The 2012 International Pulsar Timing Array (IPTA) Mock Data Challenge (MDC) is designed to test current Gravitational Wave (GW) detection algorithms. Here we will briefly outline two detection algorithms for a stochastic background of gravitational waves, namely, a first-order likelihood method and an optimal statistic method and present our results from the closed MDC data sets.

I. GENERATING THE RESIDUALS

The first step in any data analysis pipeline used in this paper is to generate the residuals and make sure that all of the fits converged. Here we use TEMPO2 with the general2 plugin to generate the residuals. For our analysis we will also need the design matrix for each pulsar which is obtained using the designmatrix plugin developed by R. van Haasteren and modified by J. Ellis. From inspection of the residuals, it is immediately obvious that all three datasets contain strong red noise in many of the residuals. We can also see “by-eye” that the residuals of dataset 1 have similar errorbars and that the errorbars of datasets 2 and 3 seem to vary quite a bit from pulsar to pulsar.

II. NOISE ESTIMATION

The next step in our analysis pipelines is to estimate the red and white noise levels in each set of residuals. A detailed description of the process is in preparation [2]. Here we will simply review the method and present the results from the Mock Data Challenge. We model the noise in the data as a sum of three processes: a red component with a power law power spectrum, a systematic white noise component that multiplies the residual error bars (EFAC), and an extra white noise component independent of the error bars (EQUAD). We work in the time domain, so these noise processes can be completely described by their covariance matrices. Here we model the pre-fit covariance matrix as

$$\Sigma_y = \langle yy^T \rangle = C^\text{red}_y + C^\text{EFAC}_y + C^\text{EQUAD}_y.$$  (1)

Details of the various components and transformations into a post-fit basis will be described elsewhere [2]. For this discussion it is only important to note that we maximize the likelihood function

$$L(\tilde{\theta}|r) = \frac{1}{\sqrt{\det(2\pi\Sigma_r)}} \exp\left(-\frac{1}{2}r^T\Sigma_r^{-1}r\right)$$  (2)

over the parameters $\tilde{\theta} = \{A, \gamma, F, Q\}$, where $A$ and $\gamma$ are the amplitude and spectral index of the red noise power spectrum and $F$ and $Q$ are the EFAC and EQUAD parameters, respectively. In the above expression, $\Sigma_r$ and $r$ are the post-fit covariance matrix and residuals, respectively.

We have analyzed each pulsar from each dataset and compiled the results in Figures 1–3. For this four parameter search, typical runtimes are on the order of 30 minutes per pulsar. When run in parallel using MPI, these runtimes decrease to a few minutes per pulsar. In these figures we plot the red noise amplitude in familiar GWB strain units vs. the power spectral index. For example, in these units a GWB from supermassive black hole binaries SMBHBs with a characteristic spectral index of -2/3 will have a power spectral index of 13/3. We also plot the EFAC parameter vs. the EQUAD parameter. In general if the white noise were completely described by the errorbars then the EFAC parameter is unity and the EQUAD parameter is 0. However, since these datasets have equal errorbars on all TOAs, EFAC and EQUAD are degenerate and may not result in the above situation. However, from these results we can conclude that there is no evidence for any white noise other than radiometer noise contained in the error bars. Red points indicate that the amplitude or EQUAD is consistent with 0 at the one sigma level and we simply plot upper limits on those parameters (red triangles). Here we notice that MDC 1 has strong red noise in all the residuals clustered around an Amplitude of $A \sim 1 \times 10^{-14}$ and a spectral index of $\gamma \sim 4.0$. MDC 2 is centered around a higher amplitude and similar spectral index but with a larger spread. In contrast, many of the pulsars in MDC 3 show little evidence for red noise, with quite a large spread in both spectral index and amplitude.
FIG. 1: Results of our noise estimation algorithm on individual dataset 1 pulsars. The left plot is the amplitude of a red noise signal converted to units of GWB strain vs. the power spectral index. The plot on the right is the EFAC parameter plotted against the EQUAD parameter in microseconds.

III. ANALYSIS METHODS

A. Optimal Statistic

The optimal statistic (OS) used for this work is based on that published in Anholm et al. \[1\] and will be further discussed in a future work. Here we will briefly review the algorithm. In general, we wish to construct a statistic that gives some measure of the significance of the expected (Hellings-Downs curve) correlations among pulsars giving larger weight to those residuals with lower noise levels. It is also useful to construct this statistic, such that, on average its value is the amplitude of the GWB for some given spectral index. It is shown in \[1\] that this statistic is both an optimal filter and maximizes the likelihood function in the low-signal limit. The optimal statistic can be written as

\[
\hat{\Omega} = \frac{\sum_{I,J} r_I^T P_I^{-1} S_{IJ} P_J^{-1} r_J}{\sum_{I,J} \text{tr} \left[ P_I^{-1} S_{IJ} P_J^{-1} S_{JI} \right]} \tag{3}
\]

with an SNR of

\[
\hat{\rho} = \frac{\sum_{I,J} r_I^T P_I^{-1} S_{IJ} P_J^{-1} r_J}{\sqrt{\sum_{I,J} \text{tr} \left[ P_I^{-1} S_{IJ} P_J^{-1} S_{JI} \right]}} \tag{4}
\]

where \(\sum_{I,J}\) denotes the sum over pulsar pairs, \(r_I, P_I,\) and \(S_{IJ}\) are the residuals for pulsar \(I,\) the auto-covariance matrix of pulsar \(I,\) and the amplitude free cross covariance matrix of pulsar pair \(IJ.\) We use the amplitude free cross covariance matrix to ensure that \(\langle \hat{\Omega} \rangle = \Omega_{\text{true}}\) under the assumption that \(\langle r_I r_J^T \rangle = \Omega_{\text{true}} S_{IJ}.\) The SNR used here gives a measure of how strong the correlated signal is compared to an uncorrelated signal of the same strength.
FIG. 2: Results of our noise estimation algorithm on individual dataset 2 pulsars. The left plot is the amplitude of a red noise signal converted to units of GWB strain vs. the power spectral index. Note that the red points show upper limits on the amplitude and EQUAD parameters respectively. This is done because these points are consistent with 0 at the 1-sigma level. Here, 32 of the 36 pulsars show evidence for red noise in the individual case indicating the presence of a strong red noise signal in nearly all pulsars. The plot on the right is the EFAC parameter plotted against the EQUAD parameter in microseconds.

For this analysis we use two versions of the optimal statistic: the zero-order and the iterative optimal statistic. The zero-order optimal statistic consists of estimating the maximum likelihood noise parameters, as discussed above, and forming the auto-covariance matrices as input into the code. In general, this method will give a biased result for the estimate of the GWB amplitude $\Omega$ because we are ignoring the presence of a common GWB signal in all pulsars. As we can see from Figures 1–3, there is quite a large spread on the red noise parameters and this large spread adds to the bias in the optimal statistic. However, if we iterate this procedure we can obtain a much better estimate of $\Omega$ and a higher SNR. The iterative procedure is as follows:

1. Run the zero-order likelihood with the auto-covariance matrices produced from the noise estimation procedure to obtain an estimate of the GWB amplitude $\Omega_0$ (here we assume a known spectral index in the construction of the cross-covariance matrices).

2. Now fix the spectral index for both the auto and cross covariance matrices and use the estimate $\Omega_0$ from the previous step to produce new auto-covariance matrices.

3. Run the optimal statistic code again, now with the new auto-covariance matrices and produce a new estimate of the amplitude $\Omega_i$.

4. Repeat these steps until the new estimate of the amplitude changes from the previous estimate by some amount $\varepsilon$.

It is important to note that for the iterative method to be valid we must have a priori knowledge about any red noise intrinsic to the pulsars. However, for the MDC datasets, no pulsars show evidence of intrinsic red noise, so it is a good
FIG. 3: Results of our noise estimation algorithm on individual dataset 3 pulsars. The left plot is the amplitude of a red noise signal converted to units of GWB strain vs. the power spectral index. Note that the red points show upper limits on the amplitude and EQUAD parameters respectively. This is done because these points are consistent with 0 at the 1-sigma level. Here, only 15 of the 36 pulsars show evidence for red noise in the individual case indicating that many pulsars are white noise dominated. The plot on the right is the EFAC parameter plotted against the EQUAD parameter in microseconds.

approximation to assume that the only red noise source is the GWB. Results for both the zero-order and iterative optimal statistics are shown in Table II and will be discussed in the final section.

B. First-Order Likelihood

The first-order likelihood combines elements of van Haasteren et al. [4] and [1] and will be published in a future paper. For a pulsar timing array with $M$ pulsars we define the probability distribution function of the presumed Gaussian noise as multivariate Gaussian [6]

$$ p(r) = \frac{1}{\sqrt{\det 2\pi \Sigma}} \exp \left( -\frac{1}{2} r^T \Sigma^{-1} r \right), $$

(5)

where

$$ r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_M \end{bmatrix} $$

(6)
is a vector of the residual time-series, \( r_\alpha(t) \), for all pulsars,

\[
\Sigma = \begin{bmatrix}
P_1 & S_{12} & \cdots & S_{1M} \\
S_{21} & P_2 & \cdots & S_{2M} \\
\vdots & \vdots & \ddots & \vdots \\
S_{M1} & S_{M2} & \cdots & P_M
\end{bmatrix}
\]  

(7)

is a multivariate block covariance matrix, and

\[
P_\alpha = \langle r_\alpha r_\alpha^T \rangle
\]

(8)

\[
S_{\alpha\beta} = \langle r_\alpha r_\beta^T \rangle |_{\alpha \neq \beta}
\]

(9)

are the auto-covariance and cross-covariance matrices, respectively, for each set of residuals.

Here we are interested in measuring the spectral index, \( \gamma_{\text{gw}} \), and amplitude, \( \Omega \), of the stochastic background. These parameters will be common among all pulsars, however, as mentioned above, each pulsar will have intrinsic noise parameters as well: an amplitude \( A_\alpha \) and spectral index \( \gamma_\alpha \) for a power law red noise process, and EFAC and EQUAD parameters, \( F_\alpha \) and \( Q_\alpha \), for white noise processes. Therefore, we write our auto-covariance as a sum of a common GWB term and a pulsar dependent term

\[
P_\alpha = R_\alpha + S_{a\alpha},
\]

(10)

where \( R_\alpha \) is the intrinsic noise auto-covariance matrix and \( S_{a\alpha} \) is the common GWB auto-covariance matrix. It is convenient to work in a block matrix notation where

\[
\Sigma = P + S_c,
\]

(11)

where \( P \) is a block diagonal matrix with diagonals \( P_\alpha \) and \( S_c \) is block matrix with off diagonals \( S_{\alpha\beta} \) and zero matrices on the diagonal.

In general, we write the log likelihood function as

\[
\ln L = -\frac{1}{2} \left[ \text{Tr} \ln \Sigma + r^T \Sigma^{-1} r \right],
\]

(12)

where we have used the fact that \( \ln \det(A) = \text{Tr} \ln(A) \). In practice the matrix \( \Sigma \) is quite large and therefore, computationally prohibitive to invert. Since many multi-frequency residual datasets now have on the order of \( 10^3 \) points, for many modern PTAs the matrix \( \Sigma \) will be of order \( 10^4 \times 10^4 \). To avoid inverting the full covariance matrix we can expand the inverse out in a Neumann series to obtain

\[
\Sigma^{-1} = P^{-1} - P^{-1} S_c P^{-1} + O(\epsilon^2).
\]

(13)

FIG. 4: (from left to right) Results from our first-order likelihood method run on the “best 12” pulsars for closed mock datasets 1, 2 and 3. Here the solid, dashed and dash-dot lines are the 68%, 95%, and 99% contours, respectively. The points with errorbars are the estimates of the amplitude obtained from the iterative optimal statistic with 1-sigma uncertainties.
where $\epsilon$ is an order parameter. Physically, $\epsilon$ can represent the amplitude of the GWB, $\Omega$, since it is a small constant multiplier of the elements of $\mathbf{S}$. Since the diagonal elements of $\mathbf{\Sigma}$ will always be less than the diagonal terms, we can keep only up to first order in $\epsilon$. The determinant can also be approximated (to first order in $\epsilon$) as

$$\ln \det \mathbf{\Sigma} = \text{Tr} \ln \mathbf{P} + \mathcal{O}(\epsilon^2).$$

With these approximations, it is now possible to write the approximate log-likelihood

$$\ln \mathcal{L} = -\frac{1}{2} \left[ \text{Tr} \ln \mathbf{P} + \mathbf{r}^T \mathbf{P}^{-1} \mathbf{r} - \mathbf{r}^T \mathbf{P}^{-1} \mathbf{S} \mathbf{P}^{-1} \mathbf{r} \right]$$

$$= -\frac{1}{2} \sum_i \left[ \text{Tr} \ln P_i + r_i^T P_i^{-1} r_i \right] + \Omega \sum_{i,j} r_i^T P_i^{-1} S_{ij} P_j^{-1} r_j,$$

where again, we have used the notation that $\sum_{i,j}$ denotes a sum of all unique pulsar pairs and $S_{ij}$ is the amplitude-free cross covariance matrix. The results of evaluating the first order likelihood for MDC’s 1, 2, and 3 are shown in Figure 4 and will be discussed in the next section.

IV. EXECUTIVE SUMMARY

Here we will discuss our results in detail and use this information to make a decision as to what GWB signals are present in the data. As mentioned above, first we ran the noise estimation analysis on all pulsars of all datasets. For the purposes here we do not directly use the red noise information obtained from this search in subsequent analyses, however, we do record the white noise levels for all pulsars and find that the errorbars give a good estimate of the total white noise level. We will use these white noise values throughout our GWB analyses as the white noise components are not correlated with red noise components, thus, the measurements of the white noise are independent of red noise model assuming the data is well described by a two component noise model.

We have shown above that the first-order likelihood is designed for signals that are noise dominated. However, this is not the case for most residuals in the three datasets. For both MDC 2 and MDC 3 we have split the datasets up into thirds sorted by the white noise level. Using the open datasets and the noise estimation results as a gauge, we have found that the best 12 pulsars (sorted by the white noise level) is sufficiently noise dominated (across the 12 pulsar span) that the first-order likelihood performs well with no discernible biases. We have not used all 36 pulsars because we see biases in our results in the open datasets that we do not currently understand when using all 36 pulsars.

However, for MDC 1 all of the pulsars are in the large signal limit thus we simply choose a random group of 12 pulsars just to be consistent with the searches carried out for MDC’s 2 and 3. We also use 12 pulsars in an attempt to keep the number of search parameters to a minimum (26 parameters in the case of 12 pulsars) and we have found that our analysis method using MultiNest is slowly convergent in large dimensional parameter spaces and in fact becomes computationally prohibitive in the 36 pulsar case. Other samplers are being explored but were not tested in time for this challenge.

With these datasets now chosen we have run our first-order likelihood assuming three different models:

1. Stochastic GWB with amplitude $A$, spectral index $\gamma_{gw}$ with Hellings-Downs correlation coefficients plus intrinsic red noise with a power law red noise with amplitude $A_i$ and spectral index $\gamma_i$ and white noise given by the noise estimation estimates. We use flat priors in the GW amplitude and spectral index parameters and the noise amplitude and spectral index with $A \in (0, 5 \times 10^{-13})$, $\gamma_{gw} \in [1, 7]$, $A_i \in (0, 5 \times 10^{-15})$, and $\gamma_i \in [1, 7]$. This results in a 2 + 2M dimensional search, where $M$ is the number of pulsars used (two parameters for the GWB and 2M for the intrinsic noise parameters).

2. Null hypothesis where all red noise processes are intrinsic. Again we assume an intrinsic amplitude $A_i$ and spectral index $\gamma_i$ with noise estimated white noise parameters and no GWB signal. This results in a 2M parameter search.

3. Common red noise signal with amplitude $B_{\text{common}}$ and spectral index $\gamma_{\text{common}}$ but no spatial correlations and intrinsic red noise parameters as described above. Again, this results in a 2 + 2M parameter search.

For each model we evaluate the Bayesian evidence and then compute the Bayes factors $B_0$ and $B_{\text{common}}$, which compare the GWB hypothesis to the null hypothesis and uncorrelated common red noise hypothesis, respectively. All of these searches were run on the Nemo Computing Cluster at UWM using the MultiNest [3] algorithm. Typical runtimes using 20 CPUs for each dataset (12 pulsars) are around 2 hours.
TABLE I: “best 12” pulsars for each dataset.

| MDC1       | MDC2       | MDC3       |
|------------|------------|------------|
| J0030+0451 | J1939+2134 | J1939+2134 |
| J1738+0333 | J0437-4715 | J0437-4715 |
| J1741+1351 | J1713+0747 | J1713+0747 |
| J1744-1134 | J1909-3744 | J1909-3744 |
| J1751-2857 | J1744-1134 | J1744-1134 |
| J1853+1303 | J1910+1256 | J1910+1256 |
| J1857+0943 | J1853+1303 | J1853+1303 |
| J1732-5049 | J1955+2908 | J1955+2908 |
| J1909-3744 | J1741+1351 | J1741+1351 |
| J1918-0642 | J1640+2224 | J1640+2224 |
| J1939+2134 | J1600-3053 | J1600-3053 |
| J1955+2908 | J1955+2908 | J1955+2908 |

TABLE II: Table of results from many analysis methods. Here we have estimates of the GWB amplitude $A$ and the corresponding SNR using the zero-order optimal statistic and iterative optimal statistic methods described above. The uncertainty on $A$ represent the 1-$\sigma$ confidence levels. We have also listed the results of our first-order likelihood method, also described above.

| Zero Order OS$^a$ | Iterative OS | First-Order Likelihood$^b$ |
|-------------------|--------------|---------------------------|
| $A \times 10^{-14}$ | SNR          | $A \times 10^{-14}$ | $\gamma$ | $\ln B_0$ | $\ln B_{\text{common}}$ |
|-------------------|--------------|-------------------------|----------|------------|---------------------------|
| MDC 1             | 0.45         | 4.3                     | 0.92 ± 0.2 | 9.2         | $1.9_{-0.16}^{+0.19}$ | $4.26_{-0.27}^{+0.22}$ | 33.5 | 16.8 |
| MDC 2             | 1.8          | 2.0                     | 4.3 ± 1.4 | 9.5         | $6.87_{-0.65}^{+0.40}$ | $4.16_{-0.15}^{+0.18}$ | 153.4 | 10.5 |
| MDC 3             | 0.15         | 1.2                     | 0.63 ± 0.38 | 5.5         | $0.88_{-0.22}^{+0.20}$ | $3.58_{-0.44}^{+0.48}$ | 31.6 | 20.6 |

$^a$ An estimate of the amplitude is required to fully calculate the uncertainty, thus we do not include it here.

$^b$ Note that these results are only for the “best 12” pulsars.

The results of the first order likelihood run on the “best 12” pulsars is shown in Table I and the list of pulsars used in each dataset is shown in Table I. Here we quote the maximum of the marginalized posterior distribution for both the amplitude (written in characteristic strain units) and spectral index as well as the 68% credible regions and the logarithm of the two Bayes factors, $B_0$ and $B_{\text{common}}$. Firstly, we note that all three datasets (using the “best 12” pulsars) have large Bayes factors in favor of a GWB model over a model with just intrinsic red noise. We also note that all 3 datasets also strongly favor a GWB model over an uncorrelated common red noise signal. As we will note shortly, we can use the optimal statistic in combination with the first-order likelihood to make a model decision in situations where the evidence for a particular model is weak. We also see that the most likely spectral index is consistent with $13/3$ (SMBHB spectrum) for all three datasets. We also note that there was no evidence for red noise in MDC1 or MDC2, however, we did find evidence for intrinsic red noise in closed dataset 3. Although there is evidence for red noise, we cannot accurately characterize it in many cases. For PSRs J1939+2134, J0437-4715 and J1955+2908, the 1-sigma contours do are not consistent with red noise of 0 amplitude, however for all others we are unable to characterize the red noise. Contour plots of the red noise parameters for MDC3 are shown in Figure 5.

We have also run both the zero-order and iterative optimal statistics on the full datasets. Typical run times are on the order of 5 minutes on a modern workstation. As mentioned above, for the zero order OS we use the noise estimation results shown in Figures 1–3 to construct the auto-covariance matrices. The amplitude (written in characteristic strain units) and corresponding SNR are shown in the second and third columns of Table I. We see that there is a significant detection for MDC 1 and marginal detections for MDC’s 2 and 3 using this method. Intuitively, this makes sense because there is more white noise in pulsars from MDC’s 2 and 3 and thus a larger spread in the estimated amplitudes and spectral indices of the GWB (based on the first order likelihood results we assume that any measured red noise is due to the GWB) and thus a more biased result from the optimal statistic. When we now use...
FIG. 5: Noise contours for the “best 12” pulsars of MDC3. The x-axis is the amplitude measured in GW strain units normalized by $10^{-14}$ and the y-axis is the power spectral index. The solid, dashed, and dot-dashed lines represent the 1, 2, and 3 sigma confidence contours, respectively.

the iterative method we make confident detections for all three datasets and recover amplitudes that are consistent with the first-order likelihood at the 95% level or better.

[1] Anholm, M., Ballmer, S., Creighton, J. D. E., Price, L. R., & Siemens, X. 2009, Phys. Rev. D, 79, 084030
[2] Ellis, J. A., et al. 2012, in preparation
[3] Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
[4] van Haasteren, R., Levin, Y., McDonald, P., & Lu, T. 2009, MNRAS, 395, 1005
[5] For this analysis, $\varepsilon$ was chosen to be 0.005.
[6] Here we will assume that all noise processes are stochastic and that there are no deterministic signals present in the data
[7] Here we use a red noise amplitude that is analogous to the traditional GW strain amplitude
[8] Remember, a Bayes factor $> 10$ is strong evidence against the competing hypothesis.