



Applying clock comparison methods to pulsar timing observations

Siyuan Chen, François Vernotte, Enrico Rubiola

► To cite this version:

Siyuan Chen, François Vernotte, Enrico Rubiola. Applying clock comparison methods to pulsar timing observations. Monthly Notices of the Royal Astronomical Society, 2021, 503 (3), pp.4496-4507. 10.1093/mnras/stab742 . hal-03022716

HAL Id: hal-03022716

<https://hal.science/hal-03022716>

Submitted on 22 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Applying clock comparison methods to pulsar timing observations

Siyuan Chen¹,^{1,2,3}★ François Vernotte^{2,4} and Enrico Rubiola^{2,4,5}

¹Station de Radioastronomie de Nançay, Observatoire de Paris, PSL University, CNRS, Université d'Orléans, F-18330 Nançay, France

²FEMTO-ST, Department of Time and Frequency, UBFC and CNRS, F-25030 Besançon, France

³Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, LPC2E UMR7328, Université d'Orléans, CNRS, F-45071 Orléans, France

⁴Observatoire des Sciences de l'Univers THETA, UBFC and CNRS, F-20510 Besançon, France

⁵Division of Quantum Metrology and Nanotechnology, Istituto Nazionale di Ricerca Metrologica INRiM, I-10135 Torino, Italy

Accepted 2021 March 9. Received 2021 February 24; in original form 2020 March 20

ABSTRACT

Frequency metrology outperforms any other branch of metrology in accuracy (parts in 10^{-16}) and small fluctuations ($<10^{-17}$). In turn, among celestial bodies, the rotation speed of millisecond pulsars is by far the most stable ($<10^{-18}$). Therefore, the precise measurement of the time of arrival (TOA) of pulsar signals is expected to disclose information about cosmological phenomena, and to enlarge our astrophysical knowledge. Related to this topic, Pulsar Timing Array projects have been developed and operated for the last decades. The TOAs from a pulsar can be affected by local emission and environmental effects, in the direction of the propagation through the interstellar medium or universally by gravitational waves from super massive black hole binaries. These effects (signals) can manifest as a low-frequency fluctuation over time, phenomenologically similar to a red noise, while the remaining pulsar intrinsic and instrumental background (noise) are white. This article focuses on the frequency metrology of pulsars. From our standpoint, the pulsar is an accurate clock, to be measured simultaneously with several telescopes in order to reject the uncorrelated white noise. We apply the modern statistical methods of time-and-frequency metrology to simulated pulsar data, and we show the detection limit of the correlated red noise signal between telescopes.

Key words: gravitational waves – methods: data analysis – pulsars: general.

1 INTRODUCTION

Millisecond pulsars (MSP) are considered extremely stable astronomical clocks because of their high energy and momentum in a small size (Verbiest et al. 2009), albeit the observations can be affected by large observational white noise. Such noise can be due to the low signal-to-noise ratio of the MSP observations, and depends on the radio telescope (RT). On the other hand, red noise could originate from the MSP, the propagation through the interstellar medium or from gravitational waves (GWs) on the line of sight, hence it is common to all the RTs (Hellings & Downs 1983; Jenet & Romano 2015). The characterization of the spectral signature of GWs is a major scientific challenge because it is expected to disclose information about the astrophysical sources. As pulsars have been precisely timed for decades with up to weekly cadence, the frequency range that can be probed by Pulsar Timing Array (PTA; Foster & Backer 1990) projects is $f \in [10^{-9}, 10^{-6}]$ Hz, which is a window out of reach for the LIGO/VIRGO interferometers and at the edge of the LISA band. The most interesting and likely to be detected source is a cosmic population of super massive black hole binaries (Taylor et al. 2016 and Perrodin & Sesana 2018 for a recent review), whose interaction with the MSP signals introduces a correlated red noise in the time of arrival (TOA) series, with a phase power spectral density (PSD) proportional to $1/f^{4/3}$ (for a circular population, see

e.g. Phinney 2001; Chen, Sesana & Del Pozzo 2017). The choice of the algorithm is therefore of paramount importance to disentangle the pulsar specific small red noise (signal) from the effects of GWs in the presence of large white noise (background) in the shortest observation time. In this regard, PTAs (Desvignes et al. 2016; Perera et al. 2019; Alam et al. 2020; Kerr et al. 2020) enable the simultaneous measurement of the same pulsar with several RTs, especially the Large European Array for Pulsars (LEAP) project (Bassa et al. 2016).

The extraction of small signals from noise exploiting simultaneous measurements is a well-known method for the measurement of oscillators and atomic clocks. Specific tools have been developed over more than 50 yr using the PSD and wavelet variances (Barnes et al. 1971; Gray & Allan 1974; Rutman 1978; Allan & Barnes 1981). Among them, the cross-spectrum (CS) method (Rubiola & Vernotte 2010) deserves separate mention in view of our application. Under certain assumptions, this method rejects the background noise (uncorrelated), and converges to the oscillator noise even if it is significantly lower than the background. Wavelet variances are commonly used in the time domain, the most known of which is the 2-sample Allan variance (AVAR), after the pioneering work of Allan and Barnes (Allan 1966). Such variances enable to distinguish different types of noise defined by their exponent in the PSD, e.g. f^0 for white phase, $1/f$ for flicker phase modulation, $1/f^2$ for white frequency modulation, $1/f^3$ for flicker frequency modulation, $1/f^4$ for random walk frequency modulation noise, etc. The parabolic variance (PVAR; Benkler, Lisdat & Sterr 2015; Vernotte et al.

★ E-mail: siyuan.chen@cnrs-orleans.fr

2016) extends this concept, and exhibits (i) the highest rejection of white noise and (ii) the efficient detection red noise underneath the background with the shortest data record. Moreover, the wavelet covariance (Fest, Gros Lambert & Gagnepain 1983; Lantz et al. 2019), i.e. Allan covariance (ACOV) or parabolic covariance (PCOV), enables the rejection of the background using two (or more) uncorrelated instruments.

This article is intended to port the time-and-frequency metrology methods to pulsar astronomy. In this respect, the pulsar is the clock under test, observed simultaneously with two or more RTs playing the role of phase meters. We compare the results of different methods applied to simulated time series representing the TOAs of MSPs. In Section 1, we summarize the concept of pulsar timing as well as the spectral and variance methods from the time-and-frequency metrology. Section 2 describes the simulation process of the TOA time series, followed by the statistical methods in Section 3. Results and the conclusions are presented in Sections 4 and 5, respectively.

1.1 Pulsar timing

Pulsars are spinning neutron stars that emit radio waves from the region above their magnetic poles. If these regions are misaligned with the rotation axis and happen to point towards the Earth as they sweep space, we receive one radio pulse each rotation. The most stable pulsars have rotation period of milliseconds and are stable on a very long time-scale. Over the decades that they have been timed, very small variations have been detected, which translate into low frequency red noise in the Fourier domain. A detailed review about pulsars can be found in Lorimer & Kramer (2012).

In general, the TOA of the pulses can be described precisely by the following timing model (Hobbs, Edwards & Manchester 2006):

$$t_{\text{obs}} = t_{\text{model}} + t_{\text{red}} + t_{\text{white}}, \quad (1)$$

where t_{obs} is the observed and t_{model} is the predicted TOA considering all known pulsar properties and propagation effects. For this study, we also include the effects of dispersion measure variation over time into the perfect model. Ideally, we should be able to precisely predict the TOAs if all model parameters are perfectly known and there are no unknown sources of noise left. However, this is not the case in practice. The difference between the observed and model observation forms the residual series $x = t_{\text{red}} + t_{\text{white}}$. Therefore, we are interested in both white and red noise components.

The pulse residual series can be transformed into Fourier frequency space. The noise components are then described by a PSD. The PTA noise description can be found in Lentati et al. (2016), Caballero et al. (2016), and Arzoumanian et al. (2016). Following the notation of the latter, the white noise is described by two parameters E_k and Q_k :

$$S_w = (E_k W)^2 + Q_k^2, \quad (2)$$

where W is the initial estimate of the white coming from the radiometer noise and template matching error and E_k and Q_k are the so-called EFAC and EQUAD parameters, respectively. Note that S_w is constant over the whole frequency range.

The red noise is described by a power law with two parameters

$$S_r(f) = A_{\text{PTA}}^2 f^{-\gamma_{\text{PTA}}}, \quad (3)$$

where A_{PTA} is the amplitude and γ_{PTA} is the spectral index of the power law. The total noise PSD $S(f)$ becomes (Vernotte & Zalamansky 1999)

$$S(f) = S_r(f) + S_w. \quad (4)$$

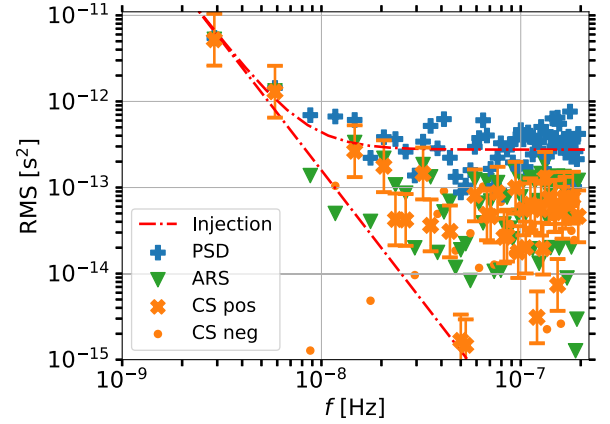


Figure 1. Average PSD (the blue pluses), averaged residual spectrum (ARS; the green triangles), and CS (the orange crosses/points) computed from one set of five different residual series in the boundary case (the red-dashed lines).

1.2 Power and cross-spectrum

In time and frequency metrology, the CS is the most often used method to measure the phase noise of oscillators. The method relies on the simultaneous measurement with two separate instruments, assuming that the background noise is statistically independent. The same hypothesis holds for pulsar observations from different radio telescopes. The telescopes are obviously independent and sufficiently far from each other for the tropospheric/ionospheric noise to be statistically independent. We summarize the main results from our tutorial on the CS method (Rubiola & Vernotte 2010).

Given a residual series x from an instrument, we can write its Fourier transform as $X = \Re(X) + i \Im(X)$. The one-sided PSD can be calculated as

$$S_x(f) = \frac{2}{T_a} X X^* = \frac{2}{T_a} (|\Re(X)|^2 + |\Im(X)|^2) \quad \text{for } f > 0, \quad (5)$$

where T_a is the acquisition time, and the factor 2 is due to energy conservation after suppressing the negative frequencies.

Measuring a clock (pulsar) signal c simultaneously with two instruments (telescopes), we get two residual series x and y , each containing the white background noise of the measurement added to c . The CS

$$S_{yx} = \frac{2}{T_a} Y X^* \quad (6)$$

converges to $S_c = \frac{2}{T_a} C C^*$. Notice that, after compensating for the differential path, c goes only into $\Re(S_{yx})$, while the $\Im(S_{yx})$ contains only the background noise of the observation. After Rubiola & Vernotte (2010), the fastest, unbiased estimator is

$$\hat{S}_c = \frac{2}{T_a} \Re(S_{yx}) = \frac{2}{T_a} (\Re(X) \Re(Y) + \Im(X) \Im(Y)) \quad (7)$$

and can give negative values. We average over all the CSs coming from the different RT pairs to produce one simultaneous observation series for the analysis.

Fig. 1 shows an example of a comparison between the different spectra. The PSD can be computed on the averaged residual series (ARS) or on each residual series individually, and then averaged (PSD). As the ARS is very similar, but slightly worse than the CS, we opt to focus the analysis on the CS method, whose results should be comparable to those from the ARS. The CS is better at rejecting

uncorrelated white noise and produces a lower level than the average PSD, which serves as a good reference.

We also investigated the effects of the choice of the window function on the Fourier transform by comparing the sinc function used above with a Blackman window. Although there is spectral leakage of the red noise, the amount is only noticeable in the lowest frequencies and is within the uncertainty of the PSD/CS value. We found minimal effects on our analysis. In fact, as the spectral value decreases, the recovered red noise parameters provide a worse match to the injected values. We also found similar results with a pre-whitening/ post-darkening procedure. Therefore, we show only the results obtained with the above equations.

1.3 Variances

Different variances have been proposed to describe the amount of variation in a given residual series x with a total time span T for time-steps τ . The two prominent ones are the Allan (Allan 1966) and more recently the PVAR (Vernotte et al. 2016), which we summarize here.

The AVAR is a simple computation comparing three observations separated by m time-steps

$$\text{AVAR}(\tau) = \frac{1}{2M\tau^2} \sum_{i=0}^{M-1} [(-x_i + 2x_{i+m} - x_{i+2m})]^2, \quad (8)$$

where $M = N - 2m + 1$ is related to the total number N of data points in the residual series x . Compared to similar variances, AVAR is the most efficient at estimating for the largest $\tau = T/2$. By contrast, the Modified AVAR is more suitable to the analysis of short-term fluctuations (Allan & Barnes 1981). The PVAR combines the original long-term computability as well as a good response to short time-scales

$$\text{PVAR}(\tau) = \frac{72}{Mm^4\tau^2} \sum_{i=0}^{M-1} \left[\sum_{k=0}^{m-1} \left(\frac{m-1-2k}{2} \right) (x_{i+k} - x_{i+m+k}) \right]^2, \quad (9)$$

where $M = N - 2m + 2$.

Similar to the CS method, we can also use two independent residual series x and y to compute the ACOV:

$$\text{ACOV}(\tau) = \frac{1}{2M\tau^2} \sum_{i=0}^{M-1} [(-x_i + 2x_{i+m} - x_{i+2m}) \times (-y_i + 2y_{i+m} - y_{i+2m})] \quad (10)$$

and the PCOV

$$\text{PCOV}(\tau) = \frac{72}{Mm^4\tau^2} \sum_{i=0}^{M-1} \left[\sum_{k=0}^{m-1} \left(\frac{m-1-2k}{2} \right) (x_{i+k} - x_{i+m+k}) \times \left(\frac{m-1-2k}{2} \right) (y_{i+k} - y_{i+m+k}) \right]. \quad (11)$$

A comparison of the two different (co)variances can be found in Fig. 2. In general, the variance plot has two prominent features: (i) at lower time-steps, the variance is dominated by white noise, which leads to a power-law decrease with fixed spectral index, (ii) at higher time-steps, red noise start to appear, which shows up as a flattening and then rising of the variance, depending on the amplitude and spectral index of the red noise. The variance curve can be parametrized as

$$\text{VAR} = A_{\text{VAR}}^2 \tau^{\gamma_{\text{VAR}}} + B_{\text{VAR}}^2 \tau^{\beta_{\text{VAR}}}. \quad (12)$$

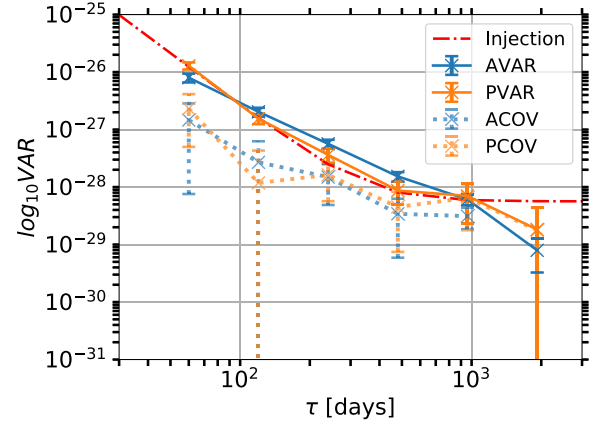


Figure 2. Averaged data Allan variance (blue) and PVAR (orange), the corresponding covariances (the light-dotted lines) and the associated uncertainties computed from one set of five residual series in the boundary case (the red-dashed line).

The relation between the PTA parameters EFAC, EQUAD, γ_{PTA} and A_{PTA} and the variance parameters is not trivial. As the white noise follows a fixed theoretical spectral index for a given variance $\beta_{\text{VAR}} = -2/-3$ for AVAR and PVAR, respectively, we can compare the overall amplitude of the variance white noise with S_w from the PTA white noise in (2)

$$S_w = \begin{cases} 2B_{\text{VAR}}^2/3 & \text{for AVAR} \\ B_{\text{VAR}}^2/(6\tau_0) & \text{for PVAR,} \end{cases} \quad (13)$$

where τ_0 is the sampling time.

The two red noise spectral indices can be related via

$$\gamma_{\text{PTA}} = \gamma_{\text{VAR}} + 3. \quad (14)$$

The red noise amplitudes follow analytic relations for given integer spectral indices (see table 1 in Vernotte et al. 2016). Under the condition that $\gamma_{\text{PTA}} \in [1.5, 4.5]$, it is also possible to find an analytic relation between A_{VAR} and A_{PTA} for non-integer γ_{PTA} , see also Vernotte, Rubiola & Chen (2020):

$$A_{\text{VAR}}^2 = \frac{A_{\text{PTA}}^2}{12\pi^2} 4\pi^2 y r^{3+\gamma_{\text{PTA}}} \mathcal{A}(\gamma_{\text{PTA}}), \quad (15)$$

where $\mathcal{A}(\gamma_{\text{PTA}})$ is a function that relates to the coefficients of the variance. Using the same approach as Walter (1994), we find the following relationships for AVAR and PVAR:

$$\mathcal{A}(\gamma_{\text{PTA}}) = \begin{cases} (4^{\gamma_{\text{PTA}}-2} - 2^{\gamma_{\text{PTA}}-1}) \pi^{\gamma_{\text{PTA}}-3} & \\ \Gamma(1 - \gamma_{\text{PTA}}) \sin(\pi - \pi \gamma_{\text{PTA}}/2) & \text{for AVAR} \\ -4^{\gamma_{\text{PTA}}} 9\pi^{\gamma_{\text{PTA}}-3} & \\ \times (2^{2-\gamma_{\text{PTA}}} (-1 - \gamma_{\text{PTA}}) & \\ + 2 + 3\gamma_{\text{PTA}} - \gamma_{\text{PTA}}^2) & \\ \Gamma(-3 - \gamma_{\text{PTA}}) \sin(\pi - \pi \gamma_{\text{PTA}}/2) & \text{for PVAR.} \end{cases} \quad (16)$$

2 SIMULATED DATA SETS

Clock comparison methods require several time series of the same pulsar observed simultaneously with different telescopes over a long time span. The LEAP project uses the five major radio telescopes

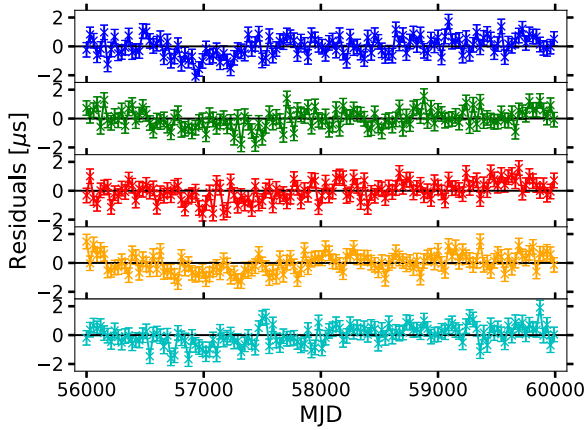


Figure 3. Example set of five different simulated residual series of a 4000-d long observing campaign with a monthly cadence in the boundary case.

in Europe: Effelsberg in Germany, Lovell in UK, Westerbork in the Netherlands, Nançay in France and Sardinia in Italy as an interferometer with monthly simultaneous observations since 2012 (Bassa et al. 2016; Smits et al. 2017).

Thus, we simulate 100 realizations of five TOA series over a 4000-d long observing campaign with a cadence of 30 d, this is of the order of the nominal LEAP observations up to today. They have been created using the `libstempo` package (Vallisneri 2015). For simplicity, we assume that all five RTs have comparable instrumental white noise, and give to all residuals the initial uncertainty $W = 500$ ns. As the two white noise parameters are degenerate, we will only focus on the dominant term, i.e. EFAC, injecting different realizations of white noise with (EFAC = 1.05, EQUAD = 0).

We also simulate the same residual series as it would be observed by LEAP under ideal conditions. In the LEAP statistical limit (LSL) of the coherently added mode, the total S/N of the LEAP observation is the sum of the individual RTs. Since we assume that the five RTs are identical, the LEAP observations have a five times higher S/N with the TOAs having a white noise level that is five times lower as $S/N \propto 1/W$. This also translates into an EFAC injection of $1.05/5 = 0.21$, while EQUAD remains at 0.

In total, we consider three cases:

- (i) white noise case: no red noise signal injected
- (ii) boundary case: ($\log_{10} A_{\text{PTA}} = -14$, $\gamma_{\text{PTA}} = 3$)
- (iii) red noise case: ($\log_{10} A_{\text{PTA}} = -13$, $\gamma_{\text{PTA}} = 3$)

For each case, we simulate 100 sets of five residual series with different red noise realization between each set. For a given set, the red noise realization is the same amongst the five residual series, but the white noise realizations differ. An example of a set of five simulated residual series can be seen in Fig. 3. As we are focusing only on the noise properties, we fix the timing model at its injected values. This means that the residual series are fully determined by the injected white and red noise. In practice, there will be covariances between the timing model and the noise parameters. Here, we only show the theoretically optimal results as a proof of concept.

3 METHODS

The standard PTA analysis is done via the `enterprise` package (Ellis et al. 2017), details of the analysis can be found in Arzoumanian et al. (2016, 2018). We use all five individual RT residual series for

the PTA analysis comparison. For the LSL analysis, we only have one residual series, but with a lower white noise level.

We compute the PSD and CS through equations (5) and (7), see Fig. 1, and compare them directly to equation (4) to get constraints on the noise parameters. The PSD can be averaged over the five residual series, whilst the CS can be computed from the average of the 10 different combinations of two of the five residual series.

The Allan and PVARs are calculated through equations (8) and (9), see Fig. 2, and matched to the parametrization in equation (12) with the use of equations (13), (14), and (15). Unlike the PSD, we use the ARS to compute the variances.

We use Bayesian statistics and a nested sampling algorithm, implemented in the `cpnest` package (Del Pozzo & Veitch 2015), to look for the posterior constraints on the white noise properties as well as the spectral index and amplitude of the red noise parameters.

A likelihood function can be derived from a weighted least-squares fitting procedure:

$$\mathcal{L} = -\frac{1}{2\sigma_i^2} (\mathcal{D} - \mathcal{F}(\phi))^2, \quad (17)$$

where σ_i^2 are the weights, \mathcal{D} are the computed data points, and $\mathcal{F}(\phi)$ are the corresponding parametrizations with the model parameters $\phi = (\text{EFAC}, \log_{10} \text{EQUAD}, \gamma_{\text{PTA}}, \log_{10} A_{\text{PTA}})$.

For the spectral analysis, we work in the logarithmic space, where \mathcal{D} corresponds to the computed \log_{10} PSD/CS values. From a large Monte Carlo simulation, we infer the uncertainties to be $\sigma_i = 0.2136$ across all frequencies in the Gaussian approximation using only the positive estimates, see appendix A.

For the (co)variance analysis, where \mathcal{D} are the computed (co)variances, we compute the weights σ_i^2 from the effective degrees of freedom for each time-step. A more detailed description can be found in Appendix B.

The choice of priors follows closely the standard PTA analysis (e.g. Lentati et al. 2015) wherever possible with uniform distributions in $\text{EFAC} \in [0.1, 1.5]$, $\log_{10} \text{EQUAD} \in [-10, -5]$, $\gamma_{\text{PTA}} \in [0, 7]$, and $\log_{10} A_{\text{PTA}} \in [-20, -10]$. The only exception are the boundaries of $\gamma_{\text{PTA}} \in [1.5, 4.5]$ for the variance analysis. This is due to the limited range of validity of equation (16; Verotte et al. 2020).

4 RESULTS

To ensure that our results are statistically robust, each of the three cases is based on 100 different realization sets and doing the analysis using all the clock comparison methods and the standard PTA/LSL comparisons. All following posterior distributions and recoveries are based on the addition of all 100 individual posterior distributions, unless stated otherwise. They should thus be representative of the constraints from the analyses of a single simulated set of five residual series.

4.1 White noise

We first explore the performance of the clock methods in the case of pure white noise in the residual series, and compare it to the traditional PTA/LSL analysis. Fig. 4 shows the PTA noise parameter marginalized 1D and 2D posterior distributions in a corner plot. As there is no red noise present, A_{PTA} and γ_{PTA} return very uninformative posteriors for all methods. γ_{PTA} simply returns a nearly flat uniform prior. The quantity A_{PTA} returns a posterior distribution that is flat for low amplitudes, but has a cut-off at an higher amplitude. This can be interpreted as a upper limit case, as red noise with a higher

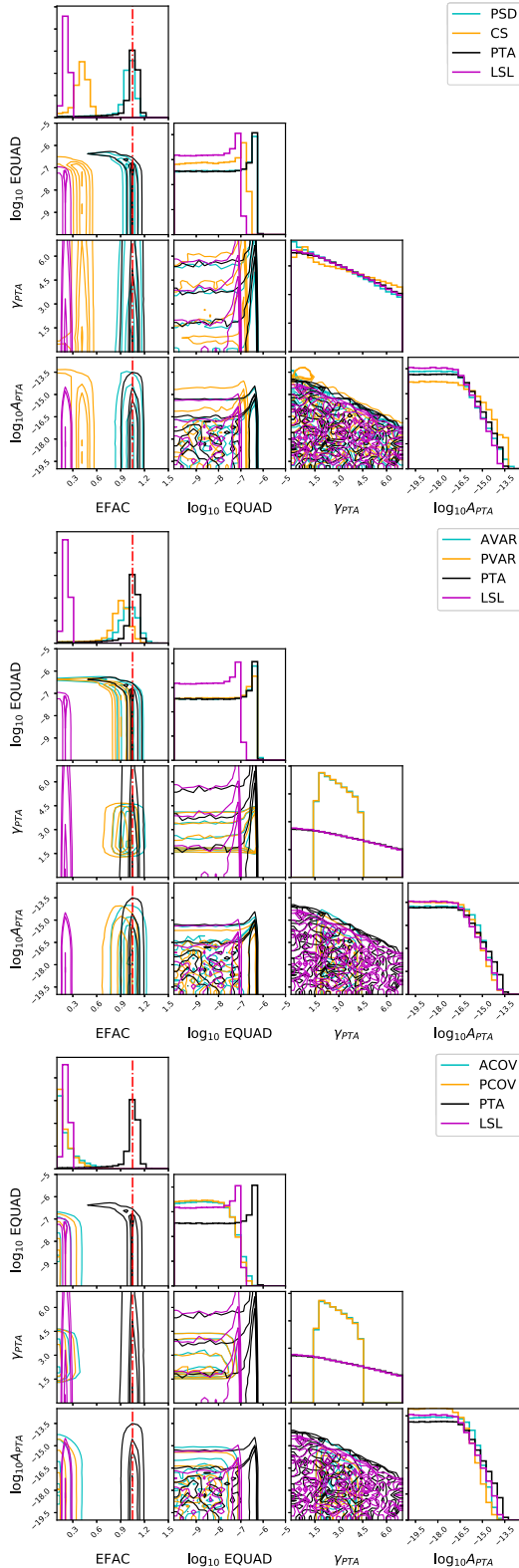


Figure 4. Comparison of the posterior distribution corner plots for the white noise case: spectral (top), variance (middle), and covariance analyses (bottom).

amplitude should have affected the residual series. Similarly, as there is no EQUAD injected, all analyses also show upper limits.

The only parameter of interest in the case of white noise is EFAC. As expected, both the PSD and PTA methods return distributions around the injected value, with the PTA distribution being more constraining than the PSD. The CS method does not seem to lower the white noise to the same level as the LSL, see top panel of Fig. 4. However, this is misleading, as the white noise is constrained from the positive CS values only. In fact, the CS method lowers the white noise to almost zero and could perform similarly or better to the LSL residuals by also taking the negative CS values into account. This would, however, come at the cost of a decreased performance in the recovery of the red noise parameters. In the middle panel of Fig. 4, both AVAR and PVAR show similar, but worse, constraints than the PTA analysis, where AVAR is closer to the injected value and PVAR has a small bias towards lower values. The bottom panel shows that the covariances reject most of the white noise, pushing the EFAC towards the lower end of the allowed values. This effect is intended as the covariances ‘try’ to find a correlated signal, but as there is only uncorrelated noise a upper limit on EFAC is placed.

However, we track the evolution of the performance of the different methods to recover white noise as the amount of red noise gradually increases as we move through the three simulated cases. The main result in Fig. 5 is that as the red noise increases, all methods seem to progressively perform worse in constraining the white noise. For the spectral and variance analysis, the posterior distributions progressively widen. Particularly in the strong red noise case, the AVAR and PVAR analysis are completely unable to recover the injected white noise (bottom middle panel). For the covariance analysis, the EFAC upper limit gradually increases, most notable in the strong red noise case.

A different visual aid is given in Fig. 6, which shows the constraints on EFAC from the individual analyses of a subset of 10 sets of realizations.

4.2 Boundary

A summary of the constraints on the PTA parameters from the standard PTA/LSL analysis and the clock comparison methods is given in Table 1. The posterior distributions for all analysis methods with the recovered spectra and variances can be found in Fig. 7. Examples from analyses of a single simulation set are shown in Fig. 8. Despite no injection of a EQUAD value, we still perform the all analyses with all four PTA noise parameters. However, the recovered EQUAD posterior distributions are all upper limits as in the white noise case above. Thus, we will omit the EQUAD parameter from our discussion from here on, as the focus will be on the red noise parameters.

4.2.1 Spectral analysis

The comparison of the marginalized 1D and 2D posterior distributions of the PTA noise parameter for the boundary case from the spectral analysis can be found in the top left corner plot of Fig. 7. The middle and bottom left-hand panels in Fig. 7 show the injected red and white noise along with the recovered PSD and CS, respectively.

The red noise signal is not well detectable for the PSD as only the lowest frequency bin in the left middle panel of Fig. 7 shows signs of deviating away from white noise. There is a small hint of the red noise. Which is sufficient to put loose constraints on the spectral

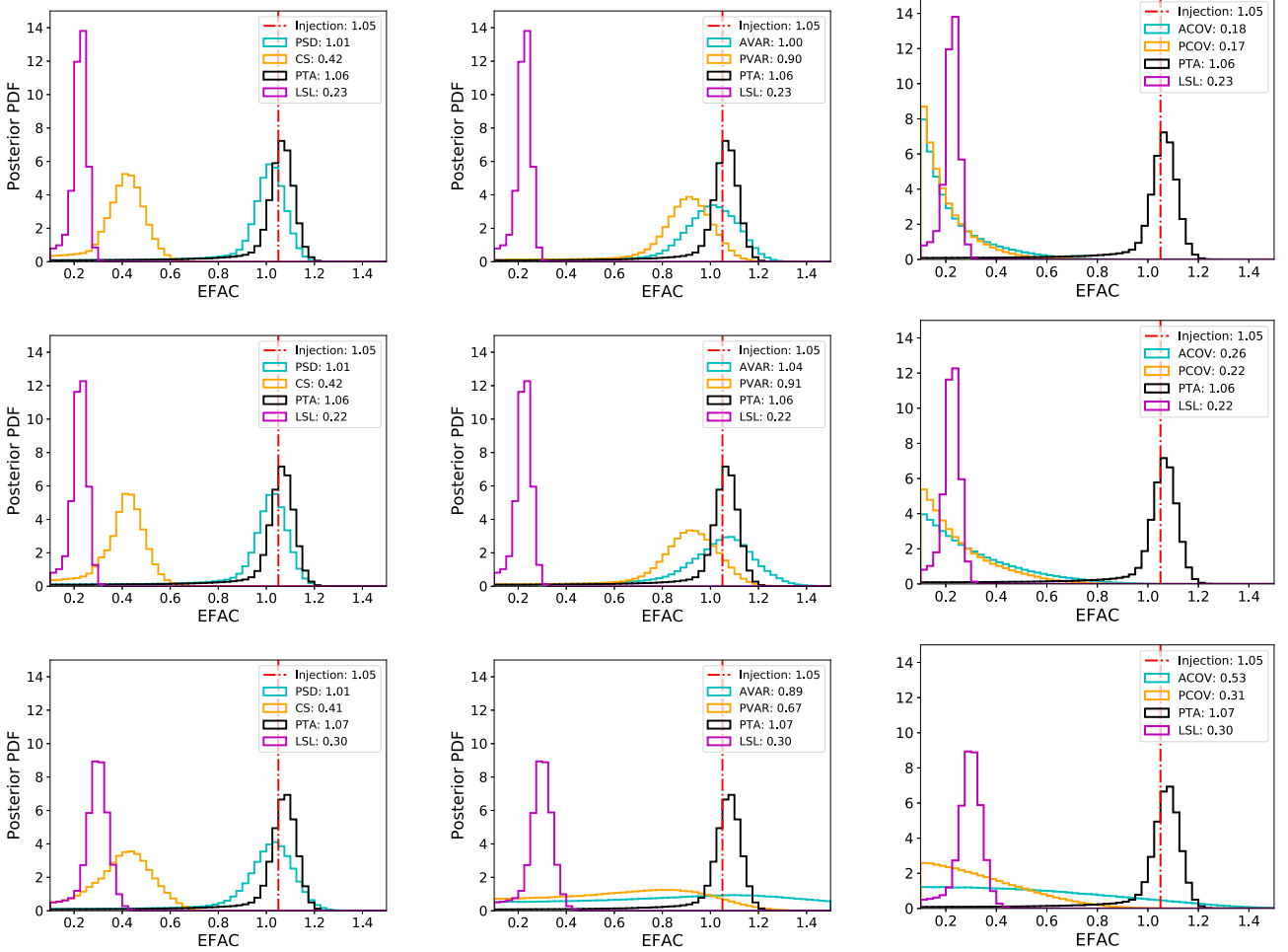


Figure 5. Evolution of the recovery of the EFAC parameter from the white noise case (top row), boundary case (middle row) to the red noise case (bottom row): spectral (left column), variance (middle column), and covariance methods (right column). The numbers in the legends represent the median values of the posterior distributions.

index γ_{PTA} and amplitude A_{PTA} from the PSD method in the top left-hand panel. The constraints on the two PTA red noise parameters are comparable to the standard PTA analysis.

On the other hand, the decrease of the white noise level through the CS improves the detection of the injected red noise and the constraints. The lowest three frequency bins contribute to the recovery of the red noise power law, see bottom left-hand panel of Fig. 7. Both the posterior distributions in the top panel of the figure and Table 1 show that the CS performs a little better with tighter constraints than both the PSD and the standard PTA methods. In the LSL, the signal is clearly detectable and the tightest constraints can be placed in the top left-hand panel.

4.2.2 Variance analysis

The middle column of Fig. 7 shows the comparison red noise parameter corner plot in the top panel and recovered AVAR (middle panel) and PVAR (bottom panel) with the injected red noise. As equation (16) is only valid for $1.5 \leq \gamma_{PTA} \leq 4.5$, the boundaries are reduced in the top row middle panel of Fig. 7. With this caveat, the red noise spectral index γ_{PTA} and amplitude A_{PTA} can still be compared, however, the cut-offs at 1.5 and 4.5 limit the comparison and can thus influence the red noise recovery.

The recovered variances in the middle and bottom panels of the middle column of Fig. 7 illustrate the fact that the highest octave of τ with very large uncertainty is unreliable and contributes very little to the overall analysis. Since we are looking at a boundary case, the injections are close or within the uncertainties of the variance computation. Therefore, they do not necessarily coincide with the median of the recovered variance, see also Table 1. Additionally, the prior choice allows for a large distribution of small amplitude red noise, producing a median variance close to white noise.

This results in a conservative estimation that there is little red noise in the boundary case with AVAR and PVAR. However, as there a small turn-up in the higher time-steps, both provide some evidence of the injected red noise. This can be seen in the top row middle panel corner plot, especially in the A_{PTA} parameter and a small improvement in constraining the spectral index. The PTA analysis is also able to pick up the red noise and performs better than the variances. With the lowest level of white noise, it is no surprise that the LSL analysis performs the best.

4.2.3 Covariance analysis

We can further focus on the red noise detection by sacrificing the constraints on the white noise. This is done in a similar way as with

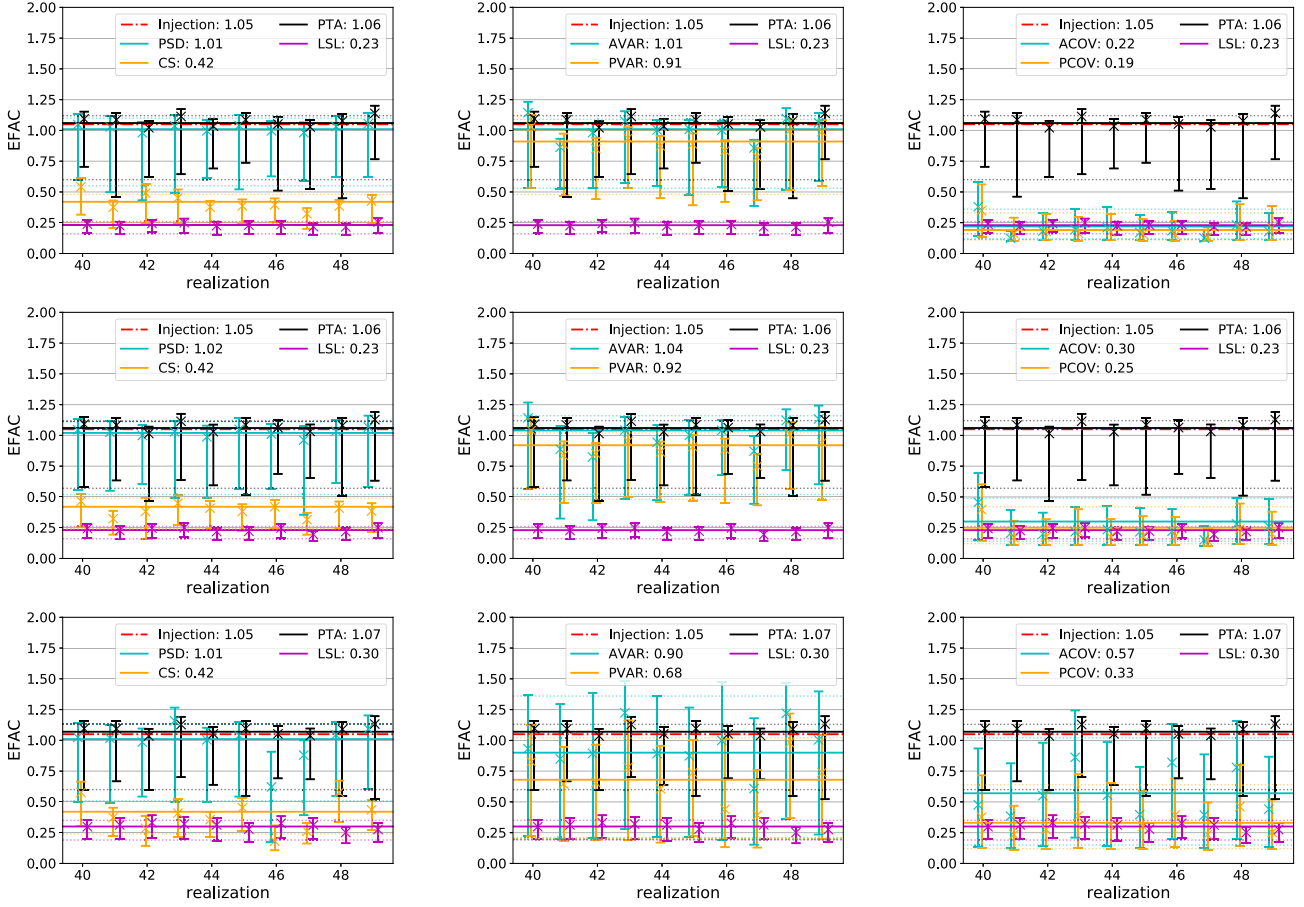


Figure 6. Examples of the recovery of the EFAC parameter from one set of realizations of the white noise case (top row), boundary case (middle row) and the red noise case (bottom row): spectral (left column), variance (middle column), and covariance methods (right column). The numbers in the legends represent the average median values of the posterior distributions.

Table 1. List of the average median posterior values and average 90 percent central region bounds for the three constrainable PTA noise parameters from the analyses of 100 realizations in the boundary case with different analysis methods.

parameter	EFAC	γ_{PTA}	$\log_{10} A_{PTA}$
Injection	1.05	3	-14
PTA	$1.06^{+0.06}_{-0.49}$	$3.55^{+2.32}_{-1.76}$	$-14.48^{+0.75}_{-1.24}$
LSL	$0.23^{+0.03}_{-0.07}$	$3.24^{+1.62}_{-1.10}$	$-14.16^{+0.35}_{-0.55}$
PSD	$1.02^{+0.09}_{-0.50}$	$3.69^{+2.59}_{-2.00}$	$-14.75^{+1.09}_{-1.75}$
CS	$0.42^{+0.07}_{-0.18}$	$3.55^{+1.44}_{-1.19}$	$-14.35^{+0.50}_{-0.71}$
AVAR	$1.04^{+0.12}_{-0.52}$	$2.78^{+1.33}_{-1.07}$	$-15.36^{+1.50}_{-2.91}$
PVAR	$0.92^{+0.10}_{-0.44}$	$2.76^{+1.48}_{-1.13}$	$-15.35^{+1.52}_{-4.03}$
ACOV	$0.30^{+0.19}_{-0.16}$	$2.99^{+1.26}_{-1.21}$	$-15.77^{+1.67}_{-3.27}$
PCOV	$0.25^{+0.17}_{-0.13}$	$2.97^{+1.35}_{-1.31}$	$-16.49^{+2.33}_{-3.16}$

the CS by computing the covariances, such that uncorrelated white noise is rejected to allow for the correlated red noise to be more prominent. The right-hand column of Fig. 7 shows the posterior distributions, recovered ACOV and PCOV from top to bottom. The same limitation on the spectral index γ_{PTA} for variances applies also to the covariances.

Despite the lower white noise from the covariances, the constraints on the red noise parameters γ_{PTA} and A_{PTA} are actually worse than their variance counterparts, as it can be seen in the top right-hand panel of Fig. 7 and Table 1. As there is no correlated white noise, the lowest time-steps become upper limits with large uncertainties and do not provide any information on the red noise. This leaves only one or two time-steps with information, which is shown in the right column middle and bottom panels. The overall recovery using all time-steps thus favours low amplitude white and red noise. Using only time-steps with small uncertainties could produce tighter, but less reliable, red noise constraints.

4.3 Red noise

The strong red noise case is of particular interest as it shows how well the different methods can in principle divide different red noise realizations in the large-signal regime. An overview on the constraints on the PTA parameters for the red noise case is given by Table 2 for the standard PTA/LSL and clock metrology analysis. The overall posterior distributions for all methods are shown in Fig. 9. The results of some example analyses of a single simulation set can be found in Fig. 10.

In general, all methods can detect the injected red noise very well. As the red noise is very large, there is very little difference in the constraints on the red noise parameters γ_{PTA} and $\log_{10} A_{PTA}$.

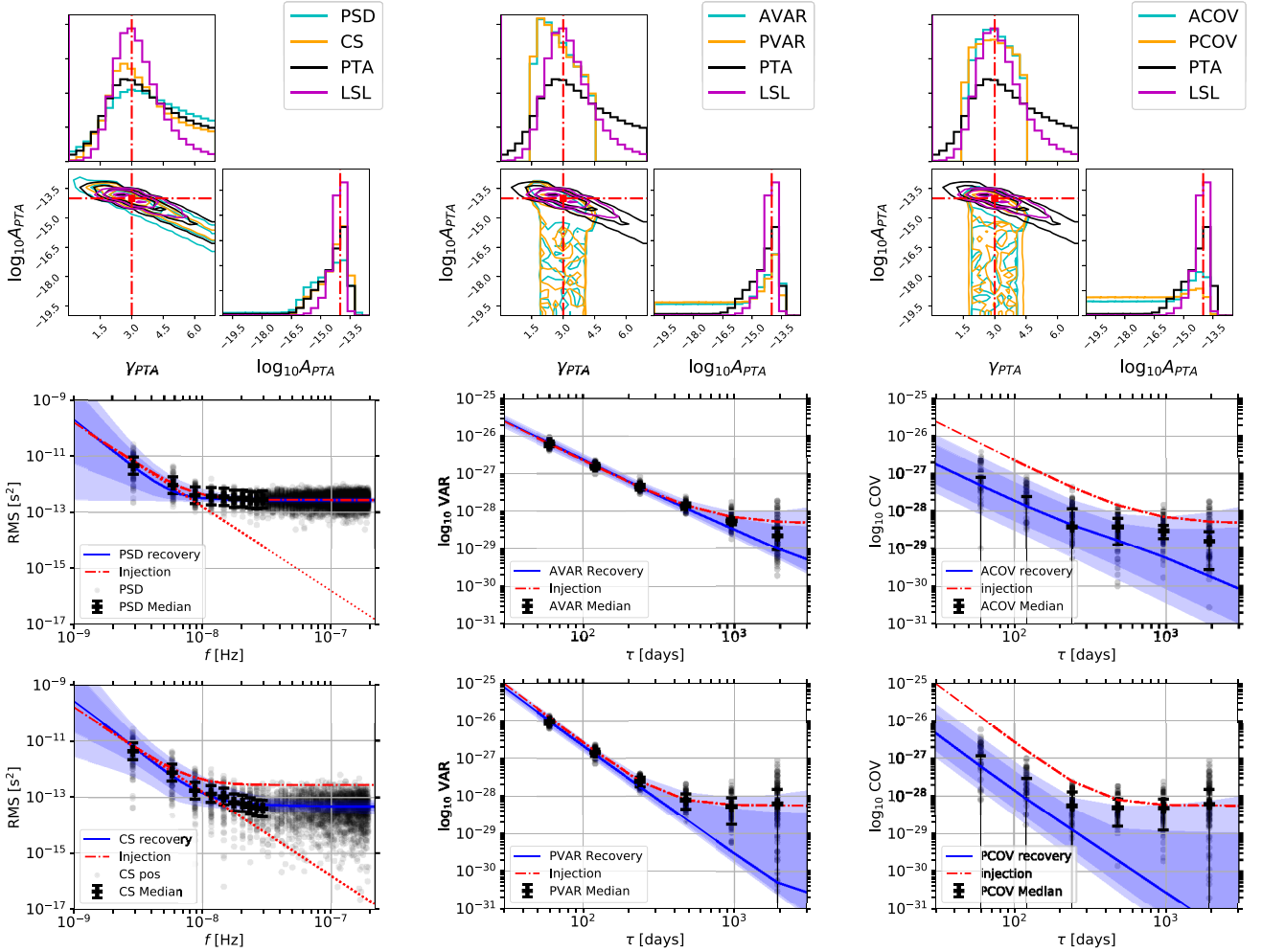


Figure 7. The left, middle, and right column show results from the spectral, variance, and covariance analysis, respectively, for the boundary case. The top row shows the posterior distribution corner plots. The middle row shows the recovered PSD (left), AVAR (middle), and ACOV (right). The bottom row shows the recovered CS (left), PVAR (middle), and PCOV (right). Each recovery figure contains the values from the 100 realizations in black dots, the injection in the red-dotted lines and the median values and central 68 and 90 per cent credible regions in blue bands.

between the PTA and the LSL analysis. Both are centred around the injected values with tight constraints, see top panels of Fig. 9. So do the spectral method recoveries, although with a very small tail, which is longer for the PSD compared to the CS. The uncertainties in the spectral methods allow for small values even at the lowest frequencies. This leads to a tail of lower amplitudes being recovered, see Table 2. Both variances and covariances perform very similar to the PTA/LSL analysis in terms of the tightness of the constraints on the red noise parameters, especially PVAR and PCOV are comparable to the LSL analysis. This is due to the fact that the higher time-steps are dominated by red noise and have therefore very similar VAR/COV values. Additionally, the absolute uncertainties σ_i are the same for both variances and covariances.

Consequently, the covariances at the lowest τ are not as well defined as the variances for the same time-steps. These large relative uncertainties and the lower COV values translate into a small bias for the red noise to have steeper spectral indices γ_{PTA} than with the variance methods. On the other hand, the variance methods are still a little bit impacted by the choice of priors, see Table 2. This can be seen in the bottom middle panel of Fig. 9, where the median recovered PVAR is lower than the median PVAR values of the 100 realizations. The same is also true for AVAR recovery.

Table 2 also shows that both AVAR and ACOV have a longer tail towards lower red noise amplitudes $\log_{10} A_{\text{PTA}}$ when compared to PVAR and PCOV. This can be the advantage of the parabolic (co)variance transitioning from a spectral index of $\beta_{\text{VAR}} = -3$ for white noise to $\gamma_{\text{VAR}} = 0$ for the injected red noise. In comparison the Allan (co)variance has a transition of one order less, i.e. from $\beta_{\text{VAR}} = -2$ to $\gamma_{\text{VAR}} = 0$.

5 CONCLUSIONS

We have shown that clock comparison methods, such as the CS and variances, can be applied to pulsar timing observations. Initial tests show that these methods produce slightly worse, sometimes comparable constraints on the pulsar noise parameters when compared to the standard PTA and optimal LSL analysis. Both perform very well in recovering the injected white noise, regardless of the level of red noise. The red noise recovery itself is broadly consistent with the injections, with different realizations giving similar constraints. In general, the metrology clock comparison methods perform well in the pure white noise case, slightly worse in the boundary case and comparable in the strong red noise case. The spread of recoveries

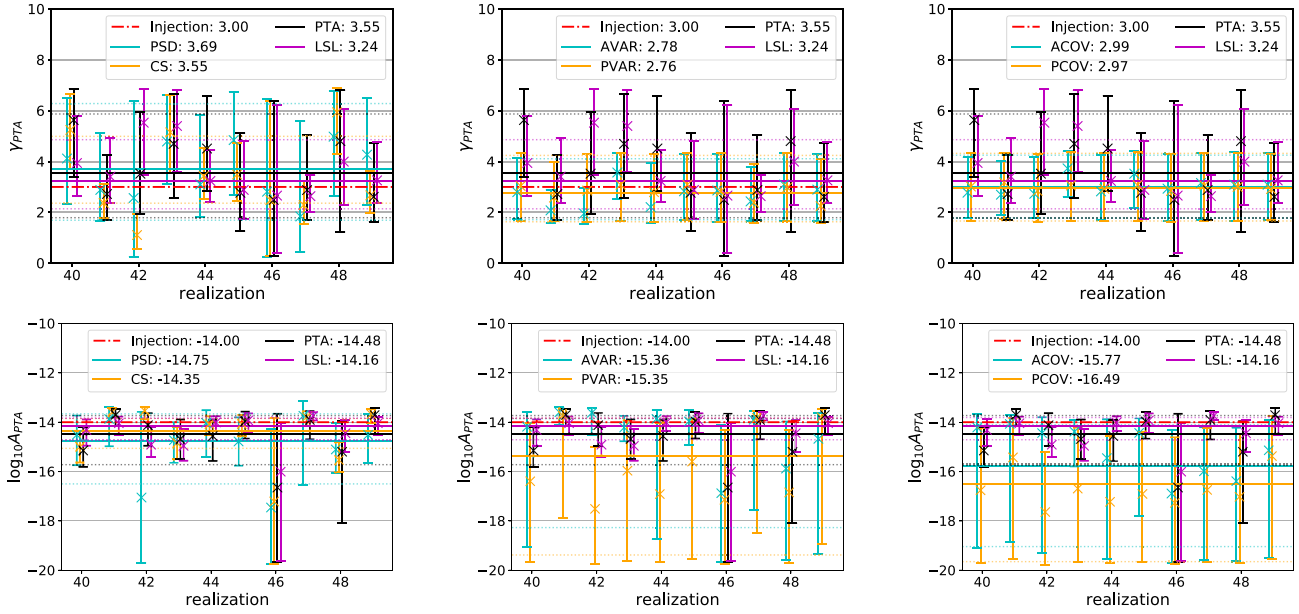


Figure 8. Examples of the recovery of the γ_{PTA} (top row) and A_{PTA} (bottom row) parameters to central 90 per cent from one set of realizations of the boundary case: spectral (left column), variance (middle column), and covariance methods (right column). The numbers in the legends represent the average median values of the posterior distributions.

Table 2. List of the average median posterior values and average 90per cent central region bounds for the three constrainable PTA noise parameters from the analyses of 100 realizations in the red noise case with different analysis methods.

Parameter	EFAC	γ_{PTA}	$\log_{10} A_{\text{PTA}}$
Injection	1.05	3	-13
PTA	$1.07^{+0.06}_{-0.47}$	$3.04^{+0.62}_{-0.52}$	$-13.02^{+0.11}_{-0.11}$
LSL	$0.30^{+0.05}_{-0.11}$	$3.09^{+0.55}_{-0.46}$	$-13.03^{+0.09}_{-0.09}$
PSD	$1.01^{+0.12}_{-0.50}$	$2.98^{+0.59}_{-0.51}$	$-13.09^{+0.13}_{-0.18}$
CS	$0.42^{+0.08}_{-0.18}$	$3.06^{+0.41}_{-0.38}$	$-13.05^{+0.07}_{-0.10}$
AVAR	$0.90^{+0.46}_{-0.69}$	$2.75^{+0.77}_{-0.63}$	$-13.06^{+0.14}_{-1.10}$
PVAR	$0.68^{+0.34}_{-0.48}$	$2.73^{+0.52}_{-0.60}$	$-13.02^{+0.08}_{-0.17}$
ACOV	$0.57^{+0.45}_{-0.42}$	$3.03^{+0.26}_{-0.69}$	$-13.27^{+0.26}_{-2.24}$
PCOV	$0.33^{+0.31}_{-0.21}$	$3.05^{+0.55}_{-0.50}$	$-13.15^{+0.12}_{-0.67}$

between different realizations is also larger than the comparison PTA/LSL analysis.

The PSD method produces slightly broader parameter posterior distributions compared to those from the PTA analysis. The CS allows for a long tail of low values, including negative values. This means a reduction of white noise to almost zero at high frequencies, but including these negative values will skew the red noise recovery. Using Gaussian distributions also decreases the possibility for low CS values and thus give slightly more optimistic constraints for A_{PTA} and γ_{PTA} . Thus, the CS method performs slightly better than the PTA analysis, but still a bit worse than the LSL analysis. Since the LSL's white noise level is only a fifth of that of the PTA observations, it can put the tightest constraints on the red noise.

Similar constraints can be put on the red noise parameters for both the variances and covariances. The (co)variance methods perform

particularly badly in recovering the white noise in the red noise case. By design, ACOV and PCOV should reject uncorrelated white noise and only leave the correlated red noise signal. Therefore, no useful information can be gained for the white noise, while some extra power may be pushed into the red noise. This lack of information may limit the performance of the covariances in precisely constraining the red noise and manifests as a small bias in γ_{PTA} or a long tail of low red noise amplitudes A_{PTA} .

The clock comparison methods complement the standard PTA analysis by giving independent constraints on the noise parameters. This can be especially valuable for certain realizations, when the different analysis methods do not converge to the same values.

We continue the study and apply the methods described to the observations from the LEAP project, where we deal with the challenges from real pulsar timing data. These include unevenly sampled data points, different levels of white noise between RTs and observations and uncertainties in the computation of the spectral densities and (co)variances.

ACKNOWLEDGEMENTS

This work is funded by the French Agence Nationale de la Recherche (ANR) Programme d'Investissement d'Avenir (PIA) under the FIRST-TF network (ANR-10-LABX-48-01) project and the Oscillator IMP project (ANR-11-EQPX-0033-OSC-IMP) and the EUR EIPHI Graduate School (ANR-17-EURE-00002), and by grants from the Région Bourgogne Franche Comté intended to support the PIA.

The Nançay radio Observatory is operated by the Paris Observatory, associated with the French Centre National de la Recherche Scientifique (CNRS). We acknowledge financial support from 'Programme National de Cosmologie et Galaxies' (PNCG) and 'Programme National Hautes Energies' (PNHE) funded by CNRS/INSU-IN2P3-INP, CEA and CNES, France.

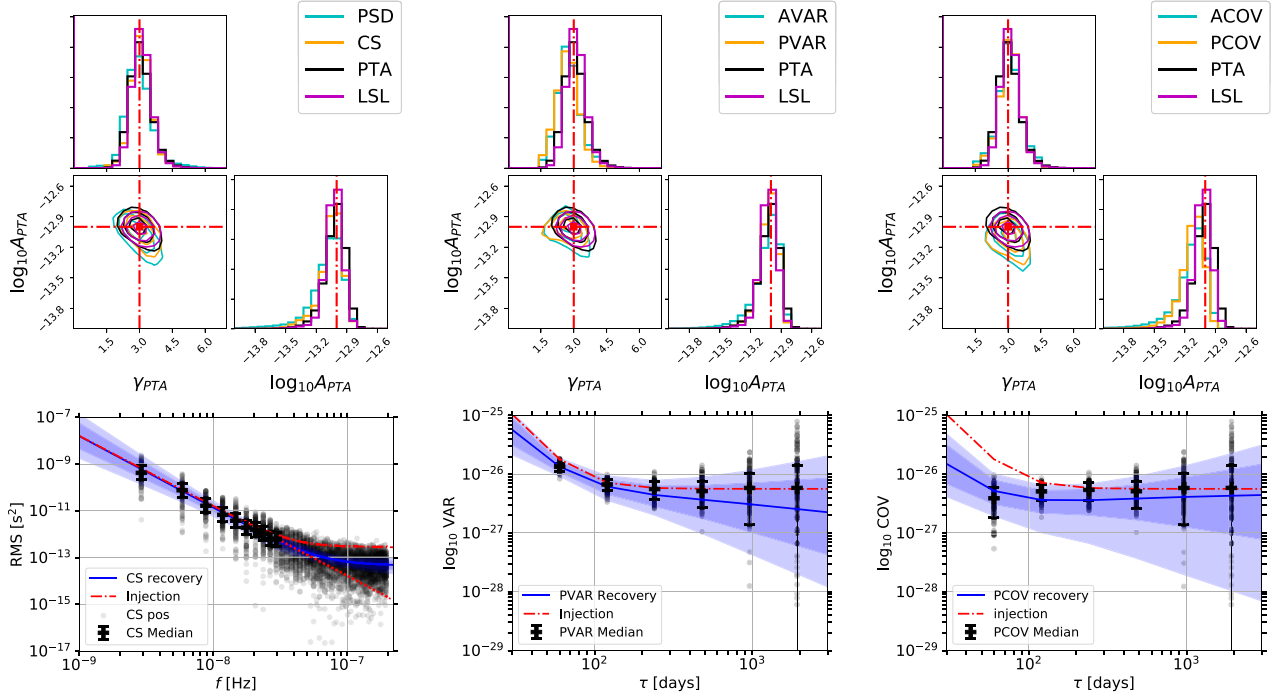


Figure 9. Top and bottom rows correspond those in Fig. 7, respectively, in the same style. However, the results are for the red noise case.

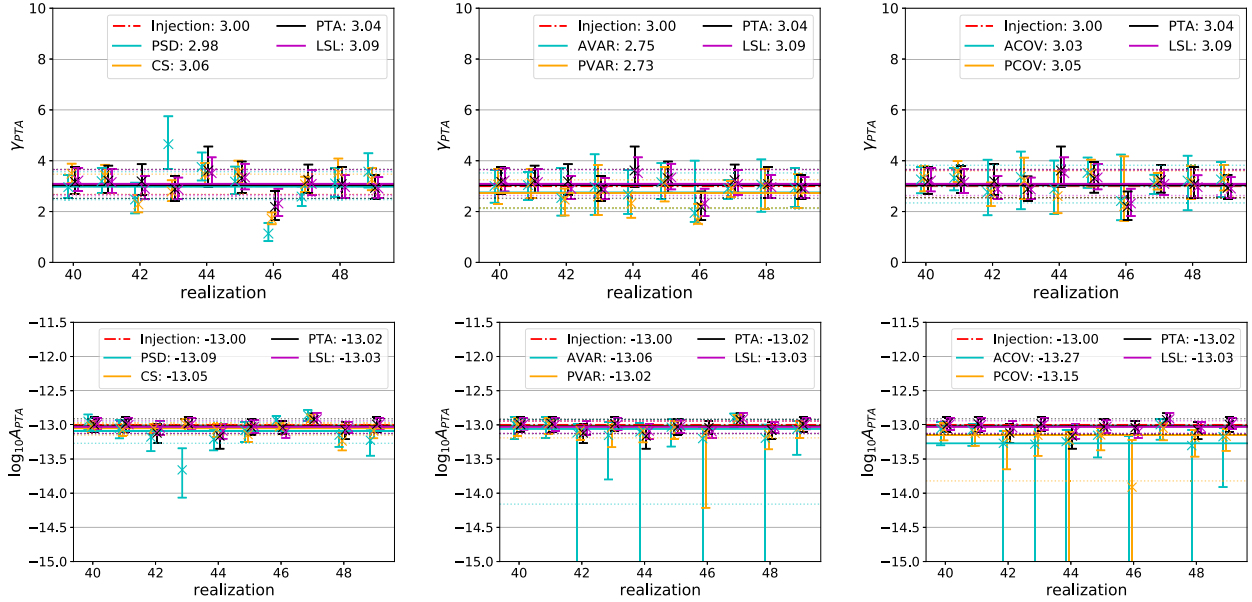


Figure 10. Examples of the recovery of the γ_{PTA} (top row) and A_{PTA} (bottom row) parameters from one set of realizations of the red noise case: spectral (left column), variance (middle column), and covariance methods (right column). The numbers in the legends represent the median values of the posterior distributions.

We acknowledge the support of our colleagues in the Large European Array for Pulsars and European Pulsar Timing Array collaboration.

DATA AVAILABILITY

The simulated residual series and posteriors can be obtained from the authors upon request.

REFERENCES

- Alam M. F. et al., 2020, *ApJS*, 252, 4
- Allan D. W., 1966, *Proc. IEEE*, 54, 221
- Allan D. W., Barnes J. A., 1981, *Proc. 35 FCS*, Ft. Monmouth, NJ, p. 470
- Arzoumanian Z. et al., 2016, *ApJ*, 821, 13
- Arzoumanian Z. et al., 2018, *ApJ*, 859, 47
- Barnes J. A. et al., 1971, *IEEE Trans. Instrum. Meas.*, 20, 105
- Bassa C. G. et al., 2016, *MNRAS*, 456, 2196
- Benkler E., Lisdat C., Sterr U., 2015, *Metrologia*, 55, 565

- Caballero R. N. et al., 2016, *MNRAS*, 457, 4421
 Chen S., Sesana A., Del Pozzo W., 2017, *MNRAS*, 470, 1738
 Del Pozzo W., Veitch J., 2015, CPNest: Parallel Nested Sampling in Python.
 Available at: <https://github.com/johnveitch/cpnest>
 Desvignes G. et al., 2016, *MNRAS*, 458, 3341
 Ellis J. A., Vallisneri M., Taylor S. R., Baker P. T., 2017, ENTERPRISE
 (Enhanced Numerical Toolbox Enabling a Robust Pulsar Inference
 Suite). Available at: <https://github.com/nanograv/enterprise>
 Fest D., Grosblambert J., Gagnepain J.-J., 1983, *IEEE Trans. Instrum. Meas.*,
 32, 447
 Foster R. S., Backer D. C., 1990, *ApJ*, 361, 300
 Gray J. E., Allan D. W., 1974, Proc. Int'l Freq. Control Symp.. Atlantic City,
 NJ, USA, p. 243
 Hellings R. W., Downs G. S., 1983, *ApJ*, 265, L39
 Hobbs G. B., Edwards R. T., Manchester R. N., 2006, *MNRAS*, 369,
 655
 Jenet F. A., Romano J. D., 2015, *Am. J. Phys.*, 83, 635
 Kerr M. et al., 2020, *Publ. Astron. Soc. Aust.*, 37, e020
 Lantz E., Calosso C. E., Rubiola E., Giordano V., Fluhr C., Dubois B., Vernotte
 F., 2019, *IEEE Trans. Ultras. Ferroelectr. Freq. Contr.*, 66, 1942
 Lentati L. et al., 2015, *MNRAS*, 453, 2576
 Lentati L. et al., 2016, *MNRAS*, 458, 2161
 Lesage P., Audoin C., 1973, *IEEE Trans. Instrum. Meas.*, 22, 157
 Lorimer D. R., Kramer M., 2012, Handbook of Pulsar Astronomy. Cambridge
 Univ. Press, Cambridge
 Perera B. B. P. et al., 2019, *MNRAS*, 490, 4666
 Perrodin D., Sesana A., 2018, Radio Pulsars: Testing Gravity and Detecting
 Gravitational Waves. Springer, Cham, p. 95
 Phinney E. S., 2001, *MNRAS*, preprint (arXiv:astro-ph/0108028)
 Rubiola E., Vernotte F., 2010, preprint (arXiv:1003.0113)
 Rutman J., 1978, *Proc. IEEE*, 66, 1048
 Smits R. et al., 2017, *Astron. Comput.*, 19, 66
 Taylor S. R., Vallisneri M., Ellis J. A., Mingarelli C. M. F., Lazio T. J. W.,
 van Haasteren R., 2016, *ApJ*, 819, L6
 Vallisneri M., 2015, libstempo – a Python Wrapper For Tempo2. Available
 at: <https://github.com/vallis/libstempo>
 Verbiest J. P. W. et al., 2009, *MNRAS*, 400, 951
 Vernotte F., Lantz E., 2012, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*,
 59, 523
 Vernotte F., Zalamansky G., 1999, *IEEE Trans. Ultrason. Ferroelectr. Freq.*
Control, 46, 1545
 Vernotte F., Lenczner M., Bourgeois P.-Y., Rubiola E., 2016, *IEEE Trans.*
Ultrason. Ferroelectr. Freq. Control, 63, 611
 Vernotte F., Rubiola E., Chen S., 2020, preprint (arXiv:2005.13631)
 Walter T., 1994, *IEEE Trans. Instrum. Meas.*, 43, 69

APPENDIX A: ESTIMATES AND WEIGHTS OF THE CROSS-SPECTRUM

For a given frequency f_0 , we consider the noise level S_w from uncorrelated observational noise of the RTs and the signal level S_r from the pulsar red noise, see equation (4). We perform a large Monte Carlo simulation with 10^7 realizations at the given frequency with varying levels of noise and signal.

The large simulation allows for the evaluation of the estimates of the CS. We determine empirically the statistics and conclude that the most reliable representation can be obtained from the logarithmic CS estimates distinguishing between positive and negative estimates. The empirical distributions are described by two parameters: μ_{emp} and σ_{emp} are the mean and variance of the log of the CS estimates, respectively. Despite the asymmetry of the probability density functions we choose to model them with Gaussians with mean μ_{mod} and variance σ_{mod} and find the best-fitting values to both positive and negative CS histograms (see inset of Fig. A1).

By varying the noise and signal levels, we obtain numerical expressions for the four parameters as a function of S_w and S_r . They

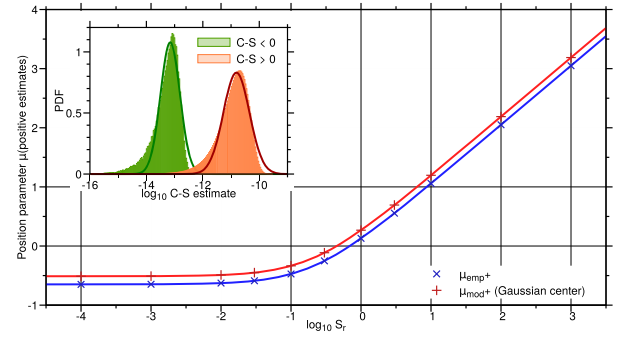


Figure A1. Dependence on S_r of the log mean (in blue) and of the Gaussian centre (in red) for positive CS estimates when $S_w = 1$. Inset: Gaussian fit of the histograms of the log of the CS estimates ($S_w = 3 \times 10^{-13}$, $S_r = 10^{-11}$).

Table A1. Dependence of the position and width parameters of the CS estimates on S_w and S_r .

Parameter	Negative estimates	Positive estimates
μ_{emp}	$\log_{10} \left(\frac{0.1949 S_w + 3.876 S_r}{S_w + 4.552 S_r} \right)$	$\log_{10} [0.2245 (S_w + 5 S_r)]$
μ_{mod}	$\log_{10} \left(\frac{0.2665 S_w + 3.882 S_r}{S_w + 4.404 S_r} \right)$	$\log_{10} [0.3082 (S_w + 5 S_r)]$
σ_{emp}	$\log_{10} \left(\frac{1.576 S_w + 4.147 S_r}{S_w + 3.833 S_r} \right)$	$\log_{10} (2.045) = 0.3106$
σ_{mod}	$\log_{10} \left(\frac{1.251 S_w + 3.835 S_r}{S_w + 3.640 S_r} \right)$	$\log_{10} (1.635) = 0.2136$

can be found in Table A1. The factor $(S_w + 5 S_r)$ shows that we can retrieve the correct number of degrees of freedom, i.e. 5, in a purely empirical way.

The important results are the following:

- (i) the width parameters $\sigma_{\text{emp}, \text{mod}}$ can be considered as constants,
- (ii) the position parameters for negative estimates $\mu_{\text{emp}, \text{mod}}^-$ only depend on S_w ,
- (iii) the position parameters for positive estimates $\mu_{\text{emp}, \text{mod}}^+$ depend on S_w as well as S_r , and
- (iv) the shift between μ_{emp} and μ_{mod} is constant and can be evaluated from the 10^7 realizations to be $\log_{10}(1.373) = 0.138$.

This shift between simulation and model can be explained by the long tail towards lower numbers of the empirical distributions that decrease μ_{emp} compared to μ_{mod} .

The (almost) constancy of the width parameters confirms that the log of the CS estimates is a reliable representation. Moreover, the position parameters of the negative estimates do not depend on S_r . This implies that the negative estimates do not carry any information about the signal level and should not be taken into account. The variation of the signal and noise levels makes f_0 representative for any frequency. Additionally, since the PSD is similar to the positive CS estimator, we can apply the same Gaussian model.

APPENDIX B: WEIGHTS AND UNBIASING OF THE VARIANCES

The uncertainty σ_i on each time-step i , whether variances or covariances, can be calculated from the average values of the variances VAR

$$\sigma_i^2 = 2 \text{VAR}^2 / \nu, \quad (\text{B1})$$

Table B1. Effective degrees of freedom and unbiasing factors for the variances in the red noise case.

τ	60	120	240	480	960	1920
Assumed γ_{PTA}	3	3	3	3	3	3
AVAR ν	80	38	18	7.7	3.0	1.1
PVAR ν	82	40	19	8.6	3.4	1
AVAR unbiasing factor	1	1	1	1.14	1.45	3.56
PVAR unbiasing factor	1	1	1	1.12	1.45	3.56

Table B2. Effective degrees of freedom and unbiasing factors for the variances in the white noise and boundary case.

τ	60	120	240	480	960	1920
Assumed γ_{PTA}	0	0	0	0	0	0
AVAR ν	66	65	63	58	46	5.8
PVAR ν	100	49	23	11	4.4	1
AVAR unbiasing factor	1	1	1	1	1	1.19
PVAR unbiasing factor	1	1	1	1.1	1.31	3.56

where ν are the effective degrees of freedom, whose expressions depend on the type of noise. The expressions used in the analysis for AVAR can be found in Lesage & Audoin (1973): for white phase modulation ($\gamma_{\text{PTA}} = 0$)

$$\nu \approx \frac{(N+1)(N-2m)}{2(N-m)} \quad (\text{B2})$$

and flicker frequency modulation ($\gamma_{\text{PTA}} = 3$):

$$\nu \approx \frac{5N^2}{4m(N+3m)}. \quad (\text{B3})$$

The equation for PVAR is given in Vernotte et al. (2020) as

$$\nu \approx \frac{35}{Am/M - B(m/M)^2}, \quad (\text{B4})$$

where A and B are constants that depend on the type of the modulation. They are given as $A = 23$ for $\gamma_{\text{PTA}} = 0$ and $A \approx 28$ for $\gamma_{\text{PTA}} = 3$, whilst $B \approx 12$ for all spectral indices. As pointed out in Vernotte et al. (2020), the formula is unreliable for the highest τ . However, from the theory we know that $\nu = 1$ at the largest time-step for PVAR. We also notice a mixture of white and red noise for our realizations, especially for the boundary case. Consequently, we approximate the degrees of freedom for the boundary case using those from the white noise case as a conservative approach. The degrees of freedom used for our analysis can be found in Tables B1 and B2 and are the same for the covariances.

Finally, Vernotte & Lantz (2012) have shown that the largest time-steps are statistically biased towards lower VAR/COV values. The factors that need to be applied for any time-step τ with its degrees of freedom ν can be found in table II in Vernotte & Lantz (2012). We also show the factors used in this paper in Tables B1 and B2.

This paper has been typeset from a \LaTeX file prepared by the author.