsilon)], \tag{9}$$

where ϵ is the significance threshold for the limit, e.g., 0.95 for a 95% UL. The expectation value yields

$$\langle A^2_{\rm UL} \rangle = A^2_{\rm GWB} + \sqrt{2}\, \sigma_0\, \mathrm{erfc}^{-1}[2(1 - \epsilon)] \tag{10}$$

$$= A^2_{\rm GWB} \left(1 + \frac{\sqrt{2}\ \mathrm{erfc}^{-1}[2(1 - \epsilon)]}{\langle \rho \rangle} \right). \tag{11}$$

Defining $\varepsilon \equiv \sqrt{2}\ \mathrm{erfc}^{-1}[2(1 - \epsilon)]$, and given the time dependence of the S/N in Equations (4) and (6), the ULs for the optimal statistic should evolve as

$$\langle A^2_{\rm UL,weak} \rangle = A^2_{\rm GWB} + \frac{\varepsilon}{A^2_{\rm GWB}\, \delta_{\rm weak}} T^{-\frac{13}{3}}, \tag{12a}$$

$$\langle A^2_{\rm UL,int} \rangle = A^2_{\rm GWB} + \frac{\varepsilon}{A^2_{\rm GWB}\, \delta_{\rm int}} T^{-\frac{1}{2}}, \tag{12b}$$

where $\delta_{\rm weak}$ and $\delta_{\rm int}$ are shorthand for the coefficients of time, T, in Equations (4) and (6), except for $A_{\rm GWB}$.

Since these relationships are based on a frequentist statistic, it is prudent to compare them to the ULs obtained from a Bayesian analysis on simulated data sets, as it is well known that frequentist and Bayesian ULs can have different interpretations (see, e.g., Rover et al. 2011).

4.2. Simulations

We simulated NANOGrav-like data sets following Taylor et al. (2017b) using the Python wrapper for TEMPO2 (Hobbs et al. 2006) and PTA simulation package `libstempo` (Vallisneri 2015).[44] We ran a UL analysis on each simulated data set with an injected GWB of known amplitude. The simulated data sets are based on the noise properties and epochs of observation for the 11 yr data set. A UL analysis was run for 200 different realizations of a GWB at $A_{\rm GWB}$ of 1×10^{-16}, 1×10^{-15}, and 3×10^{-15}.

In Figure 1 the results are summarized for the simulations with an injection of $A_{\rm GWB} = 1 \times 10^{-15}$. The mean of the ULs and the 90% confidence interval are shown, along with the level of the injection and a fit to the theoretical evolution for the UL in the weak regime, given by Equation 12(a). The curve is fit to the mean values greater than 5 yr into the data set by varying the $\delta_{\rm weak}$ parameter.

Such a close fit from 5 to 11 yr is *extraordinary* given that many of the other parameters in this relationship are changing with time, e.g., the average cadence and average TOA error. In part, it is these large changes in parameters at the beginning of the data set that are responsible for the poor fit to the theoretical prediction at early times. This fit for $\delta_{\rm weak}$ is then used in Figure 2 to compare the theoretical evolution of Equation 12(a) to the evolution of the UL with larger and smaller injections of the GWB. With the same value for $\delta_{\rm weak}$, only changing the injection strength, $A_{\rm GWB}$, accordingly, the Bayesian analyses of these simulations follow the theoretical predictions at late times.

[44] https://github.com/vallis/libstempo

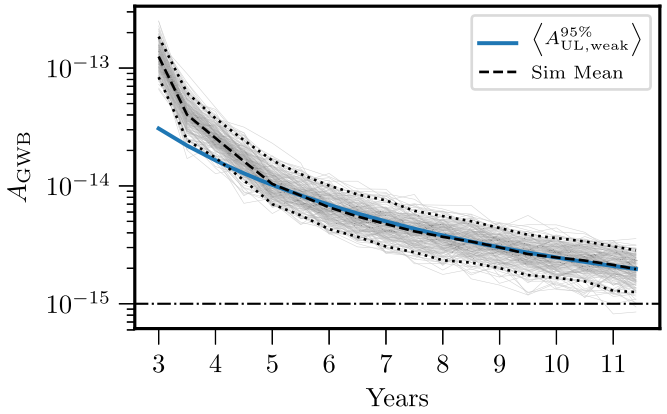


Figure 1. The 95% ULs over time for a set of NANOGrav-like data sets with an injection of $A_{\rm GWB} = 1 \times 10^{-15}$. The mean UL and 90% confidence intervals are shown for the 200 injections of a GWB. The blue curve is the predicted relationship between the UL, true value for the amplitude of the background, and total time of observation. It has been fit to the mean values of the ULs in the range of 5–11.4 yr by varying $\delta_{\rm weak}$.

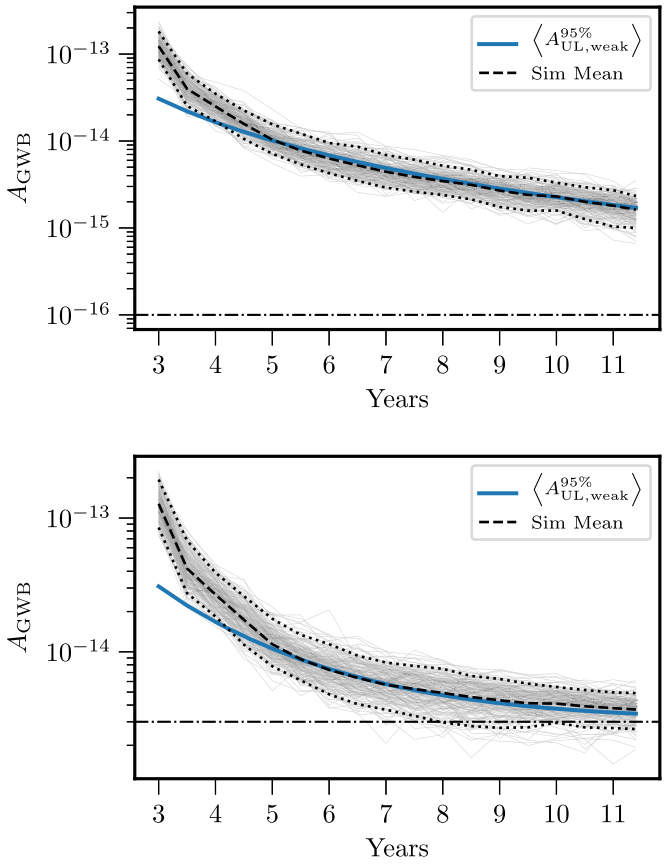


Figure 2. The 95% ULs over time for a set of NANOGrav-like data sets with injections of $A_{\rm GWB} = 1 \times 10^{-16}$ and $A_{\rm GWB} = 3 \times 10^{-15}$ are shown. The mean UL and 90% confidence intervals are shown for the 200 injections of a GWB. The blue curve is the predicted relationship between the UL, true value for the amplitude of the background, and total time of observation. Here the theoretical curve has not been fit to the data. Rather we use the value for $\delta_{\rm weak}$ used in Figure 1 and change the value of $A_{\rm GWB}$ according to Equation 12(a).

Note that since the published UL in NG11b is $A_{\rm GWB} = 1.45 \times 10^{-15}$, the sensitivity of this data set is such that a GWB with $A_{\rm GWB} = 3 \times 10^{-15}$ would be counted as a detection in the full data set. Still one can calculate a 95% UL in this case, and these simulations act as an exercise to demonstrate that the UL scalings derived in Section 4.1 hold robustly in this higher-amplitude case. In addition to the work cited on detection scaling laws earlier in the section, forthcoming work looks at injections into the real data as a way of characterizing the evolution of sensitivity in our data sets.

Armed with a general understanding of the expected evolution of GWB analysis statistics, we move on to the sliced analysis of the NANOGrav 11 yr data set.

5. Standard Analysis Results

Here we report the results of both our standard detection and UL analyses of the NANOGrav 11 yr data set for a GWB across the sliced data set.

The expectations laid out in Section 4 are that in this type of analysis, as more data are added (i.e., we have more information about the system we are studying), the posteriors for $A_{\rm GWB}$ should get narrower. In broad strokes this means that the Savage–Dickey Bayes factor approximation would start near 1 and begin to increase as more data are added. The S/N should also increase according to the evolution described in Siemens et al. (2013), while the UL should steadily decrease with more data until the data become sufficiently informative to run up against the actual signal amplitude, as shown in Section 4. In both of our Bayesian analyses, the evolution of these statistics does not conform to these reasonable predictions.

In the case of a varied spectral index analysis we expect similar behavior, but as a detection becomes imminent, the significance of a steep spectral index common process will increase. In the case of the varied spectral index analysis, rather than seeing a steep spectral index near $\gamma = 13/3$, a shallow $\gamma \approx 0$ process appears early in the observation period.

5.1. Fixed Spectral Index Analyses

The Bayes factor in the bottom panel of Figure 3 remains between 0.5 and 1.5 until 7.0 yr into the data set, and then it increases dramatically for a few slices, before decreasing again. As can be seen in the top panel of Figure 3, the UL decreases monotonically until 6.5 yr into the data set (slice ending about MJD 55,590 (2011.08)) and sharply increases over the next year until the 7.5 yr slice (slice ending about MJD 55,956 (2012.08)), before beginning to decrease again. We refer to this period of time as the "kink" for brevity.

The frequentist statistics are mixed. While the ULs calculated from the optimal statistic noise-marginalized posteriors are a bit lower during the "kink," the same trend (particularly the increase) in the UL can be seen centered around the 7.5 yr slice. The S/Ns are, however, drastically different from the Bayes factor trends and do not show the dramatic increase in this era. Since these calculations involve spatial correlations, they are often used as the quickest estimate of detection capabilities of a given data set. The fact that they are so different in this era from the Bayesian results is troubling and is the impetus for most of the remainder of this paper.

A full sliced analysis using HD spatial correlations was undertaken on the data set. These analyses hint at an even stronger detection of a GWB during the "kink" era. The posteriors for this analysis indicate a GWB detection too strong to be estimated using the Savage–Dickey Bayes factor approximation. Since these analyses take up to 10 times longer

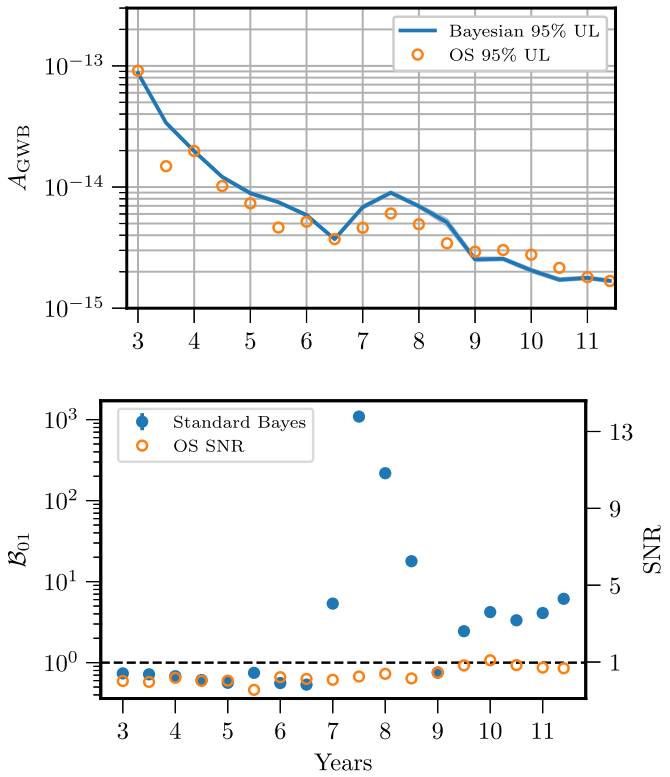


Figure 3. Results of a standard PTA GW data analysis. In both plots the blue solid line/circles show the Bayesian results at each slice, while the open orange circles show the frequentist results. Note the large rise in the UL and the Bayes factor for the Bayesian analysis starting in the 6.5 yr slice and peaking in the 7.5 yr slice. The UL calculated from the noise-marginalized optimal statistic has a similar trend to the Bayesian UL, but with lower values. The S/N calculated from the optimal statistic shows no sign of a detection in the same era.

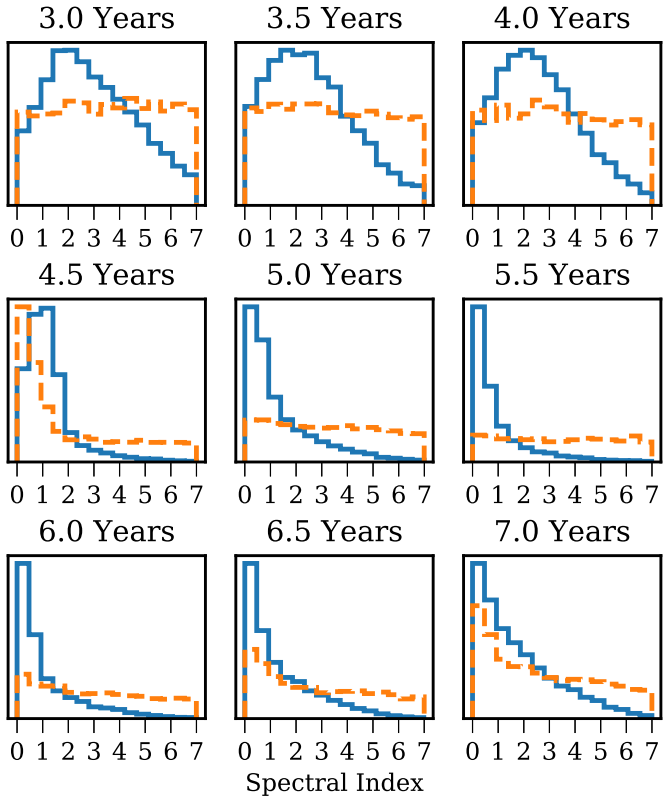


Figure 4. Comparison of posteriors for the spectral index of a power law red-noise process across the first six slices. The solid blue line is the posterior on the spectral index for a common process between all pulsars. The dashed orange line shows the posterior for the spectral index in the single-pulsar red-noise model for PSR J1713+0747. Note the appearance of a strong low spectral index in both posteriors for the 4.5 yr slice.

to run, we restricted our follow-up analyses to searching for an uncorrelated common red-noise process among all pulsars.

5.2. Varied Spectral Index

One other analysis that needs to be summarized is that with a varying spectral index for the common red-noise process model for the GWB. This analysis shows a drastic change in the posteriors of the parameters, but this change is *not* contemporaneous with the "kink" in the analyses, where the spectral index is set to $\gamma = 13/3$. Here we see that the spectral index describing the common red noise between the pulsars changes dramatically in the 4.5 yr slice ending at MJD 54,860 (2009.08) and butts up against $\gamma = 0$ (Figure 4).

The feature slowly dissipates until roughly the 7.0 yr slice. This change in the spectral index *is* contemporaneous with the "first" chromatic timing event in PSR J1713+0747 (Demorest et al. 2013; Lam et al. 2018b). Investigating further, we see that a similar feature exists in the single-pulsar red-noise analysis for this pulsar, shown by the orange dashed curve in Figure 4. We will return to this feature and investigate this correlation with the "first" PSR J1713+0747 ISM event in Section 7.

Additionally, the recovered value of γ coincident with the "kink" does not significantly vary. Rather than moving toward larger values, indicative of the recovery of a "steeper" process, the spectral index recovery remains broad and stagnant. While this is in no way dismissive of a GW-related event occurring during this time, it is further evidence of the anomalous behavior of our Bayesian analyses throughout the "kink."

6. Investigating the Anomalous Signal

In this section we investigate the anomalous GWB signal in our sliced analysis, which includes the "kink" in the UL analysis and the large spike in Bayes factor contemporaneously. We run through a number of the diagnostic analyses that were completed and summarize our mitigation strategy.

6.1. Slice-specific Simulations

While the evolution of our UL and Bayes factors seems to be unexpected, these types of statistics are expected to evolve stochastically depending on how the particular noise and Gaussian process realizations interact with the data. This is evident from the simulations run in Section 4. Therefore, it is no surprise that the UL or Bayes factor is nonmonotonic in moving from any given slice of the data to the next. The parameter space for a PTA is large, and the data sets for the individual pulsars may interact in complex ways with various lengths of observation, red-noise frequencies used, and WN parameter characterization. However, such a large and continuous rise in the Bayes factors and ULs seems to be worth further investigation.

The "kink" in the UL time series of Figure 3 lies far outside of the 90% confidence intervals of the simulations in Section 4. Noting that noise parameter characterization can change substantially over time, this large deviation inspired a *new* set

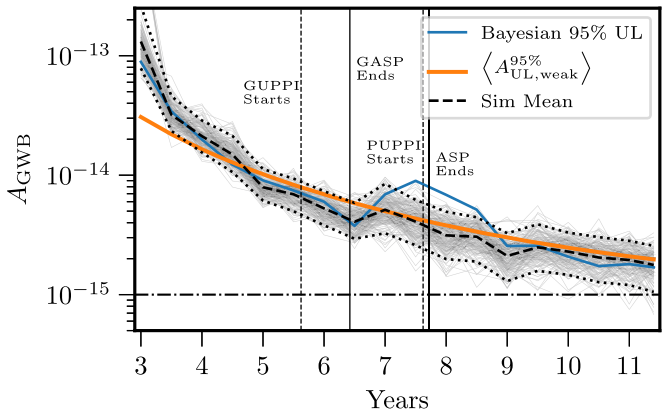


Figure 5. UL analyses of simulated NANOGrav 11 yr–like data sets with slice-dependent noise characteristic. Data were simulated for each slice based on the noise parameters recovered from the sliced analyses on the real data set. This is in contrast to the simulations done in Section 4, where the noise parameters from the full data set were used to simulate a full data set, which was then sliced for the UL analyses. One can see that the difference in noise parameters makes the "kink" of the blue trace somewhat less significant. The orange line shows the expected evolution and is identical to the blue line in Figure 1. The vertical lines show when the new GUPPI/PUPPI back ends came into use and when the old GASP/ASP back ends were phased out.

of simulations, in order to better characterize the degree to which the "kink" is just a statistical fluctuation. This set of simulations uses the same techniques as in Section 4.2, with one major difference. Rather than use the noise parameters from the full 11 yr analysis to simulate the pulsar data sets, the noise parameters retrieved from a single-pulsar analysis *at each slice* were used to build a simulation of that slice. This gave us an injection that was "truer" to the knowledge at the time of a particular slice. In practice, this involves making a whole new data set for each slice. This analysis was done on a set of data sets with an injection of $A_{\rm GWB} = 10^{-15}$.

The mean UL and 90% confidence interval are plotted in Figure 5. Besides the mostly monotonic trend in the average UL, there is a characteristic plateau that starts just around the increase in the real data. This plateau appears near the changeover in our pulsar back end at the GBT from GASP to GUPPI. GUPPI and PUPPI allow for much wider band observations of radio pulses, which allows us to do more accurate mitigation of DM fluctuations. While no causal relation has been found between this changeover and the "kink," we will present a number of results in Section 6.7 that summarize further investigations.

Even with this plateau in the same era as the "kink," the values of the UL in the 7.5 yr slice are outside of the 90% confidence interval of the simulations done. The fact that the "kink" is lessened in significance when simulating with noise estimated from each slice suggests either significant covariance between noise and GW signal at these epochs or nonstationary noise features. The results from these simulations, along with the large Bayes factors, prompted us to examine which pulsars seemed to be most responsible for the signal.

6.2. Dropout Analysis

It is important to point out that a number of different types of GW signal search are done by the NANOGrav Collaboration on each new data set. In addition to searching for various types of stochastic backgrounds, a search for single sources of GWs from SMBBHs (Aggarwal et al. 2019) and a search for GW memory (Aggarwal et al. 2020) from binary coalescences[45] were performed. The most recent studies involving these searches cover both the NG9a and NG11a data sets. In both cases there was mild evidence for GW signals in NG9a, which decreases in the analyses of NG11a. In the case of the single-source search of Aggarwal et al. (2019), PSR J0613−0200 was found to be responsible for the spurious signal and both PSR J0030+0451 and PSR J1909−3744 were responsible for the burst-with-memory anomaly in NG9a. As mentioned in those manuscripts, both noise in the individual pulsars and the low-sensitivity sky positions of PSR J0030+0451 and PSR J0613−0200 were to blame. We will see that these pulsars again appear as culprit pulsars in this analysis. These pulsars were tagged as being the source of the spurious signals using a Bayesian dropout analysis. In none of the cases above was a robust detection of GWs found.

Following Aggarwal et al. (2019), a dropout analysis was undertaken on the sliced data set. The dropout analysis is a new technique for isolating spurious noise sources from particular pulsars in a PTA GW analysis (Aggarwal et al. 2019). In a dropout analysis the signal being analyzed is coded with a so-called *dropout parameter*. These parameters multiply the signal amplitude and are sampled in the MCMC analysis. They are binary in the sense that, depending on the sample, the dropout parameter is either 1 or 0, turning the signal on/off. This allows one to use the Bayesian analysis to determine the signal model in terms of which pulsars prefer the signal to be turned on in the analysis. See S. J. Vigeland et al. (2019, in preparation) for more details.

Here we have used the GWB dropout analysis in `enterprise_extensions` to look at which pulsars favor a common red-noise signal across the slice analysis. The only pulsar with an odds ratio significantly higher than 1 in the 7.5 yr slice (i.e., more samples favor the presence of a stochastic background) is PSR B1937+21. It has long been known that this pulsar has a great deal of red noise (Kaspi et al. 1994), so much that it was not included in the analysis by Arzoumanian et al. (2016). As can be seen in Figure 6, removing this pulsar from the analysis decreases the Bayes factor and the UL during this era. Therefore, *some* of the spurious signal in the "kink" era is due to this pulsar. However, while the UL and Bayes factors decrease across this set of time slices, the main features of the "kink" are still present in the statistics.

6.3. Single-pulsar GWB Statistics

In search of other pulsars that might be responsible for this spurious signal, an exhaustive analysis of the GWB statistics was done for each individual pulsar. This analysis is often carried out to characterize the robustness of PTA GW statistics and has been used in the past to rank the sensitivity of pulsars to the GWB. These individual pulsar ULs are in turn used to do cumulative analyses where pulsars are added until the UL asymptotes to a stable and more robust value.

In an attempt to track down pulsars that could possibly be culprits in causing the "kink," we ran individual Bayes factor and UL analyses on all 34 pulsars in the GW analysis, over the slices in question. This was not done initially because it is computationally intensive, requiring upward of 1000 individual

[45] A rudimentary version of a generic burst search, based on the signal model and search algorithm of Ellis & Cornish (2016), was also done on this data set in the course of these investigations, with no significant evidence for GWs.

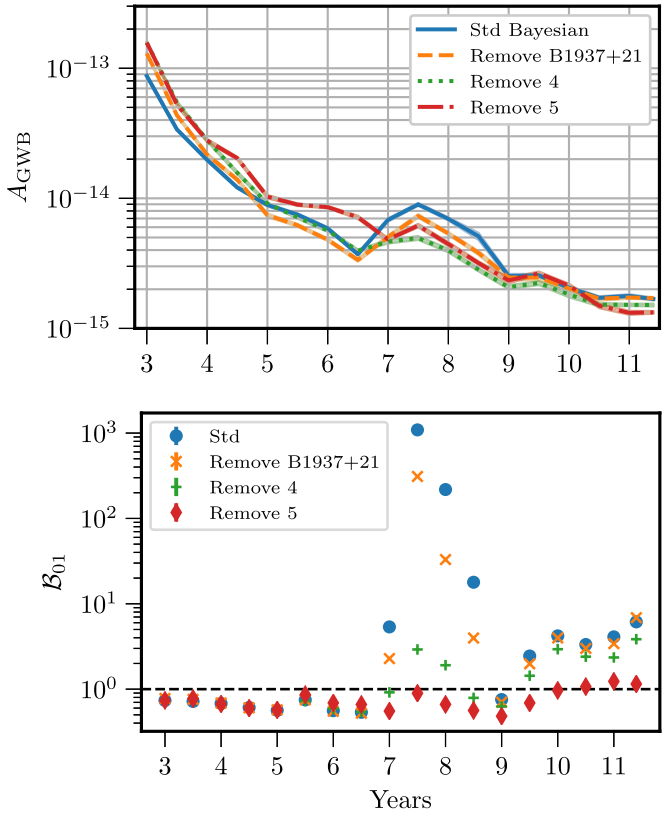


Figure 6. ULs and Bayes factors are shown for sliced analyses of the NANOGrav 11 yr data set, where some combination of culprit pulsars were removed. The solid blue line and blue circles are from the standard Bayesian analysis and are the same as in Figure 3. The orange dashed line and orange crosses are from an analysis where PSR B1937+21 has been removed. The green dotted line and green plus signs are from an analysis where PSR B1855+09, PSR B1937+21, PSR J0030+0451, and PSR J0613−0200 have been removed. The red dotted–dashed line and red diamonds are from an analysis where PSR B1855+09, PSR B1937+21, PSR J0030+0451, PSR J0613−0200, and PSR J1909−3744 have been removed.

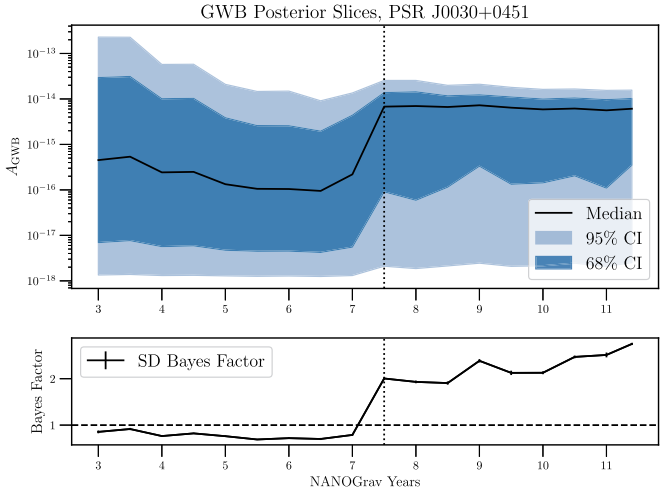


Figure 7. Evolution of the GWB statistics in PSR J0030+0451. The top panel shows the median and 68% and 95% confidence intervals for the posterior on $A_{\rm gwb}$ at each slice. The bottom panel shows the Savage–Dickey Bayes factor calculated at each slice. Note, in both cases, the jump that occurs at the 7.5 yr slice.

analyses. Plots of all of the pulsars' statistical evolution can be found online, but Figure 7 summarizes those of one interesting candidate source of the anomalous statistics, PSR J0030+0451. Four culprit pulsars were identified in these analyses: PSR B1855+09, PSR J0030+0451, PSR J0613−0200, and PSR J1909−3744. These pulsars all show either a similar feature to the full PTA analysis or a sharp rise in the Bayes factor in the 6.5–8.5 yr time span. This same feature does not appear in the UL and Bayes factor time series for PSR B1937+21, but we group it with these pulsars since it affects the GWB statistics in this era, as determined by the dropout analysis. After finding these candidates, the most straightforward test of their responsibility for the "kink" was to remove each of them and run the analysis on the remaining pulsars in the PTA.

Analyses were done with each of the pulsars removed individually and in all subsets of the culprit pulsars. We only discuss the results of removing either all of the pulsars or four of the pulsars while keeping PSR J1909−3744 in the analysis, since these are the most interesting cases.

Figure 6 shows these cases. When four of the culprits (not including PSR J1909−3744) were removed from the analysis, the "kink" strongly decreases in the time series and the evolution of the UL falls well within the 90% confidence interval of the simulations in Figure 5. The Bayes factor time series still shows a remnant feature during this period, but when PSR J1909−3744 is removed, the Bayes factor time series decreases even further. The removal of this pulsar drastically reduces our sensitivity to the GWB across the time span of interest. This is not surprising since PSR J1909−3744 is one of our most precisely timed pulsars, but in the next section it will be important to include this pulsar in our mitigation strategy.

6.4. Free Spectral Red-noise Models

While it is important to identify which pulsars have such strong, spurious GW detections, we are primarily interested in finding a way of mitigating the noise in these pulsars so they can be included in the full PTA analysis. In Section 6.3 we demonstrated that removing a small subset of the pulsars from the GW analysis removed the "kink." While it is reassuring that we have isolated the spurious detection of a GWB to a handful of pulsars, we would like to devise a mitigation strategy for this type of noise and understand the root cause more thoroughly, especially since one of our most sensitive pulsars is included in the list.

As discussed in Section 3, a free spectral model is another tool in the standard PTA data analysis toolbox that we can use to model the red noise in these individual pulsars. The free spectral model uses a free parameter for the amplitude of the Fourier basis red noise at each frequency modeled, which provides a much larger parameter space for the noise by not restricting the noise in a given pulsar to follow any type of functional dependence in frequency. Hence, such a model can model the noise in the lowest-frequency bins of an individual pulsar *and* any unmodeled higher-frequency noise independently, which the power-law model is unable to do.

In order for the free spectral model to cover the same number of frequencies as the power-law model, upward of 30 parameters need to be added per pulsar. Historically, this is the primary reason for not including free spectral models in PTA noise analyses, as the increased parameter space becomes computationally infeasible to search over. Additionally, forthcoming work (J. Simon et al. 2019, in preparation) shows that these models do not compete well in a Bayesian model selection framework because the large Occam penalty of the

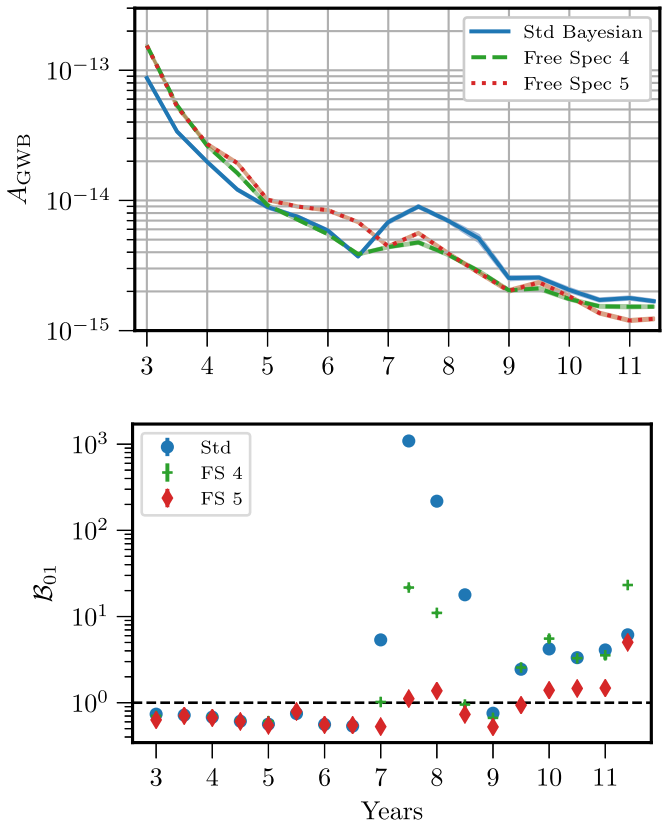


Figure 8. ULs and Bayes factors are shown for sliced analyses of the NANOGrav 11 yr data set where some combination of culprit pulsars were analyzed with a free spectral-noise model. The solid blue line and blue circles are from the standard Bayesian analysis and are the same as in Figure 3. The green dotted line and green plus signs are from an analysis where PSR B1855+09, PSR B1937+21, PSR J0030+0451, and PSR J0613-0200 have free spectral-noise models. The red dotted–dashed line and red diamonds are from an analysis where PSR B1855+09, PSR B1937+21, PSR J0030+0451, PSR J0613−0200, and PSR J1909−3744 have free spectral-noise models. Note the differences in the 11.4 yr slice for these analyses, as compared to Figure 6.

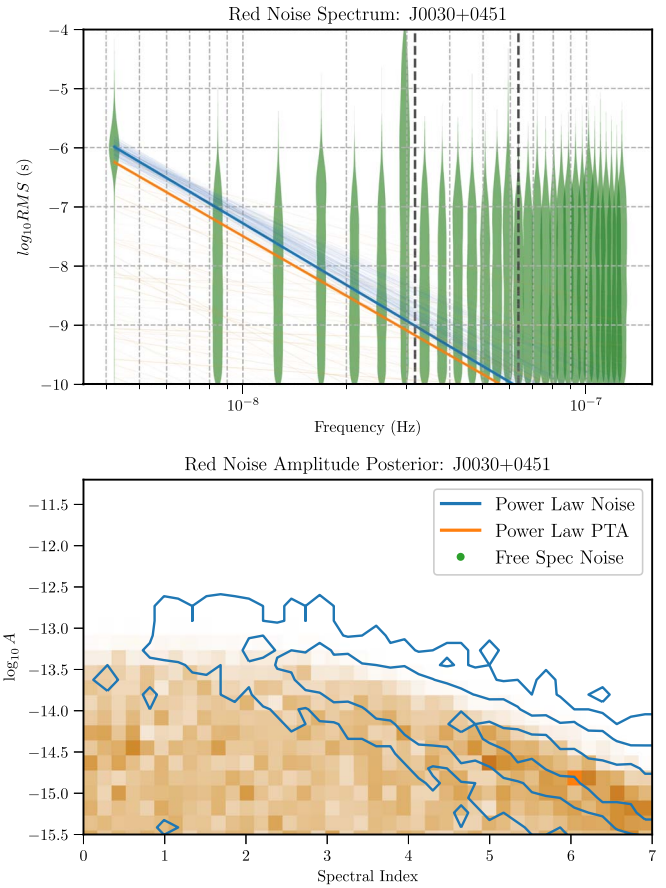


Figure 9. Noise model posteriors for PSR J0030+0451 at the 7.5 yr slice. The top panel shows realizations of the power-law red noise as straight lines in the log–log plot. The thick traces are the power law for the maximum likelihood values, while the fainter traces show a representative sample of power-law realizations from the Bayesian analysis. The legend in the bottom panel is also accurate for the top panel. Power-law red-noise realizations from the individual noise analyses (blue) and full PTA analysis (orange) are shown. The (green) violin plots show the posteriors for the free spectral-noise model at each frequency. One can judge the significance of a detection by how separated the violin plot is from the lowest amplitudes. The crowding at higher frequencies stems from the linearly spaced frequencies on a log scale. The vertical dashed lines show frequencies at 1 and 2 $\rm yr^{-1}$. The bottom panel shows the 2D posteriors for the power-law noise models, in γ and $\log_{10} A_{\rm GWB}$. The blue contours show the posterior from the individual noise run, while the orange heat map shows the posterior for the full PTA run for the same pulsar, PSR J0030+0451. Note that while the individual pulsar noise run shows closed contours for the power-law model, the full PTA has a very diffuse, nonsignificant posterior. It is suspected that this RN power has moved into the common red-noise process, which shows a strong detection in the 7.5 yr slice.

additional parameters cancels out the ability of these many parameters to describe the noise accurately.

With those caveats in mind, we undertook another set of PTA GW analyses using free spectral models for the culprit pulsars. The results are summarized in Figure 8. These results demonstrate that the free spectral model is effective at mitigating the spurious noise features in both the UL and Bayes factor analyses of the NANOGrav 11 yr data set. It is salient to compare the longest slices, i.e., >10 yr, where the Bayes factor is slowly increasing. With all of the culprit pulsars removed in Figure 6, the Bayes factor still remained at or below 1 in these late slices, showing no early signs of any type of signal. However, even though the use of a free spectral noise model for the five culprit pulsars mitigates the spurious features identified in this paper, it begins to favor the signal model in the long run, focusing on the points >10 yr in Figure 8. If this growth in the Bayes factor were the beginning indications of a real signal, its growth would be indicative of the amplitude of the underlying GWB. A separate, ongoing investigation is addressing this question by injecting GWB signals into the 11 yr data set and will be published separately.

Red-noise amplitude spectral densities for the 7.5 yr slice are shown in Figure 9 for PSR J0030+0451. The free spectral parameter posteriors are compared to a sample of the power-law posterior amplitude spectral densities. The thick straight lines are the power-law spectrum for the maximum likelihood values for these power-law parameters. The bottom panel shows the 2D posteriors for the power-law parameters, in both the single-pulsar noise run and the full PTA analysis. In the top panel note that the only significant free spectral parameter (i.e., sufficiently separated from the minimum amplitude) is the one for the lowest frequency. This points to a possible cause for the anomalous signal we are seeing from a few pulsars. The lowest frequency will model all power within a δf defined by the inverse of the time span, but it is limited by the second-lowest frequency, where there is no substantial evidence for power. One conjecture, partly substantiated by comparing the different 2D power-law posteriors, is that the power law is able to find this power at low frequencies, but since the signal is only

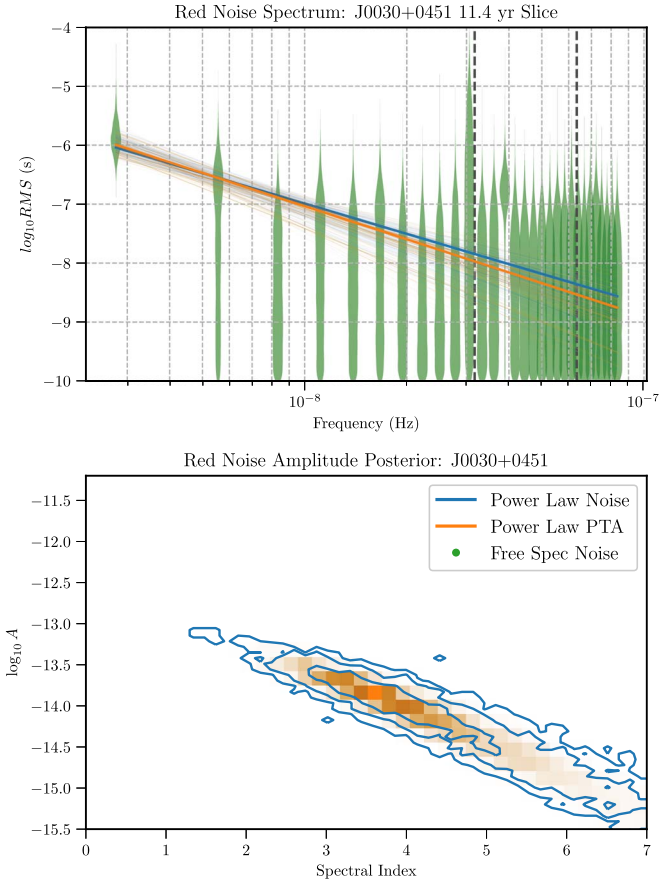


Figure 10. Noise model posteriors for PSR J0030+0451 for the full data set. The various elements are the same as in Figure 9, but the results reveal how the "kink" is in part mitigated with time. Looking at the top panel, in addition to the lowest frequency, the second-lowest-frequency posterior for the free spectral-noise model is also above the WN floor. The power-law model is not as hindered by the WN floor, and as can be seen in the bottom panel, the power-law noise model effectively holds much of the red-noise power in this pulsar, rather than allowing it to all move into the common red-noise process.

significant at one frequency, this power is allowed to transfer between the pulsar red-noise model and the GWB common red-noise process.

Compare these results to the same data products from the full NANOGrav 11 yr data set in Figure 10. The second-lowest-frequency free spectral parameter is more significant, and the power-law model is much more consistent between the single-pulsar noise analysis and the full PTA analysis.

6.5. Anomalous Signal Toy Model

To try to further understand the spurious signal seen in the 7.5 yr slice, we ran a number of simulations using PSR J0030+0451-like data sets in an attempt to duplicate the jump in the Bayes factor seen in Figure 7. In Figure 11 we show the results of the simulations and analyses. In each case the noise parameters obtained from the analysis of the real PSR J0030+0451 data were used to try to replicate the same Bayes factor. In the first four cases no GWB background was injected. Two different types of noise injection were used, corresponding to either the power-law noise model values for this pulsar or the free spectral noise model parameters. When the same model is used in the analysis as the injection model, the Bayes factor is near 1. However, when a free spectral noise model is used in the injection and a power-law noise model is used in the

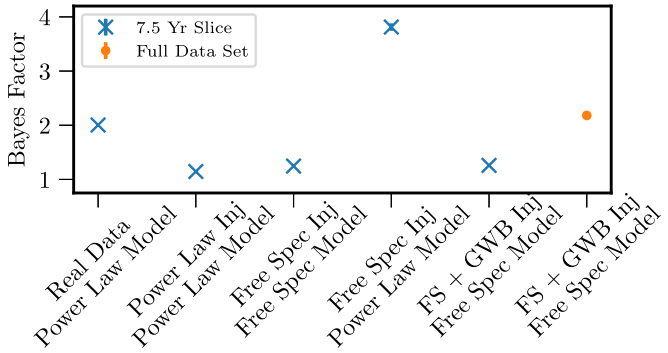


Figure 11. Bayes factor for the GWB amplitude from various noise injections and analyses. Error bars are included but are smaller than the markers in most cases. The labels show what type of injection and noise model were used. Note that using a free spectral model injection while using a power-law model in the analysis results in a higher significance detection of a GWB.

analysis, the Bayes factor roughly triples. This is further evidence in support of the main conclusion in this work—the use of an inaccurate noise model can lead to anomalous detections of a GWB.

One may question whether the free spectral model will remove any evidence for a GWB entirely. The last two GWB Bayes factors in Figure 11 show simulations where an actual power-law GWB was injected in addition to the free spectral noise. The $A_{\rm GWB}$ injected was the maximum likelihood value from the 7.5 yr slice of PSR J0030+0451's real data set. One can see that the data analysis in the 7.5 yr slice does not detect the GWB as a separate red-noise injection. The full 11 yr data set is able to better differentiate the GWB. This supports the conclusion from Section 6.4 and born out in the full time span analysis with free spectral models shown in Figure 8. One expects earlier slices to have lower Bayes factors, as well as a slow rise in the Bayes factor as we accrue longer data sets.

Obviously these Bayes factors are still rather close to 1, i.e., even odds, but since the GWB signal significance is expected to grow slowly with time, it is the comparison and trends with which we are most concerned.

6.6. Power Accounting

One would like to quantify the difference between the noise model using a power-law prior for the Gaussian processes and the free spectral models that have effectively mitigated the "kink," removing the false-positive detection of a GWB. One way in which this can be done is to calculate the posterior probability distributions for the power in these red-noise channels. Here we compare the power by calculating it from the models used as priors for the Gaussian process over the frequencies sampled in the Gaussian process coefficients. For the power law this will be a sum of the power-law values across the sampled frequencies times the frequency bin size,

$$P = \sum_{i}^{N_f} A_{\rm RN}^2 \left(\frac{f_i}{f_{\rm yr}}\right)^{-\gamma} \ {\rm yr}^3 \Delta f. \tag{13}$$

In the case of the free spectral model the amplitudes are indexed, i.e.,

$$P = \sum_{i}^{N_f} \rho_i^2 \ {\rm yr}^3 \Delta f. \tag{14}$$

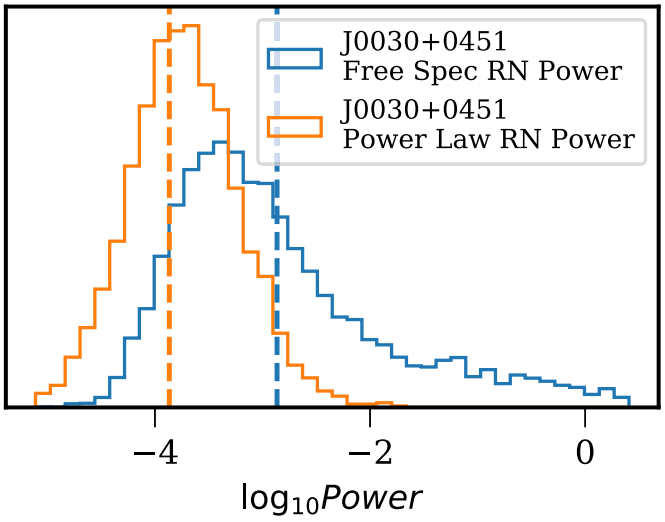


Figure 12. Power posteriors for PSR J0030+0451 comparing the power in the power-law model vs. the free spectral model. The mean values for the power are different by roughly an order of magnitude.

In Figure 12 we show the calculated posteriors for the power using these two models on the data from PSR J0030+0451, one of the main culprit pulsars in Section 6.4. The WN parameter posteriors are basically unchanged for this pulsar between the two models, so the free spectral model is effectively absorbing power that is otherwise unmodeled. While the lowest frequency is most likely the issue for this pulsar, as mentioned in Section 6, the majority of the power in the free spectral model is at high frequencies and is probably not to blame for the spurious GWB detection in the previous section. However, since the power posteriors for all of the culprit pulsars are different by approximately an order of magnitude between the power-law and free spectral models, this noise has been flagged as an obvious area for improvement in our per pulsar noise modeling. A number of in-progress projects and a forthcoming paper (J. Simon et al. 2019, in preparation) are devoted to mitigating noise of this sort in a number of NANOGrav pulsars.

6.7. Ineffective Analyses

While the use of the free spectral model has mitigated the spurious GW signal of Figure 3, there are a number of additional investigated analyses that had either a neutral or a minimal effect on the GW statistics. Here we summarize them, in part to inform the interested expert and in part to motivate their use in upcoming work that investigates more comprehensive noise models for pulsars.

The standard analyses were run with various versions of the JPL solar system ephemeris and the Bayesian solar system ephemeris model, BayesEphem (NG11b). All results were qualitatively the same as results shown in the sections above between DE421, DE430, DE436 (Folkner et al. 2009, 2014; Folkner & Park 2016), and using BayesEphem. In particular, all results showed the same anomalous GW statistics in the 6.5–8.5 yr slices.

The choice of frequencies sampled by the Gaussian processes that are modeling the underlying stochastic signals has been shown to affect the signal analysis (van Haasteren & Vallisneri 2015; Ellis & Cornish 2016). A number of strategies for choosing the frequencies, including using a log spacing rather than a linear spacing and choosing the frequencies uniformly across the slices, were carried out. These had some effect in mitigating the "kink" but were not as effective as the methods described in previous sections.

Lastly, the proximity in time of the "kink" to the changeover in the back ends used for observing is intriguing; however, analyses testing any causation were inconclusive. These included a number of analyses with different combinations of the overlapping GASP/ASP and GUPPI/PUPPI data. Here it is difficult to separate the effect of narrower bandwidths from the change in GW statistics. We also modeled back-end red noise, similar to the "band noise" of Lentati et al. (2016), but rather than restricting the noise to a specific observing frequency band, we restricted the noise to a specific back-end type. While the red-noise parameters seem to be significantly different between back ends in some pulsars, the use of these models did not help to mitigate the spurious GW statistics.[46]

7. Mitigating Low Spectral Index Noise

Here we turn our attention to mitigating the transient WN feature described at the end of Section 5. The analysis in which the spectral index is varied intimates PSR J1713+0747 as a strong candidate for this WN event in the common red-noise process. Sharp noise features in the time-series domain can manifest as low spectral index noise in the power spectral density; hence, an attempt at mitigating the noise in this pulsar was undertaken.

The recent observation of a second ISM event in PSR J1713 +0747 has prompted new work on chromatic noise models for our pulsars (Lam et al. 2018b). This second event does not occur during the time span of the NG 11 yr data set, but the first ISM event occurring near MJD 54,750 (2008.78) and first observed in Demorest et al. (2013) does, between the 4.0 and 4.5 yr slices in this analysis. As has been shown in other recent publications (Aggarwal et al. 2019, 2020), unmodeled noise in a single pulsar can appear in a common PTA signal. In order to investigate whether the first ISM event in PSR J1713+0747 is the cause of the significant WN appearing in the common red-noise signal, we ran the analysis over again using the same DM noise model[47] first used in Lam et al. (2018b). That model consists of a timing-model fit for a linear and quadratic trend in the DM (DM1 and DM2), chromatic red noise modeled with a Gaussian process, and a phenomenological model for the dip in the DM variations. In Lam et al. (2018b) this consisted of two exponential dips modeled as

$$\Delta \mathrm{DM}_{\mathrm{dip}} = -\mathcal{T}\Theta(t_0)\exp\left(-(t-t_0)/\tau\right), \tag{15}$$

where $\Theta(t_0)$ is the Heaviside function and the amplitude ($\mathcal{T}$), time of occurrence (t_0), and decay time (τ) were fit for in the Bayesian analysis. In the present work we only fit for one exponential dip, to model the first ISM event. The Gaussian process and exponential dips are implemented in `enterprise`. This model replaces the piecewise DMX model used in NG11a and NG11b. This model was also studied in depth in Wang et al. (2019), where a Bayesian cross-validation study

[46] Anyone interested in other analyses undertaken during this research, or seeing the results of those discussed in this subsection, should feel free to contact the first author for more information.

[47] This type of model has been used previously in other analyses (Lentati et al. 2016).

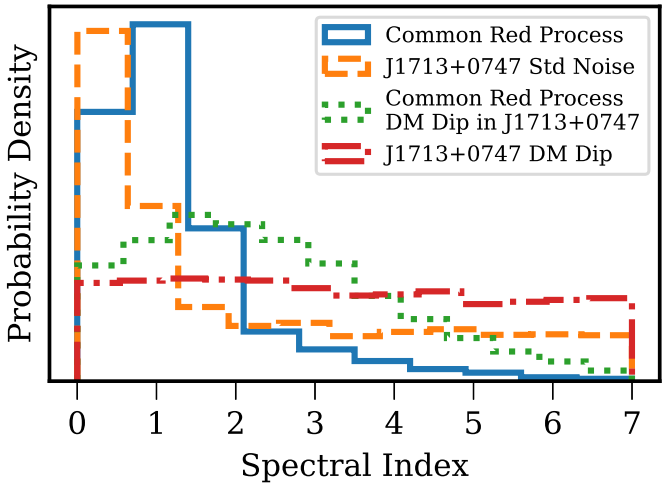


Figure 13. Comparison of posteriors for the spectral index of a power-law red-noise process. Posteriors are shown for both the full PTA analysis and the individual PSR J1713+0747 noise analysis. Note that using a more tailored noise model on this one pulsar has a significant impact on the spectral index of the common process in the full PTA.

showed convincing evidence for the preference of this model. The better performance of this model for DM variations in this particular case is explained by a lack of DMX bins in the time span around the minimum of the rapid fluctuation. From MJD 54,707 (2008.66) until 100 days past the event there are only five DMX bins. This coarse sampling is possibly inadequate for such a relatively short-timescale, high-amplitude event. The lack of interobserving band TOAs also limits the precision of the ΔDM measurement, because the measurement is done within a single, narrow receiver band.

The results of this newer model on the posteriors for the spectral index are dramatic, as can be seen in Figure 13.

The posteriors show reduced significance of an unmodeled WN transient, i.e., the posterior is no longer butted up against the $\gamma = 0$ end of the prior. This noise is mitigated not only in the individual pulsar noise but also in the common process, revealing that this event can have an important effect on the GWB analysis and can be mitigated with a more tailored noise model.[48]

With the appearance of a second ISM event in PSR J1713 +0747 in the NANOGrav 12.5 yr data set, this type of model will be necessary to properly use this pulsar in our GW analyses.

8. Summary and Conclusions

Here we have used the standard tools of PTA GW analysis to investigate the evolution of GW statistics in the NANOGrav 11 yr data set. After finding transient features in the sliced data GW analyses, we undertook an in-depth analysis to characterize and mitigate any possible sources of noise that might lead to these features. The transient GWB detection that peaks during the 7.5 yr slice was found to be due, in large part, to five "culprit" pulsars. These pulsars were identified by a combination of their individual red-noise analyses, single-pulsar GWB upper limits, and a new Bayesian PTA data analysis technique known as the dropout method. In order to test whether these pulsars were responsible for the large GWB signal, a new set of GW statistics were derived for the NANOGrav 11 yr data set with TOAs from these pulsars removed. Once a set of pulsars was identified to be responsible for these artifacts, the false signal was mitigated using a free spectral noise model. This demonstrates the importance of characterizing the noise in pulsars correctly and demonstrates that incorrect noise models can lead to false positives in our Bayesian analyses. These results highlight a number of strategies important when searching PTA data for a stochastic GWB:

1. Study the noise *evolution* of individual pulsars.
2. Look at the evolution of single-pulsar UL and detection analyses to see *when* various signals become significant and *how long* they remain significant.
3. Attempt to use more tailored noise models for a given pulsar.

This last point is especially important for the type of transient WN feature seen in the 4.5 yr slice. As shown in Section 7, this transient feature, causing the varying spectral index analysis to prefer a spectral index of zero, and contemporaneous with the first ISM event in PSR J1713+0747, was mitigated using a phenomenological model for the chromatic time delays consisting of a Gaussian process + exponential dip.

This work has shown that the standard tools for GW analysis in pulsar timing data, while sufficient to mitigate some of the noise features in NG11a, need to be updated, as the sensitivity of our detector has revealed a new noise floor. These considerations have moved NANOGrav to undertake a program of Bayesian model selection using a full suite of individual, tailored noise models for our pulsars that pay closer attention to the astrophysics causing the noise in each case. This will be presented in an upcoming paper (J. Simon et al. 2020, in preparation) presenting these new models and the results of model selection on the most sensitive NG11a pulsars.

Author contributions. This paper is the result of the work of dozens of people over the course of more than 13 years. We list specific contributions below. J.S.H. ran the sliced analyses and led the paper writing. J.S., S.R.T., M.T.L., S.J.V, K.I., and J.S. K. contributed substantially to paper writing, discussion, and interpretation of results. M.T.L. helped with analyses. J.S.H. and S.R.T. developed the formalism in Section 4. Z.A., K.C., P. B.D., M.E.D., T.D., J.A.E., E.C.F., R.D.F., E.F., P.A.G., G.J., M.L.J., M.T.L., L.L., D.R.L., R.S.L., M.A.M., C.N., D.J.N., T. T.P., S.M.R., P.S.R., R.S., I.H.S., K.S., J.K.S., and W.Z. ran observations and developed the 11 yr data set.

The NANOGrav project receives support from National Science Foundation (NSF) PIRE program award No. 0968296 and NSF Physics Frontier Center award No. 1430284. NANOGrav research at UBC is supported by an NSERC Discovery Grant and Discovery Accelerator Supplement and by the Canadian Institute for Advanced Research. J.S.H. would like to thank Joseph Romano for a number of in-depth discussions concerning the results of the numerous analyses done for this project. M.V. and J.S. acknowledge support from the JPL RTD program. 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. S.R.T. was partially supported by an

[48] It should be noted that there appears to be a large amount of power at lower spectral indices, shallower than expected from either pulsar spin noise or a GWB. This points to a systemic issue with the current noise models used in PTA GW analysis and will be confronted in an upcoming NANOGrav publication (J. Simon et al. 2020, in preparation).

appointment to the NASA Postdoctoral Program at the Jet Propulsion Laboratory, administered by Oak Ridge Associated Universities through a contract with NASA. S.R.T. thanks ERS for fruitful discussions. J.A.E. was partially supported by NASA through Einstein Fellowship grants PF4-150120. S.B.S. was supported by NSF award No. 1458952. P.T.B. acknowledges support from the West Virginia University Center for Gravitational Waves and Cosmology. This work was supported in part by National Science Foundation grant No. PHYS-1066293 and by the hospitality of the Aspen Center for Physics. Portions of this work performed at NRL are supported by the Chief of Naval Research. This research was performed in part using the Zwicky computer cluster at Caltech supported by NSF under MRI-R2 award No. PHY-0960291 and by the Sherman Fairchild Foundation. A majority of the computational work was performed on the Nemo cluster at UWM supported by NSF grant No. 0923409. Parts of the analysis in this work were carried out on the Nimrod cluster made available by S.M.R. Data for this project were collected using the facilities of the National Radio Astronomy Observatory and the Arecibo Observatory. The National Radio Astronomy Observatory and Green Bank Observatory are facilities of the NSF operated under cooperative agreement by Associated Universities, Inc. The Arecibo Observatory is operated by the University of Central Florida, Ana G. Méndez-Universidad Metropolitana, and Yang Enterprises under a cooperative agreement with the NSF (AST-1744119). This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. Some of the algorithms used in this article were optimized using the Blue Waters allocation "Accelerating the detection of gravitational waves with GPUs." The Flatiron Institute is supported by the Simons Foundation.

Software: `enterprise` (Ellis et al. 2017), `PTMCMCSampler` (Ellis & van Haasteren 2017), `enterprise_extensions` (Taylor et al. 2018), `TEMPO2` (Hobbs et al. 2006), `libstempo` (Vallisneri 2015), `Scipy` (Jones) and `Numpy` (Oliphant).

ORCID iDs

J. S. Hazboun https://orcid.org/0000-0003-2742-3321
J. Simon https://orcid.org/0000-0003-1407-6607
S. R. Taylor https://orcid.org/0000-0003-0264-1453
M. T. Lam https://orcid.org/0000-0003-0721-651X
S. J. Vigeland https://orcid.org/0000-0003-4700-9072
K. Islo https://orcid.org/0000-0001-5123-9917
J. S. Key https://orcid.org/0000-0003-0123-7600
P. T. Baker https://orcid.org/0000-0003-2745-753X
P. R. Brook https://orcid.org/0000-0003-3053-6538
S. Burke-Spolaor https://orcid.org/0000-0003-4052-7838
S. Chatterjee https://orcid.org/0000-0002-2878-1502
J. M. Cordes https://orcid.org/0000-0002-4049-1882
N. J. Cornish https://orcid.org/0000-0002-7435-0869
F. Crawford https://orcid.org/0000-0002-2578-0360
K. Crowter https://orcid.org/0000-0002-1529-5169
H. T. Cromartie https://orcid.org/0000-0002-6039-692X
M. DeCesar https://orcid.org/0000-0002-2185-1790
P. B. Demorest https://orcid.org/0000-0002-6664-965X
T. Dolch https://orcid.org/0000-0001-8885-6388
R. D. Ferdman https://orcid.org/0000-0002-2223-1235
E. Fonseca https://orcid.org/0000-0001-8384-5049
N. Garver-Daniels https://orcid.org/0000-0001-6166-9646
P. Gentile https://orcid.org/0000-0001-8158-683X
D. Good https://orcid.org/0000-0003-1884-348X
A. M. Holgado https://orcid.org/0000-0003-4143-8132
E. A. Huerta https://orcid.org/0000-0002-9682-3604
R. Jennings https://orcid.org/0000-0003-1082-2342
M. L. Jones https://orcid.org/0000-0001-6607-3710
A. R. Kaiser https://orcid.org/0000-0002-3654-980X
D. L. Kaplan https://orcid.org/0000-0001-6295-2881
L. Z. Kelley https://orcid.org/0000-0002-6625-6450
L. Levin https://orcid.org/0000-0002-2034-2986
A. N. Lommen https://orcid.org/0000-0003-4137-7536
D. R. Lorimer https://orcid.org/0000-0003-1301-966X
R. S. Lynch https://orcid.org/0000-0001-5229-7430
D. R. Madison https://orcid.org/0000-0003-2285-0404
M. A. McLaughlin https://orcid.org/0000-0001-7697-7422
S. T. McWilliams https://orcid.org/0000-0003-2397-8290
C. M. F. Mingarelli https://orcid.org/0000-0002-4307-1322
C. Ng https://orcid.org/0000-0002-3616-5160
D. J. Nice https://orcid.org/0000-0002-6709-2566
T. T. Pennucci https://orcid.org/0000-0001-5465-2889
N. S. Pol https://orcid.org/0000-0002-8826-1285
S. M. Ransom https://orcid.org/0000-0001-5799-9714
P. S. Ray https://orcid.org/0000-0002-5297-5278
X. Siemens https://orcid.org/0000-0002-7778-2990
R. Spiewak https://orcid.org/0000-0002-6730-3298
I. H. Stairs https://orcid.org/0000-0001-9784-8670
D. R. Stinebring https://orcid.org/0000-0002-1797-3277
K. Stovall https://orcid.org/0000-0002-7261-594X
J. Swiggum https://orcid.org/0000-0002-1075-3837
J. E. Turner https://orcid.org/0000-0002-2451-7288
M. Vallisneri https://orcid.org/0000-0002-4162-0033
R. van Haasteren https://orcid.org/0000-0002-6428-2620
C. A. Witt https://orcid.org/0000-0002-6020-9274
W. W. Zhu https://orcid.org/0000-0001-5105-4058

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, PhRvL, 116, 061102
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, PhRvL, 119, 161101
Aggarwal, K., Arzoumanian, Z., Baker, P. T., et al. 2019, ApJ, 880, 116
Aggarwal, K., Arzoumanian, Z., Baker, P. T., et al. 2020, ApJ, 889, 38
Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2018a, ApJ, 859, 47
Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2014, ApJ, 794, 141
Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2015, ApJ, 813, 65
Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2016, ApJ, 821, 13
Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2018b, ApJS, 235, 37
Bailes, M., Barr, E., Bhat, N. D. R., et al. 2018, arXiv:1803.07424
Burke-Spolaor, S., Taylor, S. R., Charisi, M., et al. 2019, A&ARv, 27, 5
Caballero, R. N., Lee, K. J., Lentati, L., et al. 2016, MNRAS, 457, 4421
Chen, S., Sesana, A., & Conselice, C. J. 2019, MNRAS, 488, 401
Cordes, J. M., & Lazio, T. J. W. 2002, arXiv:astro-ph/0207156
Cordes, J. M., & Shannon, R. M. 2010, arXiv:1010.3785
Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, ApJ, 762, 94
Desvignes, G., Caballero, R. N., Lentati, L., et al. 2016, MNRAS, 458, 3341
Detweiler, S. 1979, ApJ, 234, 1100
DuPlain, R., Ransom, S., Demorest, P., et al. 2008, Proc. SPIE, 7019, 7019D
Ellis, J., & Cornish, N. 2016, PhRvD, D93, 084048
Ellis, J., & van Haasteren, R. 2017, jellis18/PTMCMCSampler: Official Release (v1.0.0), Zenodo, doi:10.5281/zenodo.1037579
Ellis, J. A., Vallisneri, M., Taylor, S. R., & Baker, P. T. 2017, ENTERPRISE https://github.com/nanograv/enterprise
Folkner, W. M., & Park, R. S. 2016, JPL Planetary and Lunar ephemeris DE436, Tech. rep. 42-196, Jet Propulsion Laboratory, Pasadena, CA https://naif.jpl.nasa.gov/pub/naif/JUNO/kernels/spk/de436s.bsp.lbl
Folkner, W. M., Williams, J. G., & Boggs, D. H. 2009, IPNPR, 42-178, 1

Folkner, W. M., Williams, J. G., Boggs, D. H., Park, R. S., & Kuchynka, P. 2014, IPNPR, 42-196, 1
Foster, R. S., & Backer, D. C. 1990, ApJ, 361, 300
Hellings, R. W., & Downs, G. S. 1983, ApJ, 265, L39
Hobbs, G. 2013, CQGra, 30, 224007
Hobbs, G. B., Edwards, R. T., & Manchester, R. N. 2006, MNRAS, 369, 655
Jones, M. L., McLaughlin, M. A., Lam, M. T., et al. 2017, ApJ, 841, 125
Joshi, B. C., Arumugasamy, P., Bagchi, M., et al. 2018, JApA, 39, 51
Kaspi, V. M., Taylor, J. H., & Ryba, M. F. 1994, ApJ, 428, 713
Keith, M. J., Coles, W., Shannon, R. M., et al. 2013, MNRAS, 429, 2161
Kelley, L. Z., Blecha, L., Hernquist, L., Sesana, A., & Taylor, S. R. 2017, MNRAS, 471, 4508
Lam, M. T., McLaughlin, M. A., Cordes, J. M., Chatterjee, S., & Lazio, T. J. W. 2018a, ApJ, 861, 12
Lam, M. T., McLaughlin, M. A., Cordes, J. M., Chatterjee, S., & Lazio, T. J. W. 2018b, ApJ, 861, 132
Lam, M. T., Cordes, J. M., Chatterjee, S., et al. 2017, ApJ, 834, 35
Lee, K. J. 2016, in ASP Conf. Ser. 502, Frontiers in Radio Astronomy and FAST Early Sciences Symp. 2015, ed. L. Qain & D. Li (San Francisco, CA: ASP), 19
Lee, K. J., Bassa, C. G., Janssen, G. H., et al. 2014, MNRAS, 441, 2831
Lentati, L., Alexander, P., Hobson, M. P., et al. 2013, PhRvD, 87, 104021
Lentati, L., Alexander, P., Hobson, M. P., et al. 2014, MNRAS, 437, 3004
Lentati, L., Taylor, S. R., Mingarelli, C. M. F., et al. 2015, MNRAS, 453, 2576
Lentati, L., Shannon, R. M., Coles, W. A., et al. 2016, MNRAS, 458, 2161
MacKay, D. J. C. 2002, Information Theory, Inference & Learning Algorithms (Cambridge: Cambridge Univ. Press)
Madison, D. R., Cordes, J. M., Arzoumanian, Z., et al. 2019, ApJ, 872, 150
McLaughlin, M. A. 2013, CQGra, 30, 224008
McWilliams, S. T., Ostriker, J. P., & Pretorius, F. 2014, ApJ, 789, 156
Ng, C. 2018, in IAU Symp. 337, Pulsar Astrophysics the Next Fifty Years, ed. P. Weltevred (Cambridge: Cambridge Univ. Press), 179
Phinney, E. S. 2001, arXiv:astro-ph/0108028
Reardon, D. J., Hobbs, G., Coles, W., et al. 2016, MNRAS, 455, 1751
Romano, J. D., & Cornish, N. J. 2017, LRR, 20, 2
Rosado, P. A., Sesana, A., & Gair, J. 2015, MNRAS, 451, 2417
Rover, C., Messenger, C., & Prix, R. 2011, in Proc. PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding (Geneva: CERN), 158
Sampson, L., Cornish, N. J., & McWilliams, S. T. 2015, PhRvD, 91, 084055
Sazhin, M. V. 1978, SvA, 22, 36
Sesana, A. 2013, MNRAS, 433, L1
Shannon, R. M., & Cordes, J. M. 2017, MNRAS, 464, 2075
Shannon, R. M., Ravi, V., Lentati, L. T., et al. 2015, Sci, 349, 1522
Siemens, X., Ellis, J., Jenet, F., & Romano, J. D. 2013, CQGra, 30, 224015
Simon, J., & Burke-Spolaor, S. 2016, ApJ, 826, 11
Taylor, J. H., & Weisberg, J. M. 1982, ApJ, 253, 908
Taylor, S. R., Baker, P. T., Hazboun, J. S., Simon, J. J., & Vigeland, S. J. 2018, Enterprise_extensions https://github.com/stevertaylor/enterprise_extensions
Taylor, S. R., Gair, J. R., & Lentati, L. 2013, PhRvD, 87, 044035
Taylor, S. R., Lentati, L., Babak, S., et al. 2017a, PhRvD, 95, 042002
Taylor, S. R., Simon, J., & Sampson, L. 2017b, PhRvL, 118, 181102
Taylor, S. R., Vallisneri, M., Ellis, J. A., et al. 2016, ApJL, 819, L6
Vallisneri, M. 2015, libstempo https://github.com/vallis/libstempo
van Haasteren, R., & Levin, Y. 2013, MNRAS, 428, 1147
van Haasteren, R., & Vallisneri, M. 2014, PhRvD, 90, 104012
van Haasteren, R., & Vallisneri, M. 2015, MNRAS, 446, 1170
Verbiest, J. P. W., Lentati, L., Hobbs, G., et al. 2016, MNRAS, 458, 1267
Vigeland, S. J., Islo, K., Taylor, S. R., & Ellis, J. A. 2018, PhRvD, 98, 044003
Vigeland, S. J., & Siemens, X. 2016, PhRvD, D94, 123003
Wang, H., Taylor, S. R., & Vallisneri, M. 2019, MNRAS, 487, 3644
Williams, C. K., & Rasmussen, C. E. 2006, Gaussian Processes for Machine Learning (Cambridge, MA: MIT Press)
Wolszczan, A., & Frail, D. A. 1992, Natur, 355, 145
Yao, J. M., Manchester, R. N., & Wang, N. 2017, ApJ, 835, 29