CrossMark

# Modeling the Uncertainties of Solar System Ephemerides for Robust Gravitational-wave Searches with Pulsar-timing Arrays

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

[1] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA; Michele.Vallisneri@jpl.nasa.gov
[2] Theoretical AstroPhysics Including Relativity (TAPIR), MC 350-17, California Institute of Technology, Pasadena, CA 91125, USA
[3] Department of Physics and Astronomy, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN 37235, USA
[4] Infinia ML, 202 Rigsbee Avenue, Durham, NC 27701, USA
[5] Center for Gravitation, Cosmology and Astrophysics, Department of Physics, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, 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] X-Ray Astrophysics Laboratory, NASA Goddard Space Flight Center, Code 662, Greenbelt, MD 20771, USA
[9] Department of Physics and Astronomy, Widener University, One University Place, Chester, PA 19013, USA
[10] Department of Astronomy, Cornell University, Ithaca, NY 14853, USA
[11] Cornell Center for Advanced Computing, Ithaca, NY 14853, USA
[12] Department of Physics, Montana State University, Bozeman, MT 59717, USA
[13] Department of Physics and Astronomy, Franklin & Marshall College, P.O. Box 3003, Lancaster, PA 17604, USA
[14] University of Virginia, Department of Astronomy, P.O. Box 400325, Charlottesville, VA 22904, USA
[15] Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
[16] Department of Physics, Lafayette College, Easton, PA 18042, USA
[17] National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro, NM 87801, USA
[18] Department of Physics, Hillsdale College, 33 E. College Street, Hillsdale, MI 49242, USA
[19] Department of Physics, University of East Anglia, Norwich, UK
[20] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[21] Department of Physics, McGill University, 3600 University Street, Montreal, QC H3A 2T8, Canada
[22] University of Washington Bothell, 18115 Campus Way NE, Bothell, WA 98011, USA
[23] NCSA and Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[24] Department of Physics, Columbia University, New York, NY 10027, USA
[25] Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Northwestern University, Evanston, IL 60208, USA
[26] School of Physics and Astronomy, Rochester Institute of Technology, Rochester, NY 14623, USA
[27] Laboratory for Multiwavelength Astronomy, Rochester Institute of Technology, Rochester, NY 14623, USA
[28] Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester, M13 9PL, UK
[29] Green Bank Observatory, P.O. Box 2, Green Bank, WV 24944, USA
[30] Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269-3046, USA
[31] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA
[32] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George St., Toronto, ON M5S 3H4, Canada
[33] Hungarian Academy of Sciences MTA-ELTE "Extragalatic Astrophysics Research Group", Institute of Physics, Eötvös Loránd University, Pázmány P.s. 1/A, 1117 Budapest, Hungary
[34] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[35] Naval Research Laboratory, Washington DC 20375, USA
[36] Department of Physics, Oregon State University, Corvallis, OR 97331, 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 Sciences, Beijing 100101, People's Republic of China

## Abstract

The regularity of pulsar emissions becomes apparent once we reference the pulses' times of arrivals to the inertial rest frame of the solar system. It follows that errors in the determination of Earth's position with respect to the solar system barycenter can appear as a time-correlated bias in pulsar-timing residual time series, affecting the searches for low-frequency gravitational waves performed with pulsar-timing arrays. Indeed, recent array data sets yield different gravitational-wave background upper limits and detection statistics when analyzed with different solar system ephemerides. Crucially, the ephemerides do not generally provide usable error representations. In this article, we describe the motivation, construction, and application of a physical model of solar system ephemeris uncertainties, which focuses on the degrees of freedom (Jupiter's orbital elements) most relevant to gravitational-wave searches with pulsar-timing arrays. This model, BAYESEPHEM, was used to derive ephemeris-robust results in NANOGrav's 11 yr stochastic-background search, and it provides a foundation for future searches by

NANOGrav and other consortia. The analysis and simulations reported here suggest that ephemeris modeling reduces the gravitational-wave sensitivity of the 11 yr data set and that this degeneracy will vanish with improved ephemerides and with pulsar-timing data sets that extend well beyond a single Jovian orbital period.

*Unified Astronomy Thesaurus concepts:* Gravitational waves (678); Astronomy data analysis (1858); Ephemerides (464); Pulsar timing method (1305); Millisecond pulsars (1062)

## 1. Introduction

Pulsar timing exploits the remarkable regularity of millisecond-pulsar emissions to extract accurate system parameters from time-of-arrival (TOA) data sets (Lorimer & Kramer 2012), by fitting precise timing models that account for all pulse delays and advances, from generation near the neutron stars to detection at the radio telescopes (Lommen & Demorest 2013). The largest time-dependent term in the model is the Rømer delay (Rømer 1676) due to the motion of Earth around the solar system barycenter (SSB), with magnitude ~500 s. Solar system ephemerides (SSEs), such as those produced by the Jet Propulsion Laboratory (JPL; see Folkner et al. 2009, 2014, 2016; Folkner & Park 2016, 2018), are used to convert observatory TOAs to the notional coordinate time of an inertial frame centered at the SSB. It follows that errors in our estimate of Earth's trajectory around the SSB produce a time-dependent bias in the TOAs.

Throughout many years of pulsar-timing studies and discoveries, SSE errors were always considered negligible compared to all other sources of noise and uncertainty (Foster & Backer 1990a; Edwards et al. 2006a). TOAs are now being collected with ever greater time spans and timing precisions, especially so by the pulsar-timing array (PTA) collaborations seeking to detect gravitational waves (GWs) as correlated residuals (i.e., TOAs minus timing model) in multipulsar data sets (Sazhin 1978; Detweiler 1979; Foster & Backer 1990b; Hobbs 2013; McLaughlin 2013; Desvignes et al. 2016; Verbiest et al. 2016). The estimated magnitude of the GW signature is ~100 ns or less, leading to the recent suggestion that SSE errors could measurably bias GW results obtained from these data sets (Tiburzi et al. 2016). Indeed, while SSE corrections and GWs have different spatial-correlation signatures across multiple pulsars, the two may yet be degenerate when sampled with a limited number of PTA pulsars (Roebber 2019).

Within the North American Nanohertz Observatory for Gravitational Waves (NANOGrav), we encountered this bias first hand in early 2017, when we observed that switching between the more recent SSEs issued by JPL (specifically DE421, DE430, DE435, and DE436: Folkner et al. 2009, 2014, 2016; Folkner & Park 2016) led to discrepant results (GW upper limits and detection statistics) in the search for a stochastic signal from the population of supermassive black-hole binaries, formed in the centers of galaxies following major-merger events, as performed on the NANOGrav 11 yr data set (Arzoumanian et al. 2018a, 2018b). See Figure 1 for the Bayesian posterior probabilities of the GW-background (GWB) amplitude, as obtained with a range of SSEs. We verified that other recent PTA results (Shannon et al. 2015; Arzoumanian et al. 2016) would be similarly affected.

In this article, we report on BAYESEPHEM, the physical model of SSE uncertainties that we developed and integrated with the NANOGrav GW analysis so that we could produce SSE-robust results (Arzoumanian et al. 2018a) by marginalizing our Bayesian statistics over the SSE model parameters.

Our model complements published SSEs, which do not generally include usable time-domain representations of orbit uncertainties. We adopted the conservative goal of bridging the JPL SSEs so that our analysis would yield the same GW-amplitude posteriors, no matter which SSE was used to offset the TOAs initially. As discussed in Section 4, the crucial element in our approach turns out to be the modeling of uncertainties in Jupiter's orbit, which create apparent SSB motions with periods comparable to the duration of our data set, and with amplitudes of ~50 m. These correspond to ~170 ns delays, within the GW sensitivity of our PTA. By marginalizing GW posteriors over a set of SSE correction parameters that include Jupiter's orbital elements, we achieve our bridging criterion for the SSEs adopted in Arzoumanian et al. (2018a) and indeed also for the newer DE438 (Folkner & Park 2018); see Figure 1. By contrast, while we include SSB corrections due to perturbations in the masses of the outer planets (Champion et al. 2010), we conclude that the current sensitivity of our data set is insufficient to constrain these masses better than recent spacecraft tracking and Doppler data sets (Jacobson et al. 2000, 2006; Jacobson 2009, 2014; Caballero et al. 2018).

In our 11 yr analysis (Arzoumanian et al. 2018a), we declined to adopt an aggressive stance that would have given more credence to the more recent available SSEs (DE435 and DE436), which are based on longer sets of solar system observations that fully cover the span of our PTA data set, and on more sophisticated analysis techniques. The resulting GW-amplitude posteriors imply less evidence for GWs than those obtained with earlier SSEs (see Table 1) and intermediate estimates for 95% amplitude upper limits (see Table 2). Nevertheless, even if DE435 and DE436 are described as having only minor differences (Folkner et al. 2016; Folkner & Park 2016), the resulting posteriors are still at variance—especially so in the low GW-amplitude limit, which affects the Savage–Dickey Bayes ratio used as our GW detection statistic. Thus, uncertainty modeling remains important even if we concentrate on newer SSEs.

This article is organized as follows. In Section 2, we summarize the SSE production, and we describe the history and stated accuracy of the JPL SSEs adopted in our work; in Section 3, we discuss the TOA delays induced by SSE errors, which are partially absorbed by the timing model, and we identify Jupiter's and Saturn's orbits as the drivers behind GW-posterior discrepancies; in Section 4, we formulate BAYESEPHEM and give details about its implementation in our PTA data-analysis software, ENTERPRISE (Ellis et al. 2019); in Section 5, we report on the Bayesian posteriors (for the GW amplitude and for orbital-correction parameters) obtained with BAYESEPHEM for NANOGrav's 11 yr data set, reproducing and expanding the results of Arzoumanian et al. (2018a); in Section 6, we present simulations that probe the reduction in GW sensitivity due to BAYESEPHEM; in Section 7, we discuss other approaches toward SSE uncertainty modeling; and last, in Section 8 we offer our brief conclusions.
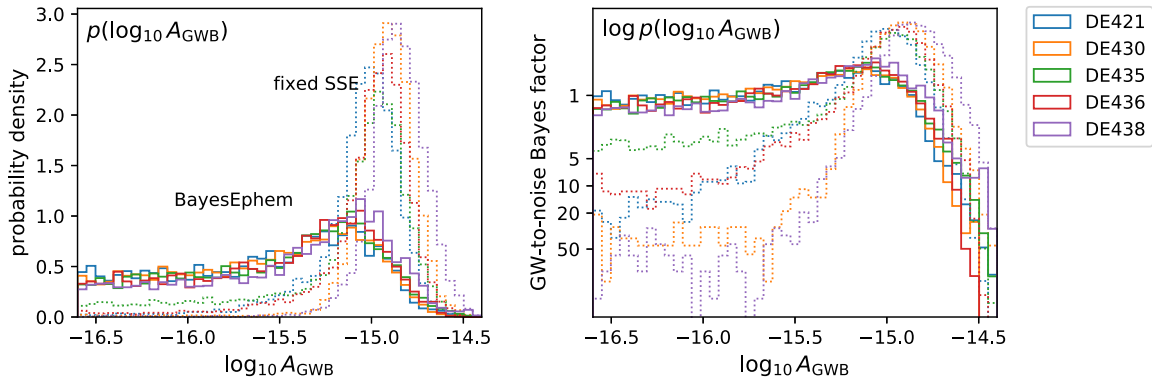
**Figure 1.** (Left) Bayesian posteriors for the amplitude $A_{GWB}$ (at $f = yr^{-1}$) of the GW stochastic background, modeled as a fixed power law with slope $\gamma = 11/3$, as appropriate for a population of inspiraling and GW-emitting supermassive black-hole binaries. The posteriors are computed for the NANOGrav 11 yr data set using individual JPL ephemerides (dotted lines) and BAYESEPHEM (solid lines). The GW model ("model 2A" in Arzoumanian et al. 2018a) does not include Hellings–Downs correlations, which however modify results only marginally. The incorporation of explicit SSE uncertainties into the analysis via BAYESEPHEM leads to substantially lower evidence of a GW background and to much greater consistency among the different SSEs. (Right) Same posteriors as in the left panel, shown with a logarithmic vertical scale that can be mapped to approximate GW vs. noise-only Bayes factors, by way of the Savage–Dickey formula $P(A_{GWB} = 0)/P(A_{GW} \neq 0) = p(A_{GWB} = 0|data)/p_{prior}(A_{GWB} = 0)$ (Dickey 1971). The application of BAYESEPHEM brings all Bayes factors close to unity, indicating no evidence for GWs once SSE uncertainties are taken into consideration. Bayes factors and 95% $A_{GWB}$ upper limits calculated from these curves are listed in Tables 1 and 2.

**Table 1**
Bayesian Evidence for a GWB with Different SSEs

| SSE | No BAYESEPHEM | Set III | Keplerian |
|-----|---------------|---------|-----------|
| DE421 | 10.6 | 0.68 | 0.68 |
| DE430 | 23.7 | 0.72 | 0.71 |
| DE435 | 2.0 | 0.76 | 0.72 |
| DE436 | 6.2 | 0.80 | 0.72 |
| DE438 | 40.7 | 0.91 | 0.87 |

**Note.** Savage–Dickey Bayes factors in favor of a common $\gamma = 13/3$ red-noise process ("model 2A" of a GWB in Arzoumanian et al. 2018a) in NANOGrav's 11 yr data set, obtained by fixing the fiducial SSE shown in column 1 and omitting (column 2) or applying BAYESEPHEM in its Set III (column 3) or Keplerian-element (column 4) formulations. The numbers shown here were drawn from newly reproduced Monte Carlo runs and differ from those of Arzoumanian et al. (2018a) by small sampling errors.

**Table 2**
95% Bayesian Upper Limits for $A_{GWB}$

| SSE | No BAYESEPHEM | Set III | Keplerian |
|-----|---------------|---------|-----------|
| DE421 | $1.54 \times 10^{-15}$ | $1.32 \times 10^{-15}$ | $1.35 \times 10^{-15}$ |
| DE430 | $1.76 \times 10^{-15}$ | $1.31 \times 10^{-15}$ | $1.33 \times 10^{-15}$ |
| DE435 | $1.59 \times 10^{-15}$ | $1.38 \times 10^{-15}$ | $1.40 \times 10^{-15}$ |
| DE436 | $1.64 \times 10^{-15}$ | $1.38 \times 10^{-15}$ | $1.41 \times 10^{-15}$ |
| DE438 | $1.94 \times 10^{-15}$ | $1.45 \times 10^{-15}$ | $1.44 \times 10^{-15}$ |

**Note.** 95% Bayesian upper limits on the amplitude of a $\gamma = 13/3$ GWB in NANOGrav's 11 yr data set, listed as in Table 1. The numbers shown here were drawn from newly reproduced Monte Carlo runs and differ from those of Arzoumanian et al. (2018a) by sampling errors $\sim 0.02$.

All computational results presented in this article were obtained with ENTERPRISE (Ellis et al. 2019), used in conjunction with the stochastic sampler PTMCMCSAMPLER[40] and the pulsar-timing package TEMPO2 (Hobbs & Edwards 2012). Code that supports all the calculations discussed here and that produces all the figures of this paper

is available as a Jupyter notebook at https://github.com/nanograv/11yr_stochastic_analysis.

## 2. Solar System Ephemerides

To derive the JPL SSEs[41] (and similarly for the French INPOP[42] and Russian EPM[43]), the orbits of the Sun, the planets, and a large number of asteroids are fit to heterogeneous data sets collected over the last few decades. Measurement techniques include spacecraft ranging and Doppler tracking, direct planetary radar ranging, very long baseline interferometry of spacecraft, and (for the Moon) the laser ranging of retroreflectors left by the Apollo missions (see Verma 2013 for a review). The masses of minor bodies are also included as fit parameters, while the masses of the planets are held fixed to values determined separately from spacecraft data for each planet (Jacobson et al. 2000, 2006; Jacobson 2014, 2009).

The parameters of the fit are the initial conditions ("epoch" positions and velocities) for all the bodies (as well as minor-body masses); from these, orbits are integrated numerically, providing a reference solution used to compute measurement residuals. The integration is repeated with minor displacements in all fit parameters, yielding variational partials for the orbits. The fit parameters are then corrected by finding the linear combination of the partials that minimizes the residuals in least-squares fashion, and the scheme is repeated until the solution converges (see, e.g., Newhall et al. 1983).

The most complex aspect of the process is the modeling of observations for data sets that are both varied in technique and unique to each planet and each new spacecraft (Moyer 2003)—a Sisyphean task. Because it is difficult to assign realistic uncertainties to many of the measurements (and, perhaps more importantly, to estimate their systematic errors), the formal errors and covariances estimated with the least-squares procedure are considered unreliable and are not published with the best-fit orbits. Instead, orbit accuracy is assessed by

---

[40] https://github.com/jellis18/PTMCMCSampler

[41] https://ssd.jpl.nasa.gov/?ephemerides
[42] https://www.imcce.fr/inpop
[43] http://iaaras.ru/en/dept/ephemeris/epm

analyzing model residuals and by comparing estimates that use different subsets of the data (Folkner & Park 2015, 2018).

In this paper, we work with the four JPL SSEs used to analyze the NANOGrav 11 yr data set (Arzoumanian et al. 2018a) as well as the more recent DE438. While we have also investigated some of the SSEs provided by INPOP (Fienga et al. 2014; Viswanathan et al. 2018), we find that they are not qualitatively different in terms of their GW constraints.

*DE421* (Folkner et al. 2009): published in 2009, based on data through 2007. The orbits of the inner planets are known to subkilometer accuracy—those of Jupiter and Saturn to tens of kilometers; Uranus and Neptune are not well determined. The axes of the ephemeris are oriented with the International Celestial Reference Frame (Ma et al. 1998) with accuracy $\lesssim$1 mas.

*DE430* (Folkner et al. 2014): published in 2014, based on data through 2013. Orbit integration relies on a more sophisticated dynamical model. The Saturn orbit is more accurate thanks to the improved treatment of range measurements to the Cassini spacecraft (Hees et al. 2014). The axes of the ephemeris are oriented with the 2009 update of the International Celestial Reference Frame, ICRF2 (Fey et al. 2015) with accuracy $\lesssim$0.2 mas, which represents the limiting error source for the inner planets, corresponding to orbit uncertainties of a few hundred meters. The Jupiter and Saturn orbits are determined to tens of kilometers, those of Uranus and Neptune (which are constrained mainly by astrometric measurements) to thousands of kilometers.

*DE435* (Folkner et al. 2016): published in 2016; it improves the Saturn orbit using Cassini data through 2015, correcting it by ∼1.5 km. The Jupiter orbit had been updated (by ∼50 km) in the 2015 DE434 ephemeris (Park et al. 2015) by reprocessing data from six spacecraft flybys and adding data from the New Horizons flyby. In DE435, Jupiter's orbit is tweaked further (by ∼20 km) by reweighting the data sets. These changes are deemed "consistent" with the estimated orbit uncertainties (Park et al. 2015; Folkner et al. 2016).

*DE436* (Folkner & Park 2016): published at the end of 2016; it updates the Jupiter ephemeris (by ∼20 km) for use by the Juno navigation team. Saturn's orbit changes very slightly.

*DE438* (Folkner & Park 2018): published in 2018 June; it updates the Jupiter ephemeris (by ∼10 km) with Juno measurements (six ranges and four VLBI observations near perijove), and the Saturn ephemeris (by ∼1 km) with reprocessed ranges through the end of the Cassini mission. The accuracy of Jupiter's orbits is deemed "at least a factor of four better than previous ephemerides," viz. $\lesssim$10 km.

Across these ephemerides, the orbit of Earth relative to the Sun is consistent at the 3 m (∼10 ns) level, after applying an overall rotation with respect to the International Celestial Reference Frame (within the uncertainties of that "tie") and a rotation rate about the ecliptic pole.[44] However, the orbit of the Sun and therefore the orbit of Earth (both relative to the SSB) match only at the 100 m (∼300 ns) level across ephemerides (see Figure 2). This discrepancy arises from the estimated positions of Jupiter, Saturn, Uranus, and Neptune. Using a simple dynamical model of the solar system (i.e., integrating the equations of Newtonian gravity for the eight planets and the Sun), it is easy to show that if we perturb the masses or the orbits of the outer planets, we affect the Sun-to-SSB and Earth-to-SSB trajectories through the resulting redefinition of the SSB, rather than through the very minor changes in the gravitational pull of the outer planets (see also Guo et al. 2019).

The last few JPL ephemerides focus on Jupiter and Saturn, as motivated by the navigation needs of JPL missions to those planets. In DE421 to DE436, Saturn's orbit is known better than Jupiter's, because the Cassini tracking data is more complete and accurate than was possible for previous spacecraft. In particular, the high-gain antenna of the Jupiter orbiter Galileo failed to deploy, leading to low-accuracy measurements. Data from the Juno spacecraft, which has been orbiting Jupiter since 2016 July, appears to improve the Jupiter ephemeris substantially (Folkner et al. 2009).

## 3. SSE Errors as Systematics for Pulsar-timing Arrays

The search for stochastic GW signals with PTAs exploits the distinctive quadrupolar signature in the interpulsar correlations of timing-model residuals (Hellings & Downs 1983). Residuals are obtained after applying a chain of corrections that convert the TOA measured at the radio telescope to the notional emission time in the pulsar system (see, e.g., Edwards et al. 2006b):

$$t_e^{\rm psr} = t_a^{\rm obs} - \Delta_\odot - \Delta_{\rm IS} - \Delta_{\rm B}, \quad (1)$$

where $t_e^{\rm psr}$ and $t_a^{\rm obs}$ are the emission and arrival times, $\Delta_\odot$ captures corrections between the observatory and SSB frames, $\Delta_{\rm IS}$ describes corrections between the SSB frame and the pulsar-system barycenter, and $\Delta_{\rm B}$ the model corrections from the pulsar system barycenter to the pulsar frame (which are relevant for binary systems). Among these terms, $\Delta_{\rm IS}$ is very large, but changes very slowly over the duration of pulsar data sets and thus maps to a constant phase offset. Next comes the Rømer delay (Rømer 1676), corresponding to the light-travel time (∼500 s) between the observatory at $\boldsymbol{r}^{\rm obs}$ and the SSB at $\boldsymbol{r}^{\rm SSB}$:

$$\Delta_{\rm R\odot} = t_a^{\rm obs} - t_a^{\rm SSB} = -\frac{(\boldsymbol{r}^{\rm obs}(t_a^{\rm obs}) - \boldsymbol{r}^{\rm SSB}(t_a^{\rm obs})) \cdot \hat{\boldsymbol{p}}}{c}, \quad (2)$$

where $\hat{\boldsymbol{p}}$ is the unit vector in the direction of the pulsar. In fact, $\hat{\boldsymbol{p}}$ is determined from pulsar-timing data mainly through the time dependence of Equation (2).

The errors $\delta\boldsymbol{r}^{\rm obs}$ and $\delta\boldsymbol{r}^{\rm SSB}$ induce systematic TOA delays according to Equation (2). Furthermore, the position of the observatory with respect to Earth's barycenter is known to few-centimeter (sub-nanosecond) accuracy (Edwards et al. 2006b), so we may write the systematic Rømer-delay error as

$$\delta t_{\rm R\odot} = -\frac{\delta\boldsymbol{x}^{(3)}(t_a^{\rm obs}) \cdot \hat{\boldsymbol{p}}}{c}, \quad (3)$$

where $\boldsymbol{x}^{(3)} \equiv \boldsymbol{r}^{(3)} - \boldsymbol{r}^{\rm SSB}$ is the terrestrial barycenter's position in the SSE frame where the SSB is identified with the origin, and $\delta\boldsymbol{x}^{(3)}(t)$ is the systematic error (a function of time) in the SSE's estimate of $\boldsymbol{x}^{(3)}(t)$. Here and below, we index the solar system planets from Mercury to Neptune from (1) through (8), so Earth is (3). We will also continue to use $\boldsymbol{x}$ to refer to vectors in the SSE frame, with origin at the SSB.

---

[44] This accounts for differences in the estimated semimajor axis of the Earth–Moon barycenter orbit, which gives rise to a linear rate in estimated ecliptic longitude.
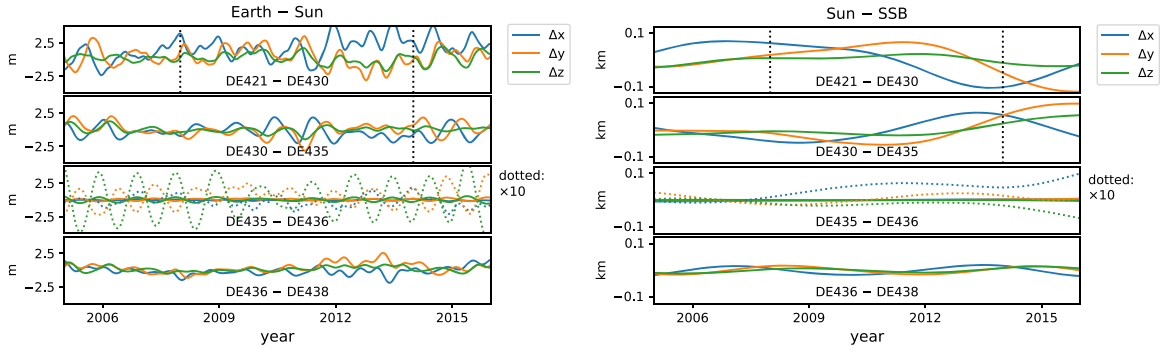
**Figure 2.** Differences in Earth–Sun position, $r^{(3)} - r^{\text{Sun}}$, and in Sun–SSB position, $r^{\text{Sun}} - r^{\text{SSB}}$, between pairs of JPL SSEs. Note the difference in the vertical scale of the two panels. Earth–SSB plots, not included here, would appear essentially the same as Sun–SSB plots, as the Earth–Sun differences are very small. For reference, 2.5 m ≈ 8 ns, and 0.1 km ≈ 300 ns, bracketing the range of amplitudes expected for GW signals in PTA data sets. We plot SSE equatorial-coordinate differences over the approximate span of NANOGrav's 11 yr data set, and we apply the best-fit frame rotation that minimizes the 3D norm of the difference. Dotted curves are zoomed ×10 vertically. Dotted vertical lines mark the end of the DE421 and DE430 fitted data sets. For clarity, we also remove a ∼100 m mean in each coordinate in the Sun–SSB differences; such a constant offset does not affect PTA likelihoods.

The time dependence of the SSE errors is essential to the effect of the ensuing residuals on GW searches in PTA data. Constant offsets, as well as linear and quadratic trends, are absorbed by redefinitions of the timing-model parameters that describe the intrinsic spin evolution of the pulsar; in other words, the TOA likelihood is affected only by $\delta t_{\text{R}\odot}$ minus the best-fitting quadratic polynomial.[45] Other free parameters in the timing model yield similar subtractions: most important, an angular misalignment between the SSE frame and the "true" sidereal frame can be absorbed by a slight coherent displacement in the position of all pulsars; the corresponding TOA corrections are sinusoidal with periods of one sidereal year.

PTAs are most sensitive to GWs with periods of a few to several years, on the order of the total time span of observations;[46] therefore, SSE errors in the orbits of the giant planets, which have comparable periods, could be mistaken for GWs. Correcting the orbit of a planet by the time-dependent vector $\delta r(t)$ corresponds to offsetting the SSB position by $(m^{\text{planet}}/m^{\text{SS}}) \times \delta r(t)$, where $m^{\text{SS}}$ is the total mass in the solar system. We then estimate

$$\delta x^{\text{SSB}}(\delta x^{(5)}) \approx 10^{-3} \times (50 \text{ km}) \approx 50 \text{ m} \approx 170 \text{ ns},$$
$$\delta x^{\text{SSB}}(\delta x^{(6)}) \approx (3 \times 10^{-4}) \times (50 \text{ km}) \approx 15 \text{ m} \approx 50 \text{ ns},$$
$$\delta x^{\text{SSB}}(\delta x^{(7,8)}) \approx (5 \times 10^{-5}) \times (5000 \text{ km}) \approx 250 \text{ m} \approx 800 \text{ ns},$$

$$(4)$$

where the quantities in nanoseconds are light-travel times equivalent to the distances. While the uncertainties due to Uranus and Neptune are larger, their orbital periods ($P^{(7)} = 84$ yr and $P^{(8)} = 165$ yr) ensure that the corresponding $\delta x^{\text{SSB}}$ appear as linear or mildly quadratic TOA trends, which are absorbed by timing models (and will continue to be, until PTA data sets approach a century in duration). By contrast, Jupiter and Saturn corrections enter the residuals with timescales ($P^{(5)} = 12$ yr and $P^{(6)} = 29$ yr) comparable to the

span of our data set—just where PTAs are most sensitive to GWs.

Likewise, the absolute location of the SSB is degenerate with the initial phase of the pulses for each pulsar, so we need not worry that the former depends strongly on the set of bodies that are included in each SSE fit. For instance, including trans-Neptunian objects relocates the SSB by ∼100 km (Folkner et al. 2014).

Champion et al. (2010) discuss pulsar timing's potential to constrain the masses of outer planets. The uncertainties in current IAU best estimates (derived from spacecraft tracking and Doppler studies: Jacobson et al. 2000, 2006; Jacobson 2014, 2009; IAU Division I Working Group on Numerical Standards for Fundamental Astronomy 2017) give rise to Rømer corrections comparable to the orbit errors. These scale as $(\delta m^{\text{planet}}/m^{\text{SS}}) \times r$, so

$$\delta x^{\text{SSB}}(\delta m^{(5)}) \approx (1.6 \times 10^{-11}) \times (5.2 \text{ au}) \approx 12 \text{ m} \approx 40 \text{ ns},$$
$$\delta x^{\text{SSB}}(\delta m^{(6)}) \approx (8.2 \times 10^{-12}) \times (9.4 \text{ au}) \approx 12 \text{ m} \approx 40 \text{ ns},$$
$$\delta x^{\text{SSB}}(\delta m^{(7)}) \approx (3.2 \times 10^{-11}) \times (19 \text{ au}) \approx 90 \text{ m} \approx 300 \text{ ns},$$
$$\delta x^{\text{SSB}}(\delta m^{(8)}) \approx (8.0 \times 10^{-11}) \times (30 \text{ au}) \approx 360 \text{ m} \approx 1.2 \text{ } \mu\text{s}.$$

$$(5)$$

These corrections yield delays with the same periods as the corresponding planets, so again they may be observable in PTA data sets for Jupiter and Saturn, but not for Uranus and Neptune.

In Figure 3, we show the difference between Rømer delays for 100 simulated pulsars randomly distributed across the sky, as computed with DE421/DE430/DE435/DE438 and with DE436, which is taken as a reference. Thus, we are plotting the systematic error that we would introduce using the other SSEs if DE436 gave exact orbits. We represent timing-model fits by subtracting the best-fitting quadratics and 1 yr-period sinusoids. The plot covers the span of the NANOGrav 11 yr data set. The differences reach ∼100 ns with typical periods of ∼10–12 yr (consistent with a Jovian attribution)—within the sensitivity range and band of GW searches. Indeed, as we see in Figure 1, the GW-amplitude posteriors for the NANOGrav 11 yr stochastic-background characterization are affected significantly by the choice of SSE.

---

[45] In a Bayesian context ,there is no single best-fitting polynomial when noise parameters are included in the inference, leading to varying weights for the fit. Nevertheless, the subtracted features remain largely irrelevant to parameter posteriors.
[46] This can be understood by noticing that the dominant source of residual noise ("radiometer" measurement noise) is white, but TOAs are sensitive to time integrals of GW strain, so the effective GW noise grows with frequency.
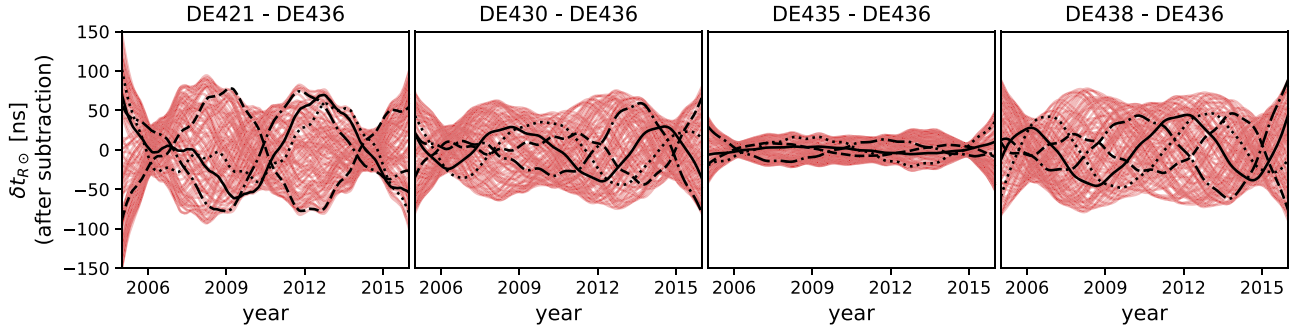
**Figure 3.** Differences in Rømer delays computed with DE421/DE430/DE435/DE438 and taking DE436 as reference for 100 simulated pulsars randomly distributed across the sky, plotted after subtracting the best-fitting quadratic and yearly sinusoid. The darker curves with the various dashed styles allow a comparison of the same pulsars across the three differences. The plot covers the span of the NANOGrav 11 yr data set. For any given pulsar, these differences are reduced to less than 10 ns after subtracting the best-fitting linear combination of the BAYESEPHEM partials, dotted into the pulsar position to obtain Rømer delays.

## 4. BAYESEPHEM: A Physical Model of Solar System Ephemeris Uncertainties for Gravitational-wave Searches in PTA Data

We set out to address the sensitivity of the NANOGrav stochastic-background analysis to SSE systematics by developing a parameterized physical model of SSE uncertainties (BAYESEPHEM), so that we can compute robust GW-parameter posteriors by marginalizing over the SSE parameters. Motivated by the discussion of Sections 2 and 3, we include the following components:

1. TOA delays generated by corrections to the masses of Jupiter, Saturn, Uranus, and Neptune, modeled as Champion et al. (2010):

$$\delta t_{R\odot}(\delta m^{(p)}) = -\frac{\delta m^{(p)}}{m^{SS}} \frac{\boldsymbol{x}^{(p)} \cdot \boldsymbol{p}}{c}. \tag{6}$$

We impose normal Bayesian priors on the $\delta m^{(p)}$, with standard deviation equal to the IAU-adopted mass-estimate uncertainties (IAU Division I Working Group on Numerical Standards for Fundamental Astronomy 2017). For each pulsar data set, we compute the $\boldsymbol{x}^{(p)}$ by evaluating the DE436 SSE at the measured TOAs $\{t_i\}$.

2. TOA delays generated by a rotation rate about the ecliptic pole (as needed to absorb Sun-to-Earth orbit differences among SSEs),

$$\delta t_{R\odot}(\omega^{\hat{z}}) = -\frac{(\mathsf{R}^{\hat{z}}(\delta\theta) \cdot \boldsymbol{x}^{(3)} - \boldsymbol{x}^{(3)}) \cdot \boldsymbol{p}}{c}, \tag{7}$$

where $\mathsf{R}^{\hat{z}}(\cdot)$ is the appropriately oriented rotation matrix and $\delta\theta = \omega^{\hat{z}}(t - t_0)$ is the rotation angle, with $\omega^{\hat{z}}$ the rate and $t_0$ set at the beginning of the NANOGrav data set. We impose uniform Bayesian priors on $\omega^{\hat{z}}$ that are commensurate with the rotation rates needed to reduce the difference between the JPL SSEs. Given that these rates are very small, we linearize Equation (7) with respect to $\omega^{\hat{z}}$; as for $\delta t_{R\odot}(\delta m^{(p)})$, we compute $\boldsymbol{x}^{(3)}$ by evaluating the DE436 at the measured TOAs $\{t_i\}$ for each pulsar data set.

Note that we do not expect this term to affect GW posteriors: both static and uniform rotations of the SSE frame are absorbed in the estimated positions and proper motions of the pulsars. We omit the former altogether and include $\omega^{\hat{z}}$ as a check.

3. TOA delays generated by perturbing Jupiter's average orbital elements, as given by

$$\delta t_{R\odot}(\delta a^\mu) = -\delta a^\mu \frac{m^{(5)}}{m^{SS}} \frac{\partial \boldsymbol{x}^{(5)}}{\partial a^\mu} \cdot \frac{\boldsymbol{p}}{c}, \tag{8}$$

where the six $a^\mu$ are the six J2000 Keplerian elements (semimajor axis, eccentricity, inclination, mean longitude, longitude of the perihelion, longitude of the ascending node) and the $\delta a^\mu$ are their perturbations. Alternatively, we can formulate the perturbations in terms of Brouwer and Clemence's "Set III" parameters (Brouwer & Clemence 1961), for which we have access to uncertainty estimates for certain JPL SSEs (Park et al. 2015; Folkner & Park 2018; see Figure 5). The two formulations are largely equivalent with respect to their effect on GW searches.

To implement the Rømer-delay perturbations, we begin with approximate values for the $a^\mu$ and their rates of change.[47] We vary these $a^\mu$ to minimize the (rms) difference between our quasi-Keplerian orbits and DE436, integrated between years 2000 and 2020. The resulting orbits are within 1% of DE436, which ensures similar accuracy for the orbit partials, more than enough for our purposes.

We compute the six partials as finite differences; because they are strongly correlated (see the top panel of Figure 4) and have different scales in their natural units, we decorrelate and normalize them by computing the singular value decomposition $\sum_\nu U_{\mu\nu} S_\nu V_{\nu i k}$ of the matrix $P_{\mu i k} = \partial x_k^{(5)}(t_i)/\partial a^\mu$, and adopting $\partial x_k^{(5)}(t_i)/\partial b^\nu \equiv V_{\nu i k}$ as new orbit partials (see bottom panel of Figure 4), where $\delta b^\nu = \sum_\mu \delta a^\mu M_{\mu\nu}$ with $M_{\mu\nu} = U_{\mu\nu} S_\nu$ (no summation intended). The resulting units are mixed. We give the orthonormalized coefficients $\delta b^\mu$ uniform priors that are broad enough to generate the range of Rømer variation seen across the JPL SSEs and to contain the support of the PTA likelihood for the NANOGrav 11 yr data set. In other words, we make the priors broad enough that the posteriors do not impinge on the boundaries.

We do not include Uranus' and Neptune's orbital perturbations, which currently lead to $\delta t_{R\odot}$ with linear

---

[47] The six $a^\mu$ plus the rate of change $\dot{a}^4 \equiv \dot{l}$ of the mean longitude specify Jupiter's orbit as the osculating ellipse at the J2000 reference epoch (MJD 2451545); the remaining five rates encode the secular evolution of Jupiter's orbit due to SSB bodies other than the Sun and to other physical effects. See https://ssd.jpl.nasa.gov/txt/aprx_pos_planets.pdf.
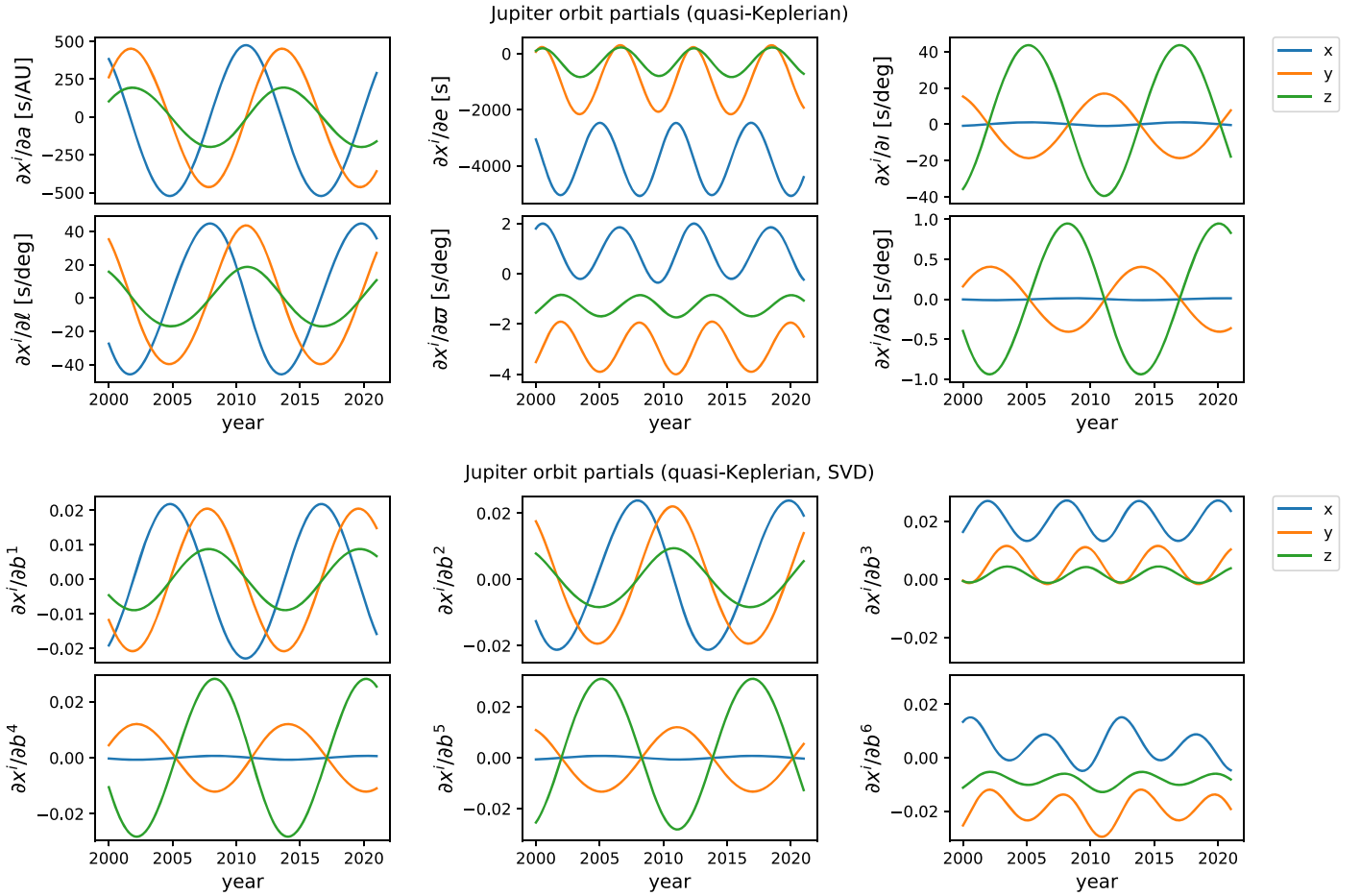
**Figure 4.** Top: perturbative partials ($x$, $y$, and $z$ coordinates, geometrized units of seconds) for Jupiter's orbit between the years 2000 and 2020, as obtained by varying the six Keplerian elements in a quasi-Keplerian model that includes their rates. The baseline elements and rates are adjusted to fit DE436 orbits. Here, $a$ is the semimajor axis of the orbit, $e$ the eccentricity, $\iota$ the inclination, $\ell$ the mean longitude, $\varpi$ the longitude of the periapsis, and $\Omega$ the longitude of the ascending node. Bottom: singular value decomposition vector time series obtained from the partials in the top panel.

time dependencies that are absorbed entirely by timing-model parameters. By contrast, we repeat the procedure that we just outlined for Saturn, but find that GW posteriors are barely affected for orbital perturbations of magnitude comparable to the differences among the JPL SSEs. Thus, in our use of BAYESEPHEM, we usually omit Saturn orbit perturbations. These may prove more important as data sets increase in length.

The components outlined above are brought together into a linear delay model $\delta t_{R\odot}(c^a)$ written as the product of the 11 dimensional correction vector $c^a \equiv \{\delta m^{(p)}, \omega^{\hat{z}}, \delta b^\nu\}$ (with $(p) = 5$, 6, 7, 8 and $\nu = 1$, ..., 6) by the $(n_{\text{TOAs}} \times 11)$ dimensional design matrix $G_{ia}$ with columns defined by the equations in this section. This model, including variational partials for Keplerian and Set III parameterizations, is available as part of the open-source software package ENTERPRISE (Ellis et al. 2019).

It is also possible to treat the linear model as a Gaussian process common to all pulsars and to marginalize analytically over the $c^a$, by building the effective correlation matrix $G_{ia}\Phi_{ab}G_{jb}$, where $\Phi_{ab}$ is the prior covariance of the $c^a$ (van Haasteren & Vallisneri 2014). The basis vector for each $c^a$ corresponds to the concatenation of the Rømer-delay perturbations that the parameter generates for each pulsar (i.e., the

projection of the same vector time series onto different $p$, at the appropriate TOAs).

## 5. Effect of BAYESEPHEM on the Search for Stochastic Gravitational Waves in NANOGrav's 11 yr Data Set

To derive the SSE-robust results reported in Arzoumanian et al. (2018a), we sampled the BAYESEPHEM corrections $c^a$ alongside the hyperparameters that describe the common GW background and the noise parameters for each pulsar (see Section 3.4 of Arzoumanian et al. 2018a and Sections 3 and 4 of Arzoumanian et al. 2016), treating $\delta t_{R\odot}(c^a)$ as a deterministic correction to the residuals.

In Figure 1, we show Bayesian posteriors for the stochastic-background amplitude $A_{\text{GWB}}$ at a fiducial frequency of $\text{yr}^{-1}$, as computed by fixing the fiducial SSE to DE421, DE430, DE435, DE436, and DE438 in turn, and either disabling (dotted lines) or applying BAYESEPHEM in the Set III formulation (solid lines). The prior on $\log_{10} A_{\text{GWB}}$ is flat in $[-18, -14]$, and the GWB spectral slope is set to $\gamma = 13/3$, as appropriate for an ensemble of binary inspirals progressing by GW emission alone. The posteriors follow from PTA likelihoods that omit Hellings–Downs correlations (as in "model 2A" rather than "3A" of Arzoumanian et al. 2018a), but results change only modestly if we include those. (Simulations show that as PTA data sets become more sensitive to GWs, a GWB
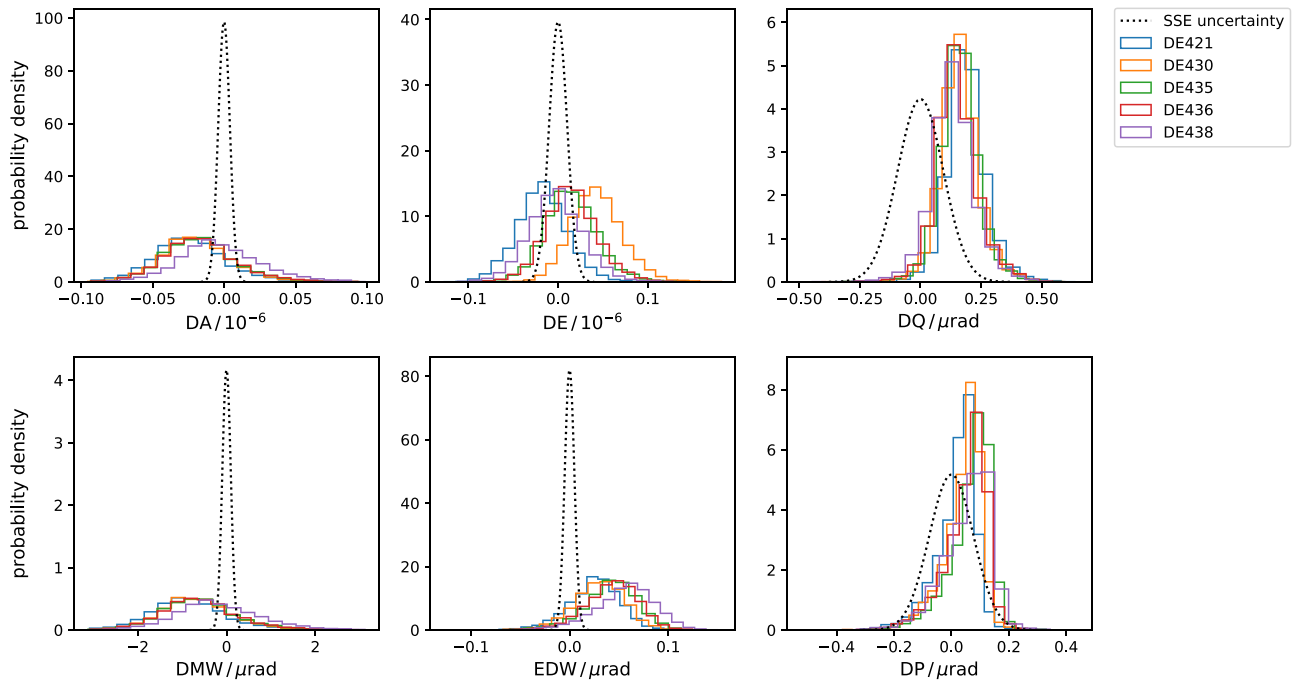
**Figure 5.** Bayesian posteriors for corrections to Jupiter's "Set III" orbital elements, as obtained in BAYESEPHEM's application to NANOGrav's 11 yr data set. The solid curves show posteriors obtained by correcting different fiducial SSEs; the dotted curve shows JPL's estimate of uncertainties in Jupiter's osculating orbital elements for DE435 and DE436 (Park et al. 2015). Parameter names follow JPL's orbit-determination package (Moyer 2003). Here, $DA \equiv \Delta a/a$, a fractional correction to the semimajor axis of the osculating orbit; $DE \equiv \Delta e$, a fractional correction to the eccentricity; $DMW = \Delta M_0 + \Delta w$, $EDW = e\Delta w$, $DP = \Delta p$, and $DQ = \Delta q$, where $M_0$ is the mean anomaly at the initial epoch (conventionally MJD 2440400.5) and $(\Delta p, \Delta q, \Delta w)$ encode a rigid rotation of the orbit, with $\Delta p$ and $\Delta q$ in the orbital plane and $\Delta w$ normal to it.

would manifest first as seemingly uncorrelated red-noise processes of comparable amplitude in multiple pulsars, and later as a Hellings–Downs-correlated process across the PTA, providing more conclusive evidence of the GW origin of the signal).

The plots in Figure 1 demonstrate the successful bridging of $A_{GWB}$, our goal in mitigating the systematic effects of SSE errors. In doing so, BAYESEPHEM removes hints that any GWB is present, as confirmed by computing model 2A Bayes factors (in favor of a common GWB process), listed in Table 1. In Table 2, we show 95% $A_{GWB}$ upper limits, computed as just discussed but with flat uninformative priors on $A_{GWB}$. The tables (as well as plots analogous to Figure 1 not shown here) confirm that the Set III and Keplerian-element formulations are equivalent with respect to GW detection.

The posterior distributions of the BAYESEPHEM mass perturbations are identical to their IAU priors: NANOGrav's 11 yr data set is uninformative compared to current estimates from spacecraft data. Likewise, the frame rotation rate $\omega \hat{z}$ fills its prior, as expected. The posterior distributions of the BAYESEPHEM corrections to Jupiter's orbital elements are shown in Figure 5, where they are compared to JPL's estimated uncertainties for DE435 and DE436, derived by comparing SSE fits that use independent subsets of the data (Park et al. 2015). The NANOGrav posteriors appear consistent with all SSEs: they have reasonably high support around the estimated-uncertainty region around $\delta a^\mu = 0$. The least-consistent parameters are those involving rotations of the orbital plane (DMW and EDW; see the caption of Figure 5), as well as the eccentricity (DE) for the oldest SSEs, DE421 and DE430. JPL's estimated uncertainties for DE438 are a factor of $\sim 4$ tighter than those for DE435/6, but they do not change this picture substantially.

The large dispersion of the BAYESEPHEM posteriors is not unexpected, given that pulsar-timing data is sensitive to Rømer delays in a rather selective fashion (see the discussion of Section 3). In terms of uncertainties on Jupiter's instantaneous position, the BAYESEPHEM posteriors map to rms 3D errors $\sim 100$ km, and thus are larger than the differences between SSEs. While BAYESEPHEM can bridge NANOGrav 11 yr $A_{GW}$ posteriors for different SSEs, the resulting orbital-element posteriors do not identify any specific systematic offset among them.

## 6. Assessing Gravitational-wave-background Detection Prospects when Using BAYESEPHEM

BAYESEPHEM is designed to model SSE uncertainties and systematics, bridging published estimates of Earth's trajectory around the SSB. To do so, BAYESEPHEM introduces new parameters that govern a spatially correlated process of amplitude comparable to the stochastic GWBs that we seek. While SSE corrections and the GWB have different spatial-correlation structures, the two may nevertheless remain degenerate when probed by a limited number of PTA pulsars (Roebber 2019). It is then natural to ask how the application of BAYESEPHEM may affect GWB-detection prospects (sensitivity and time to detection) in the weak-to-intermediate signal-to-noise regime in which spatial correlations among the pulsars carry marginal information. To sketch an answer to this question, we conducted a set of simulations, which are summarized briefly in Arzoumanian et al. (2018a) and discussed further here.

We created multiple synthetic data sets meant to replicate the sensitivity of NANOGrav's 11 yr (really, 11.4 yr) data. We used actual observation epochs for the 34 analyzed pulsars and
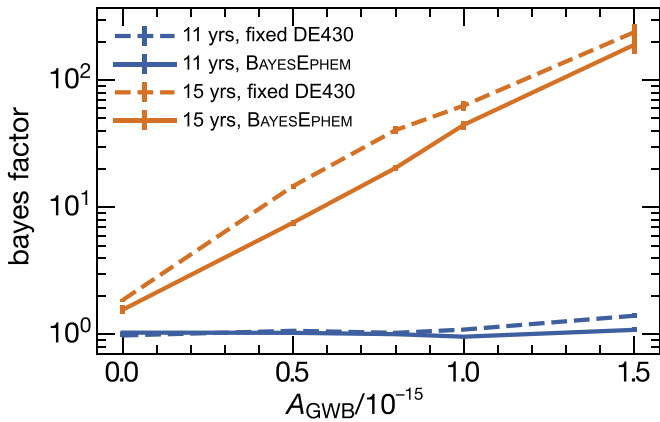
**Figure 6.** Correlated vs. uncorrelated GWB Bayes factors (i.e., model 3A vs. model 2A), estimated for a set of simulated 15 yr data sets with GWB injections at increasing amplitudes. The data sets were generated by adopting DE436 as the "correct" SSE, and analyzed using both the "wrong" DE430 ephemeris (dashed lines) and BAYESEPHEM (solid). The orange and blue lines display results for the full data sets and their 11 yr "slices," respectively. Their comparison indicates that BAYESEPHEM will not impede the ability of PTAs to make a definitive detection in the near future. Figure adapted from Arzoumanian et al. (2018a).

extended the time span to 15 yr by drawing observation times from distributions fit to the last three years of measured data. We simulated timing residuals by drawing random white- and red-noise deviates at the maximum a posteriori 11 yr levels, and again extrapolated to 15 yr from the last 3 of the 11 years. These choices ensure that future data taking is represented with cadence and precision comparable to the end of 2015—a very conservative option given ongoing receiver upgrades and the continuous addition of new pulsars to the PTA.

We calibrated these noise-only simulations by increasing noise levels until the 11 yr "slice" of the data matched the GWB upper limit (specifically the spatially uncorrelated model 2A, with uncorrected DE436) of Arzoumanian et al. (2018a). All simulated data sets were created by postulating that DE436 was the "truth." We then injected $\gamma = 13/3$ GWBs of various amplitudes and analyzed both the 11 yr slices and the full 15 yr data sets by taking DE430 as our fiducial SSE, both with and without BAYESEPHEM.

Our results were as follows: first, for noise-only data sets without BAYESEPHEM, the systematic offset between DE430 and DE436 is interpreted as a GWB for both 11 yr and 15 yr time spans, with (model 2A versus the noise-only model 1) Bayes ratios of ∼2 and ∼20 respectively. The ratios are reduced to levels consistent with noise fluctuations by the application of BAYESEPHEM.

Second, as we increase the amplitude of injected GWBs while applying BAYESEPHEM, the Bayes factors remain marginal for the 11 yr data set, as do model 3A versus model 2A factors, the substantive spatial-correlation test of GW presence. The latter are shown in Figure 6. The story is different for 15 yr data sets, where the scaling of Bayes factors with injected GWBs is comparable whether or not we apply BAYESEPHEM. Indeed, the longer time span disentangles SSB and GWB correlations, enabling detection even at the astrophysically conservative levels described by Sesana and colleagues (Sesana et al. 2016).

In summary, our simulations confirm that SSE errors can produce spurious evidence for GWBs; they suggest that, in data sets similar to NANOGrav's 11 yr, BAYESEPHEM can

overcorrect these biases, suppressing the evidence for a "true" GWB, but they imply also that this effect vanishes for longer data sets. Thus, we expect that BAYESEPHEM will not impair detection prospects in the near future—if of course nature grants us a sufficient GWB bounty.

## 7. Other Modeling Approaches

In our analysis of NANOGrav's 11 yr data set, we experimented with other modeling approaches for SSE uncertainties, which (in our implementation) did not satisfy our bridging criterion or were otherwise disfavored. We discuss them briefly here, as a reference for future investigations:

1. Planetary mass perturbations for outer-solar system planets: modeled as in Champion et al. (2010, and as in the first element of BAYESEPHEM; see Section 4), these did not affect our $A_{GWB}$ posteriors significantly (and therefore did not resolve the discrepancy among SSEs), whether introduced with best-estimate priors (IAU Division I Working Group on Numerical Standards for Fundamental Astronomy 2017) or with more relaxed assumptions.

2. Dipole-correlated Gaussian process: as proposed in Tiburzi et al. (2016), the Rømer delays for individual pulsars may be treated as Gaussian processes with common power-law or free-spectral priors (van Haasteren & Vallisneri 2014), and with dipolar spatial correlations between pulsars ($C_{ij} = \cos \theta_{ij}$, with $\theta_{ij}$ the angle between pulsars). We note that this approach is equivalent to modeling the apparent motion of the SSB (or equivalently, the error in Earth's orbit) as three Gaussian processes (along the three spatial axes) with identical uncorrelated priors, and then projecting the vector time series onto the pulsar positions to obtain Rømer delays.

   Dipole-correlated Gaussian processes were included in models 2B, 2C, 3B, and 3C of our 11 yr analysis (Arzoumanian et al. 2018a). It is unclear to us why the approach failed to bridge $A_{GWB}$ posteriors across SSEs, but the reason may be related to the choice of Gaussian-process priors appropriate to describe the range and shapes of variations among SSEs.

3. Rømer mixture: this phenomenological model describes Earth's orbit as a linear combination of its estimate in multiple SSEs, with mixture coefficients constrained by a Dirichlet prior (see, e.g., Gelman et al. 2013). This approach achieves bridging by construction, but it is difficult to interpret physically. In our 11 yr investigation, the posteriors of the mixture coefficients indicated a moderate preference for DE435/6 over DE421 and DE430.

4. Gaussian process based on numerical partials: in this approach, SSE corrections are modeled as a finite Gaussian process (Williams & Rasmussen 2006) in which basis vectors are given by orbit partials (e.g., the change in Earth's orbit as we vary the initial conditions for all the planets), as computed by numerical integration for SSE fits. The priors for the basis weights are given by the formal covariance of the SSE fit parameters. The basis vectors are then projected into Rømer delays for each pulsar.

   It turns out that only Jupiter and Saturn partials matter to GW results; when restricted to these planets, the

approach is effectively equivalent to the orbit-correction sector of BAYESEPHEM—at least for the 11 yr time span, for which numerical partials are very close to analytical osculating-orbit perturbations.

## 8. Conclusion

It is striking that the abundance and precision of current PTA data sets should be such that our sensitivity to GWs is limited by the very accuracy to which we can position Earth around the SSB. In the analysis of NANOGrav's 11 yr data set (Arzoumanian et al. 2018a), we took the conservative position that robust GW upper limits and detection Bayes factors must be reproducible with all SSEs released in the last 10 years, after introducing a sufficiently descriptive model of SSE uncertainties. To this end, we developed BAYESEPHEM, described in this article, which focuses on the SSE degrees of freedom (Jupiter's orbital elements) that measurably affect our GW search, and which produces "bridged" $A_{GWB}$ posteriors, upper limits, and Bayes factors. Because of this focus, the analysis of NANOGrav data does not update the JPL SSEs significantly (see Figure 5), although it provides uncertainty estimates entirely independently of those offered by SSE makers (Park et al. 2015; Folkner & Park 2018). We expect that it would be possible to concentrate instead on a sector of SSE corrections that do not affect GW results, but that would benefit from PTA constraints, thus producing PTA-enhanced SSEs useful beyond GW detection. For this purpose, we could adopt either the orbital-elements formalism of Section 4 or the numerical-partials approach of Section 7.

The conservative modeling attitude employed for NANOGrav's 11 yr analysis comes at the cost of a loss of GW sensitivity, as described in Section 6 (see also Roebber 2019). The GW statistics reported in Arzoumanian et al. (2018a; and here in Tables 1 and 2) are supported in our Bayesian setting, in which we decline to favor one SSE above others. However, these results should not be considered binding for future GW searches that rely on demonstrably accurate SSEs and that are based on longer data sets where SSE and GW correlations become disentangled: more precisely, data sets that cover a longer time span with a sufficient number of high-quality pulsars. Both conditions are now materializing: Juno's ongoing measurements are improving estimates of Jupiter's orbit (Folkner & Park 2018), which (we argue) is the limiting factor for GW searches, and the NANOGrav data set is progressing toward the 15 yr span, for which (we reckon) Jovian systematics decouple from GW statistics. The combined data sets assembled by the International Pulsar Timing Array (Perera et al. 2019) have already passed this mark and thus may already be immune to this problem.

The path toward an authoritative detection of low-frequency GWs with PTAs requires intense and persistent timing campaigns on the world's most sensitive radio telescopes; sophisticated inference techniques and clever high-performance-computing algorithms, to make sense of ever-growing, heterogeneous data sets with many unknown parameters; and a confident control of TOA systematics at the nanosecond level. Among the last, the error analysis and validation of high-precision SSEs will remain paramount.

### ORCID iDs

M. Vallisneri ● https://orcid.org/0000-0002-4162-0033
S. R. Taylor ● https://orcid.org/0000-0003-0264-1453
J. Simon ● https://orcid.org/0000-0003-1407-6607
W. M. Folkner ● https://orcid.org/0000-0001-5133-9934
R. S. Park ● https://orcid.org/0000-0001-9896-4585
C. Cutler ● https://orcid.org/0000-0002-2080-1468
J. A. Ellis ● https://orcid.org/0000-0003-1715-974X
S. J. Vigeland ● https://orcid.org/0000-0003-4700-9072
K. Aggarwal ● https://orcid.org/0000-0002-2059-0525
P. T. Baker ● https://orcid.org/0000-0003-2745-753X
A. Brazier ● https://orcid.org/0000-0001-6341-7178
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
H. T. Cromartie ⓘ https://orcid.org/0000-0002-6039-692X
K. Crowter ⓘ https://orcid.org/0000-0002-1529-5169
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. C. Ferrara ⓘ https://orcid.org/0000-0001-7828-7708
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
J. S. Hazboun ⓘ https://orcid.org/0000-0003-2742-3321
A. M. Holgado ⓘ https://orcid.org/0000-0003-4143-8132
E. A. Huerta ⓘ https://orcid.org/0000-0002-9682-3604
K. Islo ⓘ https://orcid.org/0000-0001-5123-9917
R. Jennings ⓘ https://orcid.org/0000-0003-1082-2342
M. L. Jones ⓘ https://orcid.org/0000-0001-6607-3710
D. L. Kaplan ⓘ https://orcid.org/0000-0001-6295-2881
L. Z. Kelley ⓘ https://orcid.org/0000-0002-6625-6450
J. S. Key ⓘ https://orcid.org/0000-0003-0123-7600
M. T. Lam ⓘ https://orcid.org/0000-0003-0721-651X
L. Levin ⓘ https://orcid.org/0000-0002-2034-2986
D. R. Lorimer ⓘ https://orcid.org/0000-0003-1301-966X
J. Luo ⓘ https://orcid.org/0000-0001-5373-5914
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. K. Swiggum ⓘ https://orcid.org/0000-0002-1075-3837
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

Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2016, ApJ, 821, 13
Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2018a, ApJ, 859, 47
Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2018b, ApJS, 235, 37
Brouwer, D., & Clemence, G. M. 1961, Methods of Celestial Mechanics (New York: Academic Press)
Caballero, R. N., Guo, Y. J., Lee, K. J., et al. 2018, MNRAS, 481, 5501
Champion, D. J., Hobbs, G. B., Manchester, R. N., et al. 2010, ApJ, 720, L201
Desvignes, G., Caballero, R. N., Lentati, L., et al. 2016, MNRAS, 458, 3341
Detweiler, S. 1979, ApJ, 234, 1100
Dickey, J. M. 1971, The Annals of Mathematical Statistics, 42, 204

Edwards, R. T., Hobbs, G. B., & Manchester, R. N. 2006a, MNRAS, 372, 1549
Edwards, R. T., Hobbs, G. B., & Manchester, R. N. 2006b, MNRAS, 372, 1549
Ellis, J. A., Vallisneri, M., Taylor, S. R., & Baker, P. T. 2019, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE, Astrophysics Source Code Library, ascl:1912.015
Fey, A. L., Gordon, D., Jacobs, C. S., et al. 2015, AJ, 150, 58
Fienga, A., Manche, H., Laskar, J., Gastineau, M., & Verma, A. 2014, arXiv:1405.0484
Folkner, W. M., & Park, R. S. 2015, Uncertainty in the orbit of Jupiter for Juno planning, Tech. Rep. IOM 343R-15-019, Pasadena, CAJet Propulsion Laboratory
Folkner, W. M., & Park, R. S. 2016, JPL planetary and Lunar ephemeris DE436 https://naif.jpl.nasa.gov/pub/naif/JUNO/kernels/spk/de436s.bsp.lbl
Folkner, W. M., & Park, R. S. 2018, Planetary ephemeris DE438 for Juno, Tech. Rep. IOM 392R-18-004, Pasadena, CAJet Propulsion Laboratory
Folkner, W. M., Park, R. S., & Jacobson, R. A. 2016, Planetary ephemeris DE435, Tech. Rep. IOM 392R-16-003, Pasadena, CAJet Propulsion Laboratory
Folkner, W. M., Williams, J. G., & Boggs, D. H. 2009, IPNPR, 178, 1
Folkner, W. M., Williams, J. G., Boggs, D. H., Park, R. S., & Kuchynka, P. 2014, IPNPR, 42, 1
Foster, R. S., & Backer, D. C. 1990a, ApJ, 361, 300
Foster, R. S., & Backer, D. C. 1990b, ApJ, 361, 300
Gelman, A., Carlin, J. B., Stern, H. S., et al. 2013, Bayesian Data Analysis (Boca Raton, FL: Chapman and Hall/CRC)
Guo, Y. J., Li, G. Y., Lee, K. J., & Caballero, R. N. 2019, MNRAS, 489, 5573
Hees, A., Folkner, W. M., Jacobson, R. A., & Park, R. S. 2014, PhRvD, 89, 102002
Hellings, R. W., & Downs, G. S. 1983, ApJL, 265, L39
Hobbs, G. 2013, CQGra, 30, 224007
Hobbs, G., & Edwards, R. 2012, Tempo2: Pulsar Timing Package, Astrophysics Source Code Library, ascl:1210.015
IAU Division I Working Group on Numerical Standards for Fundamental Astronomy 2017, Astronomical Constants: Current Best Estimates http://maia.usno.navy.mil/NSFA/NSFA_cbe.html
Jacobson, R. A. 2009, AJ, 137, 4322
Jacobson, R. A. 2014, AJ, 148, 76
Jacobson, R. A., Haw, R. J., McElrath, T. P., & Antreasian, P. G. 2000, JAnSc, 48, 495
Jacobson, R. A., Antreasian, P. G., Bordi, J. J., et al. 2006, AJ, 132, 2520
Lommen, A. N., & Demorest, P. 2013, CQGra, 30, 224001
Lorimer, D. R., & Kramer, M. 2012, Handbook of Pulsar Astronomy (Cambridge: Cambridge Univ. Press)
Ma, C., Arias, E. F., Eubanks, T. M., et al. 1998, AJ, 116, 516
McLaughlin, M. A. 2013, CQGra, 30, 224008
Moyer, T. D. 2003, Formulation for Observed and Computed Values of Deep Space Network Data Types for Navigation (New York: Wiley)
Newhall, X. X., Standish, E. M., & Williams, J. G. 1983, A&A, 125, 150
Park, R. S., Folkner, W. M., & Jacobson, R. A. 2015, The planetary ephemeris DE434, Tech. Rep. IOM 392R-15-018, Pasadena, CAJet Propulsion Laboratory
Perera, B. B. P., DeCesar, M. E., Demorest, P. B., et al. 2019, MNRAS, 490, 4666
Roebber, E. 2019, ApJ, 876, 55
Rømer, O. C. 1676, Journal des Sçavans, 11, 233
Sazhin, M. V. 1978, SvA, 22, 36
Sesana, A., Shankar, F., Bernardi, M., & Sheth, R. K. 2016, MNRAS, 463, L6
Shannon, R. M., Ravi, V., Lentati, L. T., et al. 2015, Sci, 349, 1522
Tiburzi, C., Hobbs, G., Kerr, M., et al. 2016, MNRAS, 455, 4339
van Haasteren, R., & Vallisneri, M. 2014, PhRvD, 90, 104012
Verbiest, J. P. W., Lentati, L., Hobbs, G., et al. 2016, MNRAS, 458, 1267
Verma, A. K. 2013, PhD thesis, Université de Franche-Comté, Besançon, France https://tel.archives-ouvertes.fr/tel-01124331/document
Viswanathan, V., Fienga, A., Minazzoli, O., et al. 2018, MNRAS, 476, 1877
Williams, C. K., & Rasmussen, C. E. 2006, Gaussian Processes for Machine Learning (Cambridge, MA: MIT press)