Pulse profile modeling of thermonuclear burst oscillations - I

The effect of neglecting variability


DOI
10.1093/mnras/stad1030

Publication date
2023

Document Version
Final published version

Published in
Monthly Notices of the Royal Astronomical Society

License
CC BY

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
Pulse profile modelling of thermonuclear burst oscillations — I. The effect of neglecting variability

Yves Kini, Tuomo Salmi, Anna L. Watts, Serena Vinciguerra, Devarshi Choudhury, Slavko Bogdanov, Zach Meisel and Valery Suleimanov

1Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, Amsterdam 1090GE, the Netherlands
2Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA
3Institute of Nuclear & Particle Physics, Department of Physics & Astronomy, Ohio University, Athens, OH 45701, USA
4Institut für Astronomie und Astrophysik, Kepler Center for Astro and Particle Physics, Universität Tübingen, Sand 1, Tübingen D-72076, Germany

Accepted 2023 March 31. Received 2023 March 30; in original form 2023 February 27

ABSTRACT
We study the effects of the time-variable properties of thermonuclear X-ray bursts on modelling their millisecond-period burst oscillations. We apply the pulse profile modelling technique that is being used in the analysis of rotation-powered millisecond pulsars by the Neutron Star Interior Composition Explorer to infer masses, radii, and geometric parameters of neutron stars. By simulating and analysing a large set of models, we show that overlooking burst time-scale variability in temperatures and sizes of the hot emitting regions can result in substantial bias in the inferred mass and radius. To adequately infer neutron star properties, it is essential to develop a model for the time-variable properties or invest a substantial amount of computational time in segmenting the data into non-varying pieces. We discuss prospects for constraints from proposed future X-ray telescopes.

Key words: dense matter – equation of state – X-rays: binaries – X-rays: bursts.

1 INTRODUCTION

Neutron stars (NSs) are the densest directly observable objects in the universe, which makes their cores the ideal environment to probe the properties of matter at high density and low temperature. Probing the matter properties in the core can be achieved by measuring their mass and radius, which depend on the dense matter Equation of State (EoS) (see e.g. Lattimer 2012; Oertel et al. 2017; Baym et al. 2018; Tolos & Fabbietti 2020; Yang & Piekarewicz 2020; Hebeler 2021).

For X-ray pulsars, mass and radius measurement is possible through analysis of X-ray pulsations that emanate from their surface. These coherent X-ray pulsations are produced by hot spots as they rotate in and out of view as the star spins. The intense gravitational field experienced by photons as they leave the stellar surface, and the fast rotation of the star, shape the characteristics of the observed pulsations. Therefore, by using pulse profile modelling, a technique that combines special and general relativity, it is possible to infer the stellar mass and radius in addition to other key parameters (see e.g. Bogdanov et al. 2019; Watts 2019, and references therein).

To successfully infer the stellar properties, arriving photons from an actual source are first binned by energy and rotational phase, according to an ephemesis, to produce an observed pulse profile. Then a simulated pulse profile for a given set of input parameters is generated using mathematical models that take into account the relevant physical phenomena and parameters, such as the pulsar’s rotation, magnetic field, and the emission and propagation of electromagnetic radiation. Next, the simulated pulse profile is compared to observations of the pulsar’s actual pulse profile, using statistical and other analysis techniques to assess the degree of agreement between the two. Finally, the model and input parameters are adjusted if necessary, based on the comparison between the simulated and observed pulse profiles, in order to improve the accuracy of the model. This process is then repeated until the simulated pulse profile closely matches the observed pulse profile and gives a good picture of the physical processes that cause pulsar emission.

Recently, the pulse profile modelling technique has been used to place stringent parameter constraints on some of the Neutron Star Interior Composition Explorer (NICER; Gendreau et al. 2016) primary targets; PSR J0030+0451 (Miller et al. 2019; Riley et al. 2019) and PSR J0740+6620 (Miller et al. 2021; Riley et al. 2021; Salmi et al. 2022). NICER has constrained not only the masses and radii of these Rotation-Powered Millisecond Pulsars (RMPs), but also the hot spot properties and by extension the magnetic field geometry (Bilous et al. 2019; Chen, Yuan & Vasilopoulos 2020; Kalapotharakos et al. 2021). However, pulse profile modelling can also be applied to Accretion-Powered Millisecond Pulsars (AMPs) and Thermonuclear Burst Oscillation (TBO) sources (see for example Watts et al. 2016). The latter are the primary focus of this paper.

First observed in 1976 (Belian, Conner & Evans 1976; Gridlay et al. 1976). Type I X-ray bursts are thermonuclear runaway caused by unstable burning of the accreted plasma onto the NS surface from its companion. This unstable burning occurs under specific conditions of accretion rate and chemical composition of the accreted material (Bildsten 1998; Galloway & Keek 2020; Galloway et al. 2020). The typical duration of the most common bursts varies from a few to several hundred seconds. Their light curves exhibit a fast rise in...
luminosity followed by an exponential decay. Their recurrence time ranges from a few minutes to days and the energy released during a burst is typically $10^{39} - 10^{40}$ ergs. Two decades after the first burst was discovered, coherent pulsations, commonly known as thermonuclear burst oscillations or simply burst oscillations, were found in some of the burst light curves of 4U 1728–34 (Strohmayer et al. 1996). These coherent pulsations have now been firmly confirmed in nearly 20 per cent of thermonuclear bursters (Bilous & Watts 2019).

All known TBO sources have millisecond periods, with the exception of IGR J17480–2446 (~11 Hz; Altamirano et al. 2010; Strohmayer & Markwardt 2010). Their spin rates are typically between ~ 250 Hz and ~ 600 Hz. Rapid rotation enhances a variety of effects, namely, the Doppler boost of photon energy, the angular distribution aberration, and the time delay between photons emitted at different phases. These effects introduce asymmetry and harmonic content in the observed pulse profile, which combined with the compactness information gained from the pulse amplitude, can be utilized to independently determine the mass and the radius through pulse profile modelling (e.g. Braje, Romani & Rauch 2000; Poutanen & Beloborodov 2006; Lo et al. 2013; Psaltis, Özel & Chakrabarty 2014, for discussion). A benefit of using TBO sources is that the inferred mass and radius can be independently cross-checked using other methods, for example through pulse profile modelling of the accretion pulsedions of AMPs (e.g. Salmi, Nättilä & Poutanen 2018) or by modelling the burst spectra (e.g. Suleimanov et al. 2011b; Özel 2013; Steiner, Lattimer & Brown 2013; Guillot & Rutledge 2014; Nättilä et al. 2016; Steiner et al. 2018).

Still, attempting to apply pulse profile modelling to burst oscillation sources comes with multiple challenges. First, where exactly these pulsations originate from is still up for debate (see Section 2.1). Models for surface patterns are required but the uncertainties associated with them are likely to influence the inferred source properties (Watts 2012; Bhattacharyya 2022). The second challenge stems from the burst’s inherent variability, as well as the variability associated with the oscillations. Bursts are by definition variable, as temperature evolution due to thermonuclear burning and then cooling drives the fast increase and then slower decrease in X-ray flux. The presence of oscillations may however lead to additional sources of variability in the physical emission model (e.g. relating to hot spot motion or size). If not correctly addressed, these variabilities could lead to biased inferred parameters. The third difficulty associated with bursters is the low rms fractional amplitude (FA). This increases the already wide number of parameter degeneracies intrinsic to pulse profile modelling (see e.g. Lo et al. 2013; Psaltis et al. 2014). The final obstacle comes from the bursting NS atmosphere. X-ray photons from the stellar surface of the star are reprocessed in the atmosphere and this must be accounted for. The current state-of-the-art atmosphere models (Suleimanov, Poutanen & Werner 2012, 2018) for TBO sources make assumptions about the composition of the atmosphere, among other things. Since the actual composition of the atmosphere is not known, these assumptions leave room for modelling uncertainty. Although applying pulse profile modelling to TBO sources is rendered challenging by the aforementioned factors, it may not only shed light on the fundamental interactions taking place in the NS cores, but also provide a unique opportunity to explore the climate and weather of bursters and study the physics of nuclear burning.

The purpose of this paper is to determine whether time variation in temperature and/or emitting region radius during the burst can be neglected, which would reduce computational costs in analysis. Therefore, we make use of simulations to assess how well parameters can be recovered, adapting the X-ray Pulse Simulation and Inference (X-PSI; Riley et al. 2023b) pipeline developed for NICER analysis to model burst oscillations. Our simulations must account for the lack of knowledge of the spot pattern, which introduces a lot of uncertainty. To mitigate this, we use cutting-edge nuclear burning models and tailor our analysis to the best source in this class, XTE J1814–338 (hereafter J1814; Strohmayer et al. 2003; Bhattacharyya et al. 2005).

Our analysis proceeds in two stages. First, using pulse profile simulation and assuming reasonable possible surface patterns, we construct two sets of synthetic data; with and without time variation in temperature and/or emission region radius. We then perform inference runs on the various sets of synthetic data, ignoring the temporal variation for both sets. Thus, we structure this paper as follows. First, we describe in Section 2, our modelling approach, assumptions, and synthetic data generation method. Then, we present the results of our analysis in Section 3. Finally, we discuss the implications for using pulse profile modelling for TBO sources in Section 4.

2 METHOD

Photons contained in X-ray pulsations captured by a telescope carry fundamental information about their emitting sources, interactions in the various media they traversed, and the method by which they were recorded (Pechenick, Paclas & Cohen 1983; Chen & Shaham 1989; Page 1995; Miller & Lamb 1998; Braje et al. 2000; Weinberg, Miller & Lamb 2001; Beloborodov 2002; Poutanen & Beloborodov 2006; Cadeau et al. 2007;orsink et al. 2007; Bauböck et al. 2012; Lo et al. 2013; Psaltis et al. 2014; Miller & Lamb 2015; Stevens et al. 2016; Nättilä & Pihajoki 2018; Bogdanov et al. 2019). Retrieving this information through pulse profile modelling requires: a surface pattern model, an atmosphere model, a model for the interstellar medium (ISM) extinction and the instrument response. In addition, a mathematical model of the NS’s external gravitational field and relativistic effects are required to model the propagation of the emitted photons. We employ the oblate-star Schwarzschild spacetime plus Doppler approximation (Cadeau et al. 2007;orsink et al. 2007), which is adequate for the spin rates of known TBO sources, and use it for relativistic ray tracing. For the instrument response, we make use of the Rossi X-Ray Timing Explorer (RXTE), Proportional Counter Array (Jahoda et al. 1996) instrument response file. All the surface patterns, relativistic ray tracing, and accounting for the ISM, the NS atmospheric interactions and the instrument response are done using our publicly available simulation and inference package, X-PSI (Riley et al. 2023b), version v0.7.9. X-PSI uses the Bayesian inference algorithm MULTINEST (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009; Feroz et al. 2019) to explore the large parameter space and infer the source properties.

2 We used the response matrix of burst 10 of J1814 (ObsID: 80418-01-02-00, MID: 52803.7406). All the five proportional counter units were included when generating the response file.
3 https://github.com/xpsi-group/xpsi.git
4 During sampling, we slightly modified the eval_jmarginalLikelihood function to bypass the check that determines how many sigmas below the signal from the star the data can be (at any channel) before skipping the precise likelihood computation. This requirement was deemed too restrictive for bursting sources. Some X-PSI version 1.0.0 (Riley et al. 2023a) features were also used for postprocessing.

1 Superbursts, which are believed to be caused by carbon combustion, can last for several hours and generate energy of up to $10^{42}$ ergs (see e.g. in’t Zand 2017, for review).
In this section, we discuss all of the modelling components, from photon emission to the instrument. We begin by describing possible pulsation origins, our atmosphere and ISM assumptions. Since our pulse profiles include a background component, we explain how this was incorporated into the modelling. Finally, we discuss the prior assumptions required for the Multinest Bayesian framework.

2.1 Oscillation origins and phenomenological models

For RMPs, pulsations are caused by hot spots induced by return currents at the magnetic poles (Harding &Muslimov 2001, 2002). Despite the well-known origin of the pulsations, the resulting surface pattern is uncertain (Gralla, Lupasca &Philippov 2017; Philippov &Spitkovsky 2018), resulting in temperature distributions being parametrized with simple models in NICER’s RMP analyses' (Miller et al. 2019; Riley et al. 2019; Miller et al. 2021; Riley et al. 2021; Salmi et al. 2022).

For TBO sources however, the origin of the pulses is still unclear (Watts 2012; Bhattacharyya 2022). In the presence of a high magnetic field and/or rapid rotation, fuel and/or flame confinement may cause localized burning (Cavecchi et al. 2011). This confined burning would result in a hotter region, which would explain the observed oscillations during the rise and the decay of the burst. Yet, not all sources have a significant magnetic field or sufficiently rapid rotation to confine the flame for long. For those sources, the flames would spread from their ignition points, inducing oscillations during the rise (Strohmayer, Zhang &Swank 1997). In this scenario, the amplitude of the pulsation and their presence in the tail of the burst are more difficult to explain. The explanation may potentially come from global modes, or more precisely, buoyant r-modes (Heyl 2004; Lee 2004; Piro &Bildsten 2005; Chambers et al. 2019; Chambers &Watts 2020), a cooling wake (Mahmoodifar &Strohmayer 2016), convective patterns (Garcia, Chambers &Watts 2019), or the dynamics of vortices (Spitkovsky, Levin &Ushomirsky 2002; Cavecchi &Spitkovsky 2019).

Given the uncertainties related to oscillation origins and hence the ‘true’ form of the surface pattern, we employ a phenomenological approach, similar to NICER’s RMP analysis, and tailor our analysis to J1814. J1814 is one of the best sources for pulse profile modelling due to its very stable pulsations (very little frequency drift), relatively high FA, and harmonic content in the pulsations during and between bursts (Strohmayer et al. 2003; Bhattacharyya et al. 2005; Watts &Strohmayer 2006). The stable pulsations help to minimize the complexity involved with the TBO source. Furthermore, high FA and harmonic content are expected to help overcome the degeneracy between some parameters, particularly the mass and radius (Psaltis et al. 2014).

We selected phenomenological models that allow us to reproduce some basic key properties of the light curve and oscillations of J1814, assuming different surface temperature pattern properties. We developed eight models and following is an overview of each model, grouped by surface pattern and summarized in Table 1. To conform to the naming patterns of earlier X-PSI models (Riley et al. 2019), we will prefix each model name with the symbol ST-, which stands for Single Temperature. This is mainly because we assume a single hot spot, emitting at a uniform temperature, and do not yet consider the possibility of more than one emitting hot region. While our choice of models is necessarily restricted, they nevertheless allow us to simulate variability of hot spot properties and test the effects on parameter recovery.

ST-\(H_T\) & ST-\(\tilde{S}H_T\): Both models assume a hot spot on the surface of a star with a set angular radius and a variable temperature over time. Such a scenario would be consistent with confined burning, with fuel directed to a small region of the star by a strong magnetic field, or because the combined effects of a magnetic and Coriolis force would prevent the flame from spreading. With ST-\(H_T\), we presume that the star’s non-burning region is too cool to emit X-rays in the bands covered by the instrument’s effective area. Model ST-\(S\tilde{H}T\), however, allows the rest of the star to emit with a uniform and constant temperature.

ST-\(H_{T,R}\) & ST-\(S\tilde{H}_{T,R}\): Similar to the previous models, these models assume a temperature-varying hot spot, but this time, we also allow the burning region angular radius to evolve with time. These models would be consistent with a scenario in which the flames propagate outward from the ignition point but ultimately stall due to the interaction of the magnetic and/or Coriolis forces. Similar to the earlier models, model ST-\(H_{T,R}\) prevents the non-burning region of the star from emitting, whereas ST-\(S\tilde{H}_{T,R}\) allows this to occur.

ST-\(\tilde{S}T_H\) & ST-\(\tilde{S}T_{\tilde{H}}\): These two models allow the bulk of the star’s surface temperature to vary over time, with the exception of an isolated hot spot with a constant temperature. This would correspond to a scenario where the flame spreads and engulf the entire NS surface. As for the origin of the hot spot responsible for the pulsations, we speculate that it may have existed before the explosion (e.g. accretion spot), or that it may have been produced by some unknown mechanism on the star surface, in the atmosphere, or in the accretion funnel. In the ST-\(\tilde{S}T_{\tilde{H}}\) model, the angular radius of this hotter region can extend over time, whereas in model ST-\(\tilde{S}T_H\), this is not permitted.

Table 1. Table summarizing the models. \(T_{\text{spot}}\), \(\zeta_{\text{spot}}\), and \(T_{\text{star}}\) are respectively the hot spot temperature, hot spot angular radius and star temperature. \(\sim\), \(\sim\) and \(\times\) mean respectively that the corresponding parameter varies over time, is constant and is non-existent or not considered for this model.

<table>
<thead>
<tr>
<th>Models</th>
<th>(T_{\text{spot}})</th>
<th>(\zeta_{\text{spot}})</th>
<th>(T_{\text{star}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST-(H_T)</td>
<td>(\sim)</td>
<td>(\sim)</td>
<td>(\times)</td>
</tr>
<tr>
<td>ST-(H_{T,R})</td>
<td>(\sim)</td>
<td>(\sim)</td>
<td>(\times)</td>
</tr>
<tr>
<td>ST-(S\tilde{H}_T)</td>
<td>(\sim)</td>
<td>(\sim)</td>
<td>(\sim)</td>
</tr>
<tr>
<td>ST-(S\tilde{H}_{T,R})</td>
<td>(\sim)</td>
<td>(\sim)</td>
<td>(\sim)</td>
</tr>
<tr>
<td>ST-(\tilde{S}T_H)</td>
<td>(\sim)</td>
<td>(\sim)</td>
<td>(\sim)</td>
</tr>
<tr>
<td>ST-(\tilde{S}T_{\tilde{H}})</td>
<td>(\sim)</td>
<td>(\sim)</td>
<td>(\sim)</td>
</tr>
</tbody>
</table>

2.2 Atmospheric model

As mentioned in Section 1, NSs have a thin atmosphere layer at their surface, which needs to be understood to infer the stars’ properties. As photons travel through this surface layer, they are absorbed, re-emitted, and up-scattered to different energies and directions. This causes the emergent spectra of thermonuclear bursts to
(TBs) to differ from isotropic blackbody emission (see e.g. London, Howard & Taam 1984; Lapidus, Syunyaev & Titarchuk 1986; London, Taam & Howard 1986; Ebisuzaki 1987; Madej 1991; Madej, Joss & Różańska 2004; Suleimanov, Poutanen & Werner 2011a, 2012; Worpel, Galloway & Price 2013; Suleimanov et al. 2017; Suleimanov, Poutanen & Werner 2020). In this work, we make use of the atmosphere model developed for TBs in Suleimanov et al. (2012). This model solves the radiation transfer, hydrostatic equilibrium, and energy balance equations using the exact relativistic Klein-Nishina cross-section and redistribution function for Compton scattering. The spectra predicted by this model can be approximated closely by a diluted blackbody: $F_E \approx \frac{w}{\pi} B_\mu(f_e \times T_\text{eff})$ with dilution $w$ and colour-correction $f_e$ factors depending entirely on the surface gravity, luminosity, and chemical composition. The angular distribution of the emergent radiation is found to be well approximated by the Chandrasekhar–Sobolev beaming function for an electron scattering dominated atmosphere (Chandrasekhar & Breen 1947; Sobolev 1949; Chandrasekhar 1960; Sobolev 1963): $I(\mu) \sim 1 + 2.06\mu$, where $\mu$ is the cosine of the emission angle (Suleimanov et al. 2012, 2020). Thus, for simplicity, we fitted the atmosphere model spectra with a diluted blackbody using the same procedure as in Suleimanov et al. (2012), and calculated the total emitted intensity as $I_c(\mu) = B_\mu(T_\text{eff})(0.421 + 0.868\mu)$, similarly to equation (9) in Suleimanov et al. (2020).

For a computationally efficient pulse profile calculation, a grid of atmosphere models was pre-computed into a lookup table. The format of this table is similar to those used in NICER-analysis and explained in Bogdanov et al. (2021), except for the fact that our surface gravities $\log(g/\text{cm s}^{-2})$ range from 13.7 to 14.9 with $\Delta \log(g/\text{cm s}^{-2}) = 0.1$ spacing and effective temperatures range from $6.7$ to $7.6$ $[\log(T_{\text{eff}}/1\text{K})]$ with $\Delta \log(T_{\text{eff}}/1\text{K}) = 0.02$ spacing. We also created a grid in photon energies and emission angles similar to that in Bogdanov et al. (2021) [i.e. from $-1.3$ to $2.0$ in $\log(E/\text{keV})$ and from $0$ to $1$ in $\mu$, in order to easily apply the same interpolation methods as in the previous NICER-analysis when estimating the intensity for a given temperature, surface gravity, photon energy, and emission angle. Throughout this analysis, we have applied atmosphere models with solar abundance composition ($X = 0.737, Z = 0.0134$) for synthetic data generation and statistical inference.

2.3 Interstellar absorption

As X-ray photons pass through the interstellar medium (ISM), they are attenuated and/or absorbed. This results in the initial spectrum being altered. The observed spectrum can be expressed as: $I_{\text{observe}} = I_{\text{source}} e^{-\sigma_{\text{ISM}}(E)N_H}$, with $\sigma_{\text{ISM}}(E)$ the energy dependent ISM photoionization cross section normalized to the total hydrogen column density and $N_H$ the hydrogen column density measured in atoms cm$^{-2}$. We pre-computed lookup tables of attenuation factors; $f_\lambda(E, N_H) = e^{-\sigma_{\text{ISM}}(E)N_H}$ as a function of the energy (ranging from 0.1 to 30 keV) using the X-ray absorption model TBAbs (Wilms, Allen & McCray 2000) for different hydrogen column densities.

5However, we note that the original atmosphere grid was defined in luminosities relative to the Eddington luminosity $L = L_L /L_\text{bol}$ with the following values: 0.001, 0.003, 0.01, 0.03, 0.05, 0.07, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.98, 1.0, 1.02, 1.04, 1.06, 1.08, 1.1, 1.12; and in surface gravities $\log(g/\text{cm s}^{-2})$ ranging from 13.7 to 14.9 with $\Delta \log(g/\text{cm s}^{-2}) = 0.15$ spacing. A new grid, defined as explained in the main text, was produced by interpolating to the corresponding temperature and surface gravity values.

We find that, given a reference hydrogen column density $N_H^\text{ref}$, $f_\lambda(E, N_H) \approx f_\lambda(E, N_H^\text{ref})N_H/N_H^\text{ref}$ and verified that this approximate relation provides sufficient accuracy for our analysis. Thus, for improving the computing efficiency particularly during sampling, we extract the attenuation factors using the scaling relation and use the attenuation factor corresponding to $4 \times 10^{20}$ atoms cm$^{-2}$ as a reference hydrogen column density.

2.4 Background

To effectively extract parameters using pulse profile modelling, a thorough understanding of the background distributions in the data is crucial (see Salmi et al. 2022). For burst oscillation sources, there are two key components that are expected to produce the background photons (Watts 2019). The first, unpulsed, component originates from different X-ray sources in the field of view, the instrument electronics (see e.g. Campana 2022, for a review), and most significantly the observed source’s close environment (e.g. accretion disk, accretion column, etc.). The second component, this time pulsed, may result from X-ray emission caused by accretion plasma striking the stellar surface (giving rise to accretion-powered pulsations). However, only 8 of the 19 known TBO sources have exhibited accretion-powered pulsations (persistent accretion-powered pulsations for six sources and intermittent accretion-powered pulsations for the remaining two). With the typical accretion hot spot temperature being $\sim 1$ keV (Poutanen & Gierliński 2003; Salmi et al. 2018) and the temperature at the peak of the burst being $\sim 3$ keV (Galloway et al. 2020), it is expected that the contribution of the accretion-powered pulsations will be low, especially at the peak of the burst (see also Watts, Strohmayer & Markwardt 2005). Extra contributions to the pulsed background may originate from shocks in the accretion funnel as well as the pulsed X-ray signal from the stellar surface reflecting from the accretion disk. Given our current understanding of the origin of the burst oscillation, the pulsed component may or may not be categorized as background, depending on the phenomenological model considered.

For the sake of simplicity, while generating the synthetic data, we introduced only an unpulsed background with a blackbody and power-law component. We made use of the blackbody and power-law parameters values obtained by Krauss et al. (2005) through spectral analysis of J1814. Hence, the blackbody temperature and radius were fixed to 0.95 keV and 1.6 km respectively, while the power-law index was set to 1.41 with an amplitude of $3.32 \times 10^{-2}$ photons keV$^{-1}$ cm$^{-2}$ s$^{-1}$. We assumed that these background photons were produced close to the pulsar and therefore accounted for absorption of the background photons while travelling to Earth. The current likelihood computation in X-PSI is performed by marginalization over the background. Thus, the background becomes a free implicit parameter and gets degenerate with any unpulsed component of the signal. Similar to Salmi et al. (2022), we used a flat prior for each background variable centered around the true value with hard cuts at $\pm 3\sigma$ to prevent the background from increasing the already large number of uncertainties (see Riley et al. 2021; Salmi et al. 2022, for details). Another reason for imposing such a constraint is that prior knowledge of the background level for real burst oscillation.
Table 2. Parameters and their respective prior density used for inference.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Prior density</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M (M_\odot) )</td>
<td>Gravitational mass</td>
<td>( M \sim \mathcal{U}(1.0, 3.0) )</td>
</tr>
<tr>
<td>( R_{eq} ) (km)</td>
<td>Equatorial radius</td>
<td>( R_{eq} \sim \mathcal{U}(3r_g(1.0), 16.0)^a )</td>
</tr>
<tr>
<td>( D ) (kpc)</td>
<td>Distance to Earth</td>
<td>( D \sim \mathcal{N}(d, 0.15d)^b )</td>
</tr>
<tr>
<td>( \cos(i) )</td>
<td>Cosine of Earth inclination to rotation axis</td>
<td>( \cos(i) \sim \mathcal{U}(0.0, 1.0) )</td>
</tr>
<tr>
<td>( \phi_{spot} ) (cycles)</td>
<td>Phase of the hot region</td>
<td>( \phi_{spot} \sim \mathcal{U}(-0.25, 0.75) )</td>
</tr>
<tr>
<td>( \Theta_{spot} ) (radian)</td>
<td>Co-latitude of the centre of the hot spot</td>
<td>( \Theta_{spot} \sim \mathcal{U}(0.0, \pi/2) )</td>
</tr>
<tr>
<td>( \zeta_{spot} ) (radian)</td>
<td>Angular radius of the hot spot</td>
<td>( \zeta_{spot} \sim \mathcal{U}(0.0, \pi/2) )</td>
</tr>
<tr>
<td>( \log[T_{spot}(K)/1K])</td>
<td>Hot spot effective temperature</td>
<td>( \log[T_{spot}(K)/1K] \sim \mathcal{U}(6.7, 7.6)^c )</td>
</tr>
<tr>
<td>( \log[T_{star}(K)/1K])</td>
<td>Star effective temperature (K)</td>
<td>( \log[T_{star}(K)/1K] \sim \mathcal{U}(6.7, 7.6) )</td>
</tr>
<tr>
<td>( N_H ) (10(^{20})cm(^{-2}))</td>
<td>Interstellar attenuation column density</td>
<td>( N_H \sim \mathcal{U}(0.0, 10.0) )</td>
</tr>
</tbody>
</table>

Notes: 

\(^a\) \( r_g \): Schwarzschild gravitational radius in km. 

\(^b\) \( d \): distance used to generate the synthetic data. 

\(^c\) Temperature bounds set by bursting NS model described in Section 2.2.

Sources can be obtained by assuming that it is similar to the pre-burst emission.

2.5 Priors

Table 2 provides a summary of the priors used for parameter inference. When choosing the parameter set for each model to generate the synthetic data, we stayed within the same prior bounds (see paragraph below for the specific case of the distance). A uniform prior is assumed by default for the majority of parameters. Following is a discussion of non-trivial prior choices.

We fix the joint mass-radius prior density distribution to be flat in \((M, R_{eq})\) space (see Riley, Raaijmakers & Watts 2018; Riley et al. 2019, for discussion). This is done primarily to pave the way for more frequent application of posteriors in inferring the EoS. In order to account for the range of possible TBO source masses, the mass prior support is set to [1.0, 3.0]\(M_\odot\) and the radius support to [3\(r_g(1.0), 16.0\)] km with \(r_g(M) = GM/c^2\) [\(M\) in \(M_\odot\) and \(r_g(1.0)\) in km]. We set the compactness condition to \(R_{pole}/r_g(M) > 2.9\) and account for higher-order images expected for \(R/r_g(M) < 3.52\) (see Bogdanov et al. 2021, and references therein) during both data generation and the inference. We imposed no restriction on the image order; thus, images are summed over until higher-images are no longer visible or visibility is truncated owing to a lack of numerical precision. During inference, the surface gravity requirement from the atmosphere model also affects the \((M, R_{eq})\) prior space. In fact, samples with \(\log g\)\([13.7, 14.9]\) are automatically rejected.\(^7\)

When generating synthetic data that mimics the key properties of the J1814 burst oscillations, we consider distances between 0.5 and 20 kpc; but during inference, we assume a Gaussian prior with a mean equal to the distance used to create that data set, a standard deviation equal to 15 per cent of that distance, and a hard cutoff of \(\pm 5\sigma\). This is done for two reasons. First, to focus on the effects of time-varying parameters and reduce the consequences of multiple-parameter degeneracy. Second, since distance constraints for TBO sources can be determined independently through methods such as photospheric radius-expansion analysis (see e.g. Galloway et al. 2020), Gaia (Gaia Collaboration et al. 2016) astrometry of the donor stars (Ding, Deller & Miller-Jones 2021; Moran et al. 2022), through pulse profile modelling of the accretion pulsations of AMPSs or by modelling the burst spectra mentioned in Section 1.\(^7\)

To keep our simulations manageable within the restrictions of time and computational resources, we restrict the centre of the hot spot colatitude and the inclination to the stellar northern hemisphere.

2.6 Synthetic data

Bursts, in general, have changing light curves induced by temperature fluctuations as the burning progresses. Our goal in this part is to create synthetic data that mimics this type of light curve, with burst oscillations, utilizing as much of our current knowledge as possible. A burst temperature profile, which defines how the temperature evolves over time, is required to achieve this goal when utilizing X-PSI.

Given the uncertainty associated with burning physics, we were unable to obtain from the literature a simple analytical expression for such a temperature. Bhattacharyya & Strohmayer (2006) proposed a functional (exponential) form for the temperature, but using their formula requires a few additional free parameters. Galloway et al. (2020) have since shown that a power-law model, or a Gaussian plus a power-law model, yield a more accurate fit in the tail than an exponential decay. This implies additional customizable settings. In order to reduce the number of free parameters associated with the temperature profile, we instead used a previously established temperature profile acquired through simulation.

Comparing the light curves of GS 1826–24 and J1814 (see Galloway et al. 2020, for light curves\(^b\)) suggested that GS 1826–24 is a reasonable mimic for J1814 (see Fig. 1). The burst light curves of both sources have a short rise time and a slow exponential decay. This is consistent with a helium flash followed by a rp-process burning of hydrogen (Schatz et al. 2001) and implies similar burning physics. Therefore, to obtain the temperature profile for the analysis, we used the simulated temperature profiles of GS 1826–24 obtained using the stellar and binary evolution code Modules for Experiments in Stellar Astrophysics (Paxton et al. 2011), then averaged all simulated bursts’ temperatures in the ma6 model sequence (Meisel 2018) apart from the first and the last. We exclude these two since the initial burst is typically more energetic than the average burst and the last burst was incomplete. We note that model ma6 has the same chemical composition as that of the atmosphere model used throughout the modelling. We denote by \(T_{\text{fit}}(t)\) the final temperature profile obtained, where \(t\) is the time.

\(^7\)See footnote in Section 2.2

\(^b\)https://burst.sci.monash.edu/
Figure 1. Synthetic burst light curve produced using parameter set 2 of model \( \text{ST-}\hat{S}\hat{T}_{R} \), compared to an actual burst of J1814 (burst 26) and the average light curve of GS 1826–24 normalized by the synthetic burst maximum counts. The black dashed horizontal line is the background threshold above which the chopped burst is considered.

Figure 2. Temperature (top) and angular radius (bottom) profile. \( T(t) \) is the hot spot temperature profile \( (T_{\text{hot}}(t)) \) for models: \( \text{ST-}\hat{S}\hat{T} \) and \( \text{ST-}\hat{S}\hat{T}_{R} \). \( T_{\text{hot}}(t) \) represents the hot spot temperature profile \( (T_{\text{hot}}(t)) \) for models: \( \text{ST-}\hat{T}, \text{ST-}\hat{T}_{R}, \text{ST-S}\hat{T}, \text{ST-}\hat{S}\hat{T}_{R} \) and \( \text{ST-S}\hat{T}_{R} \) and the star temperature profile \( (T_{\text{star}}(t)) \) for models: \( \text{ST-}\hat{T}, \text{ST-}\hat{T}_{R}, \text{ST-S}\hat{T}, \text{ST-}\hat{S}\hat{T} \) and \( \text{ST-S}\hat{T}_{R} \).

Since some models require two temperature profiles, one for the hot spot and one for the star, we manually created a second profile and named it \( T(t) \). We also crafted an angular radius profile denoted \( \zeta_{\text{hot}}(t) \) for every model that required it. Both \( T(t) \) and \( \zeta_{\text{hot}}(t) \) are shown in Fig. 2 and were adjusted for each model such that the resulting light curve is comparable to J1814, the number of photons is approximately \( 10^{5} \), and the FA is approximately 10 per cent.

For each model, we set the temperatures as follows:

(i) \( T_{\text{hot}}(t) = T_{\text{hot}}(t) \) for \( \text{ST-}\hat{T}, \text{ST-}\hat{T}_{R}, \text{ST-S}\hat{T} \) and \( \text{ST-S}\hat{T}_{R} \);

(ii) \( T_{\text{hot}}(t) = T(t) \) and \( T_{\text{star}}(t) = T_{\text{hot}}(t) \) for \( \text{ST-}\hat{T} \)

(iii) \( T_{\text{star}}(t) = T_{\text{hot}}(t) \) for \( \text{ST-}\hat{T}_{R} \).

We discretized all temperatures and \( \zeta_{\text{hot}}(t) \) into arbitrary 0.25 s time intervals. We then generated 0.25 s exposure time pulses for a total of 150 s and summed those pulses to obtain the desired pulse and light curve (see Fig. 1). Each 0.25-s pulse is the sum of the expected counts from the model and a Poisson fluctuation. We used the random number generator of the GNU Scientific Library (GSL; Gough 2009) to add this Poisson fluctuation. To ensure that a different noise is generated for each time segment, the seed environment variables of GSL were changed for each segment. Snapshots of the pulses corresponding to the light curve of Fig. 1 are shown in Fig. A1 of Appendix A.

For this paper, we do not allow the background level to vary during the burst. Hence, we set the expected bolometric background counts to 20 in each 0.25 s time bin. This yields background counts equivalent to 10 per cent of the total photon counts as observed in J1814 bursts. To avoid having the background dominate the data, we only consider the portion of the synthetic burst from the point where the counts during the rise for the first time exceed five times the injected background counts, to the point in the decay where the counts are for the last time at least five times the injected background counts. We refer to this part of the synthetic burst as the chopped burst and this is used for parameter inference.

A typical burst of J1814 has \( 10^{5} \) counts and 10 per cent rms FA. Therefore, a set of parameters is selected if and only if the chopped burst shape is comparable to that of the averaged of GS 1826–24 light curves, produces \( 10^{5} \) counts and 10 per cent rms FA, and if the rms FA is relatively constant during the chopped burst. To guarantee that the final requirement is met, we segment the chopped burst in bins of 5000 photons and compute the rms FA for each bin. For each model, we retained three parameter sets and summarize them in Table A1 of Appendix A.

Since a typical burst of J1814 has \( 10^{5} \) counts, it is possible, in principle, to get tighter parameter constraints by combining more than one burst from the same source. To explore if the parameter recovery is affected when multiple bursts are combined, we produced 100 synthetic bursts for each selected parameter set and merged the first 10 and 100 synthetic bursts to form a mega burst of, respectively, \( 10^{6} \) and \( 10^{7} \) photons. For each burst in the 100 synthetic bursts sequence, only the MLS seeds were allowed to vary during and between bursts. The model parameters, such as the hot spot location, angular radius, temperature profile, etc., were considered fixed from one burst to another.9

We refer to all data created using a time dependent temperature (and angular radius, depending on the model) as the variability data set.

9 This is an ideal scenario. For most sources, hot spot location, angular radius, temperature profile, etc., are expected to vary from burst to burst (see e.g. Munoz, Ozel & Chakrabarty 2002; Galloway et al. 2020). This is something that could and should be investigated in more detail in future work. However, given the similarities in some bursts of 1814, merging those bursts assuming that the temperatures and/or radii profiles are identical is a reasonable preliminary approach.
Parameter degeneracy in pulse profile modelling is well-known and can be difficult to overcome (Lo et al. 2013). Prior to submitting the variability data set to the simulation pipeline, we must ensure that the method by which each burst was generated did not introduce biases that could lead to inaccurate parameter recovery. Moreover, we use nested sampling package MultiNest to estimate the posterior probability density function (PDF) for each parameter. MultiNest samples the prior hyper-space and approximates the true PDF and depending on the nature of the problem, its settings must be adjusted for optimal performance. To do so, we created a calibration data set where for each model, the time-dependent variable has been set to a fixed value such that the photon counts and the FA remain as close as possible to their the variability data set counterpart (see Table A2 in Appendix A). The exposure time remains as it would be for the chopped burst. For the $10^6$ and $10^7$ photon count bursts, only the exposure times were extended to yield the desired counts.

In summary, we have generated two data sets. The first data set, referred to as the calibration data set, contains no temporal variation. The hot spot temperature, angular radius, and stellar temperature (if applicable to that model) were set to remain constant. For the variability data set, however, the temperatures and/or angular radius were time-dependent variables conditional on the model. We utilized both data sets and conducted inference runs based on a model that was used to generate the calibration data set, in which variability is eliminated. Our findings are presented below.

3 RESULTS

We attempted to recover the parameters by conducting an inference run on each parameter set for each model. Given three parameter sets per model, eight models per data set, two data sets, and three cases with a different number of count we performed a total of 144 inference runs. We began by examining the quality of data recovery per data set. Then we examined each model’s performance with respect to the variability data set. To estimate the quality of parameter recovery, we make use of the PP-plot, which is a tool used to compare two distributions and assessing their similarities or differences. This method is commonly used in gravitational-wave studies to evaluate the robustness of inference processes (Sidery et al. 2014; Berry et al. 2015; Gair & Moore 2015; Pankow et al. 2015; Veitch et al. 2015; Biwer et al. 2019; Romero-Shaw et al. 2020; Cornish et al. 2021; Mozzon et al. 2022; You et al. 2022). Assuming the inference process has no inherent bias, the injected (or true) value of each parameter must fall inside the $x$ per cent credible interval $x$ per cent of the time (Cook, Gelman & Rubin 2006). This implies that the fraction of the recovered parameter within a given credible interval must be uniformly distributed. Hence, the cumulative fraction of the recovered parameter should sit on the diagonal of the PP-plot, with occasional fluctuations due to the limited sample size. We therefore perform a KS-test to determine the extent to which the cumulative fraction of recovered parameters deviates from the uniform distribution.

We introduce a wrongness parameter, in addition to the PP-plots, to cross-check the quality of parameter recovery. We define wrongness as $(X_{\text{inf}} - X_{\text{med}})/\Delta X_{\text{inf}}$, where $X_{\text{inf}}$ and $X_{\text{med}}$ are, respectively, the inferred and the injected values (median of the posterior) of the parameter $X$.

$\Delta X_{\text{inf}}$ is the 68 per cent credible interval recovered from sampling. Alternatively stated, wrongness is the difference between the inferred and injected values in units of 68 per cent credible interval (defined around the median as in the posterior plots), and would correspond to the standard deviation if the posteriors were Gaussians. A perfect parameter recovery would result in a wrongness of zero. We expect, however, that, due to statistical fluctuations and systematic error during sampling, the wrongness of all runs for a particular parameter will cluster close to zero.

In the following, we focus mainly on the space-time parameters: compactness, radius, and mass; the primary parameters of interest for dense matter studies. Alongside the space-time parameters, we also inferred the geometrical parameters. Some are recovered more effectively than their EoS equivalents, while others are not. However, we provide fewer details of the geometric parameter inference since they are not the primary focus of this investigation.

3.1 Calibration data set

We gauge the validity of our MultiNest settings and check for any bias by performing parameter recovery using the calibration data set. The PP-plot in Fig. 3 shows the cumulative fraction of parameter values recovered within a credible interval as a function of the credible intervals for the calibration data set (left). The top, middle, and bottom panels are the recovery fraction for the compactness, radius, and mass, respectively. Each colour represents a given number of counts, as indicated on top of the figure. The numbers in parenthesis are the $p$-values from the KS-test for each number of counts in the corresponding color. The grey shaded regions are the ±1, 2, and 3σ expected uncertainty regions corresponding to a sample size of 24 (all the runs corresponding to one number of counts) and are derived from the binomial quantile (Cameron 2011).

For this data set, the recovered compactness and radius are extremely close to what is expected for all the counts and yield $p$-values > 0.28. The mass is also recovered effectively for $10^5$ and $10^6$ counts, but is somewhat worse for $10^7$ counts. The curve always lies above the expected recovery, which implies that the majority of injected masses are below the 0.5 probability mass of the recovered posterior. A better recovery might be obtained by raising MultiNest’s live points and/or lowering the sampling efficiency. However, the cumulative fraction of parameter values recovered within a credible interval remains inside the ±3σ expected region. In Fig. 4, we plot both mass (top panel) and radius (bottom) wrongness for all 24 runs. For the radius, the mean wrongness is $-0.12$, $-0.07$, and 0.04 for, respectively, $10^5$, $10^6$, and $10^7$ counts. The wrongness is centred near zero and decreases as the number of photons increases, which implies radius is well recovered. For the mass, the mean wrongness is 0.09, 0.19, and 0.27 for, respectively, $10^5$, $10^6$, and $10^7$ counts. The average wrongness tends to increase as the data quality improves which indicates a small bias towards high inferred masses. Nevertheless, the maximum wrongness for the mass is about 1.6, which indicates that the recovered mass is not far from the injected value.

Another method of examining parameter recovery is by looking at posterior plots. In Fig. 5, we show a graphical representation of how the posteriors behave as the quality of the data improves (as the number of counts increases). This posterior plot corresponds to $X_{\text{med}}$.
to parameter set 1 of the ST-$S_T, H_{\text{T},R}$ model calibration data set counterpart. As the number of counts increases, the credible intervals of the compactness, the radius, and the mass tighten up and become narrower around the injected (true) value. This is the expected behaviour of parameter recovery when there is no bias in the inference process.

The PP-plots, the posterior plots, and the wrongness plots suggest the absence of significant biases in our inferences of compactness, mass, and radius for the calibration data set. Therefore, there is no evident bias introduced by the data creation method. Thus, we apply the same MULTINEST settings to the variability data set.

Using the calibration data set, we examined the uncertainties in the radius and mass. The uncertainty is defined as $\Delta X/X$ where $\Delta X$ is the 68 per cent credible interval of the inferred parameter $X$. This is shown on Fig. 6 for each parameter set of each model. In general, the uncertainties decrease as more counts are collected, as expected. The radius has larger fractional uncertainties than the mass. This provides more justification for why the radius recovery seems better than that of the mass from the PP-plots.

### 3.2 Variability data set

In the right column of Fig. 3, we present the cumulative fraction of parameter values recovered inside a credible interval as a function of the credible intervals for the variability data set for the compactness (top), radius (middle), and mass (bottom). When time variability is overlooked during inference, parameter recovery for $10^5$ counts is effective. At this low data quality regime, the PP-plots remain within the expected $3\sigma$ credible interval and produce statistically acceptable $p$-values for all EoS parameters. As anticipated, the parameter recovery is lower than the calibration data set counterpart for mass and radius. For compactness, however, the recovery is better, primarily due to statistical fluctuation and our relatively small sample size.
Figure 4. Wrongness, using the calibration data set, for the mass and radius. The Wrongness \( = (X^\text{calc} - X^\text{inj})/\Delta X^\text{inj} \), is a measure of how far off the inferred value \( X^\text{calc} \) is from the injected value \( X^\text{inj} \) in units of 68 per cent credible interval (around the median); \( X \in \{M, R_\text{eq}\} \). On the x-axis are the different models and their parameter sets with three different parameter sets S1, S2, and S3 for each model.

The PP-plots diverge, however, from the expected behaviour as the data quality improves (10⁶ and 10⁷ counts). Both the 10⁶ and 10⁷ counts graphs exceed the ±3σ expected region, with the 10⁷ count case performance being much worse. More injected values than expected are found at lower credible intervals, for both 10⁶ and 10⁷ counts, but fewer injected values than expected are found at higher credible intervals. This implies an excessively narrow posterior and substantial biases. To emphasize this point, we show on the right panel of Fig. 7 the posterior of the compactness, mass, and radius. This figure shows that for 10⁵ counts, even when temporal variability is overlooked, the posteriors are as broad as their calibration counterpart and encompass nearly the whole parameter space. In a high data quality regime, especially for 10⁷ counts, the posterior is compressed to a small region of the parameter space. This is because most of the parameter vectors in the prior space fail to yield good likelihood values, resulting in the most of samples clustering in smaller regions with relatively higher likelihood. This is the consequence of using an inappropriate model to infer the parameters. To provide more evidence of such poor model performance, Fig. 8 shows the residuals corresponding to the posterior of Fig. 7. The residuals were computed using the maximum likelihood parameter vector in each case. The residuals for 10⁵ counts appear normal. Due to the rather poor quality of the data, the majority of the parameters fit the data relatively easily. However, at higher counts, particularly at 10⁷ counts, discernible patterns emerge. The model generates an excess of counts in certain regions highlighted in blue and a deficiency in others (in red), with clusters that are large enough to indicate a significant deviation from randomness. These clusters are indicative that the fitting under this data regime is inadequate. This is not unexpected, given that the model used to create the data differs from the model used for inference.

Figure 5. Example of posterior plots obtained using data corresponding to set 1, model ST-S8 \( H_{T,R} \) selected from the calibration data set. The dashed black lines are the values injected to generate the synthetic data. The green vertical band shows a 68 per cent posterior density credible interval symmetric in posterior mass around the median corresponding to 10⁷ photon counts. The two-dimensional (2D) posterior distributions show the 68 per cent, 95 per cent, and 99 per cent posterior credible region, respectively, from most opaque to least.

Figure 6. Mass and radius uncertainty at 68 per cent credible interval for the calibration data set. On the x-axis are the different models and their parameter sets with SX the parameter set X of the model. The models indicated on the x-axis correspond to their variability counterparts, for which synthetic data were generated using the method outlined in Section 2.6.

The PP-plots show, however, from the expected behaviour as the data quality improves (10⁶ and 10⁷ counts). Both the 10⁶ and 10⁷ counts graphs exceed the ±3σ expected region, with the 10⁷ count case performance being much worse. More injected values than expected are found at lower credible intervals, for both 10⁶ and 10⁷ counts, but fewer injected values than expected are found at higher credible intervals. This implies an excessively narrow posterior and substantial biases. To emphasize this point, we show on the right panel of Fig. 7 the posterior of the compactness, mass, and radius. This figure shows that for 10⁵ counts, even when temporal variability is overlooked, the posteriors are as broad as their calibration counterpart and encompass nearly the whole parameter space. In a high data quality regime, especially for 10⁷ counts, the posterior is compressed to a small region of the parameter space. This is because most of the parameter vectors in the prior space fail to yield good likelihood values, resulting in the most of samples clustering in smaller regions with relatively higher likelihood. This is the consequence of using an inappropriate model to infer the parameters. To provide more evidence of such poor model performance, Fig. 8 shows the residuals corresponding to the posterior of Fig. 7. The residuals were computed using the maximum likelihood parameter vector in each case. The residuals for 10⁵ counts appear normal. Due to the rather poor quality of the data, the majority of the parameters fit the data relatively easily. However, at higher counts, particularly at 10⁷ counts, discernible patterns emerge. The model generates an excess of counts in certain regions highlighted in blue and a deficiency in others (in red), with clusters that are large enough to indicate a significant deviation from randomness. These clusters are indicative that the fitting under this data regime is inadequate. This is not unexpected, given that the model used to create the data differs from the model used for inference.
As we did for the calibration data set, we also plot the wrongness for both the mass and radius in Fig. 9. Unlike in the calibration data set case, the wrongness moves away from zero when the data improves and yields values up to 100 for some models. This provides evidence that the posteriors have been overly constrained away from the injected values. Only a few models yield acceptable wrongness values at 10^7 counts. To check whether data generated under any specific model performs better than others, we show the PP-plot in Fig. 10 for the compactness (see Fig. A2 in Appendix A for the radius and mass). The top and bottom panels correspond, respectively, to 10^6 and 10^7 counts. On top of the figure, each model and its corresponding p-values (left in parenthesis: 10^6 counts, right in parenthesis: 10^7 counts) are highlighted. Given the small sample size per model (three inference runs), it is difficult to draw meaningful conclusions regarding the quality of the recovery based solely on this PP-plots. However, we display the performance of each model as a visual indication. Despite the substantial uncertainties, most of the models fall outside the expected 3σ region and none of the models seem to perform well.

4 DISCUSSION

The major objective of this study was to examine parameter recovery for pulse profile modelling of TBOs: both ground truth recovery, and the uncertainties when the temporal variations that we expect from TBO sources are ignored. To achieve this objective, we used cutting-edge pulse profile models (incorporating the relevant general and specific relativistic effects, bursting atmosphere physics, and interstellar absorption), coupled with a reasonable background model, and constructed realistic test cases based on one of the most promising sources for this type of analysis, J1814.

We started by generating synthetic data that mimicked J1814 burst light curves and oscillation properties. In doing so, we developed eight phenomenological models that cover our current (highly uncertain) understanding of burning physics and in particular the physical mechanism that gives rise to the burst oscillations. Then, we created two data sets: a calibration data set and a variability data set. The calibration data set contains no time variability and was primarily used to determine the optimal settings for our sampler and to assess the uncertainties that can be achieved on mass and radius estimates. The variability data set, which contains bursts with time variability, was used to test the recovery of parameters when time variability (of the emitting region temperatures and sizes, depending on the model) is ignored during inference – an approach that could reduce computational cost of inference runs. To evaluate the quality of parameter recovery, we used PP-plots, wrongness parameter (the difference between recovered and injected parameters in units of 68 per cent credible interval), posterior plots, and residuals. These tools allowed us to assess how well our sampler is able to recover the true values of the parameters in the presence of realistic levels of time variability.
4.1 Phenomenological burst oscillation models

We found that all of the phenomenological models considered are adequate to account for the main observed properties (an rms FA of the fundamental that is relatively constant throughout the burst and with an overall rms FA of 10 per cent) of the J1814 oscillations.¹³ Thus, testing each of these models would be desirable when modelling actual J1814 data, depending on whether bursts are analysed individually or merged before analysis.

Establishing a definitive link between these simple phenomenological models and the physical mechanisms responsible for oscillation is something that has not yet been done, and certainly our phenomenological models do not fully encapsulate the intricacies of theoretical oscillation mechanisms. Nevertheless, as detailed in Section 2.1, the simple models we proposed were motivated by physical models, and testing such phenomenological hot spot models via pulse profile modelling may eventually provide new insight into the burst oscillation mechanism.

4.2 Mass and radius constraints

Using our calibration data set, we found that uncertainties on the mass and radius measurements can be reduced to 10 per cent and 15 per cent (width of the 68 per cent credible interval), respectively, assuming $10^7$ counts from the star. This result is consistent with the findings of Lo et al. (2013), who predicted mass and radius constraints of 10 per cent when considering only $10^6$ counts from a hot spot located within $10^7$ of the equator. In contrast, a hot spot located within $10^7$ of the pole provided no useful constraint in their analysis. Our analysis falls somewhere in between these two cases, as the location of the hot spot in our analysis was allowed to be anywhere in the northern hemisphere, as long as it resulted in a lightcurve and oscillations that mimicked J1814’s properties. The analysis performed on the calibration data set also shows that mass has smaller fractional uncertainties than radius. This implies that the mass is typically more tightly constrained than the radius. When considering the variability data set, we determined that the parameter recovery for $10^5$ counts was nominally acceptable. This is owing to the fact that the sampling uncertainties on the mass and radius estimations are larger than the model uncertainties at this level of data quality. The average uncertainty of the mass and radius estimations was determined to be 30 per cent, which is much higher than the approximately 10 per cent uncertainty required to differentiate between different EoS models (Özel & Psaltis 2009). This implies two things. First, when performing inference on a burst

---

¹³Note that we did not attempt to match the harmonic content of the J1814 oscillations in our synthetic data generation, since our primary focus was on assessing the impact of variability. The harmonic content of our synthetic data varied substantially.
of about $10^5$ counts, temporal variability can be neglected and yet the inferred mass and radius will not show any measurable bias. Second, the mass and radius uncertainties from a $10^5$ counts run are insufficient to discriminate between EoS. Collecting more counts is therefore required. However, when more counts are collected by merging several bursts (yielding counts $\gtrsim 10^5$) and inference is performed neglecting the time variability, the mass and radius are incorrectly estimated. This occurs because, given the quality of the data, the modelling uncertainties outweigh the sampling uncertainties. In this context, ignoring time variability, given a limited computation budget, is ineffective. To correctly estimate the main properties of a star, more advanced methods are required.

### 4.3 Addressing inference bias associated with time variability

One of the main reasons for neglecting time variability is to reduce computational costs, which are known to be large, based on analysis of NICER RMP data using X-PSI (up to several hundred thousand Central Processing Unit (CPU) hours for the most complex models, see for example Riley et al. 2019). One strategy to mitigate the bias introduced by time variable parameters is to have a good, physically-motivated, model of the temporal evolution of the photospheric temperature. Parametrizing such a temperature profile in our simulation pipeline would speed up the analysis significantly. However, getting such a temperature profile from nuclear theory is difficult in practice for a real source, given the complexity in modelling X-ray bursts. To accurately model X-ray bursts, it is necessary to have comprehensive information about the thermal and compositional structure of the neutron star envelope, together with more than 1000 reaction rates involving over 300 different nuclides (Meisel, Merz & Medvid 2019, and references therein). In the absence of more accurate models, slicing the light curve into shorter segments where the variability can be neglected, and then jointly fitting the various segments, may reduce the systematic uncertainties. However, this approach is likely to come at considerable computational cost. Preliminary test runs reveal that while a run using the burst data of set 1 of our simplest model ST-$H_0$ takes approximately 13 000 core-hours, slicing the same data into an arbitrary eight pieces, and doing a joint fit allowing only the hot-spot temperature to vary from slice to slice, results in the inference run taking approximately 150 000 core-hours. We plan to further explore this procedure in future work.

To get a sufficient number of counts to probe the EoS, it is necessary to combine many bursts, since for RXTE data the number of photons in a single burst is not sufficient. Although we did not investigate the known burst to burst variability in our analysis, we anticipate that it will add to the temporal variability bias. Consequently, future X-ray missions: the enhanced X-ray Timing and Polarimetry (e-XTP; Watts et al. 2019; Zhang et al. 2019), and the Spectroscopic Time-Resolving Observatory for Broadband Energy X-rays (STROBE-X; Ray et al. 2019) missions, which will have larger effective areas covering the TBO energy range, may help us to avoid this additional complication by collecting a sufficient number of counts in fewer bursts. However, accounting for the temporal variability when analysing single bursts will still be required, even for these future instruments. With current modelling techniques, this would require allocation of a substantial computational budget.

### 4.4 Some limitations of this study

Despite the insights provided by our analysis, it is important to note that the derived uncertainties are conditional on the properties of J1814, which we used as a basis for our synthetic data sets. However, different TBO sources have distinct features, making it difficult to generalize our results. In order to mitigate these uncertainties and increase the robustness of our findings, it may be necessary to conduct a more randomized analysis on a larger sample size.

We are also not really able to reach a definitive conclusion about each individual model’s performance using the variability data set. Some models that neglect time variability may well perform better than others. However with only three samples for each model (set by computational constraints), it is possible that the poor performance of some models may be due to random fluctuation. To draw more robust conclusions, it would be necessary to expand the number of parameter sets for each model. Nevertheless, our results indicate that ignoring variability is risky.

### 4.5 Previous J1814 analysis via its thermonuclear burst oscillation properties

The first analysis to constrain the properties of a neutron star via pulse profile modelling of TBOs was conducted by Bhattacharyya et al. (2005) using J1814 bursts. Bhattacharyya et al. (2005) did not perform Bayesian inference in the way that we now do when pulse profile modelling (analysis methods have evolved substantially in the last few years in support of NICER), but they did use a relativistic ray-tracing model to fit a pulse profile obtained by combining – or ‘stacking’ – data from multi bursts. Doing this, the derived constraints on J1814’s compactness, hot spot, and geometrical properties. Several assumptions and simplifications were made during the analysis, including the following: (1) the entire analysis relied on two particular EoS (imposing a prior constraint on the mass-radius relation); (2) it was conditional on a hot spot with a fixed temperature of 2 keV; and (3) the effect of the interstellar medium absorption was not included. Most importantly, given the results of our new study, possible variability during and between the bursts that were stacked together was neglected. However, our study has found that ignoring variability during the bursts makes inference of the compactness problematic, especially when the number of stacked bursts yields a total counts of approximately $10^5$ photons and above (as in the case of J1814).

### 4.6 Future prospects

In principle, pulse profile modelling of TBO sources should be able to deliver information not only on the mass and radius but also on the TBO spot properties, including their time evolution. As we have shown in this paper, extracting this information is currently challenging. Given the complexity in accurately modelling X-ray bursts and the uncertainty on the oscillations’ origin, we proceeded in this analysis by testing different models relying on assumptions spanning a wide range of possibilities. Any theoretically motivated constraint would therefore increase the efficiency of future investigations. In particular, we would like to encourage the development of parameterized models of potential burst oscillation mechanisms, preferably coupled to models for the overall surface temperature evolution, in a form that can be used for pulse profile modelling so that we can test them more thoroughly. Alternatively, in the absence of a good physical model, slicing the burst into short time segments during which we can ignore the variability – while computationally intensive – may eventually help to constrain the wide range of possible models. Further study will be needed to determine how well we can constrain models – what would it require, for example to determine whether spot growth has a preferred direction rather than being symmetric? We also need to give more consideration to
how to model other components that might contribute to the data, such as pulsed emission from ongoing accretion-powered pulsations, or reprocessing of burst photons by the accretion flow.

Untangling these effects will be complicated, but there are also good prospects for using complementary constraints to help overcome these various issues. Not all bursts from TBO sources have detectable oscillations, but phase-averaged spectral modelling of those bursts can still yield constraints on mass and radius (Stieleman et al. 2011b; Özel 2013; Steiner et al. 2013; Guillot & Rutledge 2014; Nättäli et al. 2016; Steiner et al. 2018), as well as determining the accretion disk phenomena like kilohertz QPOs (Abolmasov, Nättäli & Poutanen 2020) or broadened iron lines (Bhattacharrya 2011, 2017; Ludlam et al. 2022). Some TBO sources also have accretion-powered pulsations, for which pulse profile modelling can also be done (e.g. Salmi et al. 2018), yielding constraints on mass, radius, and geometry. Being able to perform precisely this kind of cross-check on high-quality data is one of the drivers for future facilities such e-XTP and STROBE-X. X-ray polarimetry of such sources also offers a potential way to constrain geometries (Viironen & Poutanen 2004; Watts et al. 2019; Loktev et al. 2020; Salmi et al. 2021). If we can put all of these pieces together, then the prospects for reducing systematics and uncertainties are highly promising.

ACKNOWLEDGEMENTS

YK, TS, ALW, SV, and DC acknowledge support from ERC Consolidator Grant No. 865768 AEONS (PI Watts). The major part of the work was carried out on the HELIOS cluster including dedicated nodes funded via the above-mentioned ERC CoG. Additionally, this work used the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINIF-3731. SF contributed to this work as part of his Bachelor Project at the University of Amsterdam. The authors thank Isobel Romero-Shaw and Gregory Ashton for discussion on PP-plots. The authors also express our gratitude to Anna Bilous, Phil Uttley, Tess Jaffe, and Bryan K. Irby for providing their valuable assistance with regards to the response file. ZM was supported by the U.S. Department of Energy Office of Science under Grants No. DE-FG02-88ER40387 and DE-SC0019042, the U.S. National Nuclear Security Administration through Grant No. DE-NA0003909, and the U.S. National Science Foundation under Grant No. OISE-1927130 (International Research Network for Nuclear Astrophysics). VS acknowledges support by Deutsche Forschungsgemeinschaft (DFG; grant WE 1312/59-1).

DATA AVAILABILITY

For this analysis, we used the publicly available simulation code X-PSI (https://github.com/xpsi-group/xpsi.git). The synthetic data and inference runs are both accessible on Zenodo (https://doi.org/10.5281/zenodo.7665653).

REFERENCES

Altamirano D. et al., 2010, Astron. Telegram, 2932, 1
Baym G., Hatsuda T., Kojo T., Powell P. D., Song Y., Takatsuka T., 2018, Rept. Prog. Phys., 81, 056902
Cameron E., 2011, PASA, 28, 128
Campana R., 2022, preprint (arXiv:2205.09359)
Chandrasekhar S., 1960, Radiative transfer. Dover, New York
Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2019, Open J. Astrophys., 2, 10
Hebeler K., 2021, Phys. Rept., 909, 1

MNRAS 552, 3389–3404 (2023)
Figure A1. Three snapshots showing the evolution of the pulse profile corresponding to the light curve in Fig. 1.

Figure A2. Radius (left) and mass (right) cumulative fraction of parameter values recovered within a credible interval as a function of the credible intervals for each model using the variability data set. The grey regions, from most opaque to least, depict the cumulative 1, 2, and 3σ confidence intervals. The black solid line represents the expectation for unbiased parameter estimates. Each model and their corresponding p-values are shown on top of the figure.
### Table A1. Parameters used to produce the variability data set; $\sim$: time dependent parameter, $\times$: parameter not used for this case. $sX$ denotes parameter set $X$ of the corresponding model, $X \in \{1, 2, 3\}$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$ST-H_I$</th>
<th>$ST-H_{IR}$</th>
<th>$ST-\bar{H}_I$</th>
<th>$ST-\bar{H}_{IR}$</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M (M_\odot)$</td>
<td>1.5</td>
<td>2.1</td>
<td>1.3</td>
<td>2.0</td>
<td>1.2</td>
</tr>
<tr>
<td>$R$ (km)</td>
<td>13.0</td>
<td>12.3</td>
<td>12.2</td>
<td>14.5</td>
<td>12.5</td>
</tr>
<tr>
<td>$D$ (kpc)</td>
<td>7.5</td>
<td>3.5</td>
<td>1.0</td>
<td>5.8</td>
<td>6.1</td>
</tr>
<tr>
<td>$i$</td>
<td>30</td>
<td>13</td>
<td>12</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>$\phi_{\text{exp}}$ (cycles)</td>
<td>0.3</td>
<td>0.6</td>
<td>0.9</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>$\theta_{\text{exp}}$ (radian)</td>
<td>20</td>
<td>47</td>
<td>42</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>$\xi_{\text{exp}}$ (radian)</td>
<td>0.75</td>
<td>0.5</td>
<td>0.1</td>
<td>$\sim$</td>
<td>$\sim$</td>
</tr>
<tr>
<td>$\log [T_{\text{exp}}(K)/1K]$</td>
<td>$\sim$</td>
<td>$\sim$</td>
<td>$\sim$</td>
<td>$\sim$</td>
<td>$\sim$</td>
</tr>
<tr>
<td>$\log [T_{\text{ext}}(K)/1K]$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$N_H$ ($10^{23}$ cm$^{-2}$)</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

### Table A2. Parameters used to produce the constant data set; $\times$: parameter not used for this case. $sX$ denotes parameter set $X$ of the corresponding model, $X \in \{1, 2, 3\}$. The models indicated correspond to their variability counterparts, based on which synthetic data were generated using the method outlined in Section 2.6.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$ST-H_I$</th>
<th>$ST-H_{IR}$</th>
<th>$ST-\bar{H}_I$</th>
<th>$ST-\bar{H}_{IR}$</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M (M_\odot)$</td>
<td>1.5</td>
<td>2.1</td>
<td>1.3</td>
<td>2.0</td>
<td>1.2</td>
</tr>
<tr>
<td>$R$ (km)</td>
<td>13.0</td>
<td>12.3</td>
<td>12.2</td>
<td>14.5</td>
<td>12.5</td>
</tr>
<tr>
<td>$D$ (kpc)</td>
<td>7.5</td>
<td>3.5</td>
<td>1.0</td>
<td>5.8</td>
<td>6.1</td>
</tr>
<tr>
<td>$i$</td>
<td>30</td>
<td>13</td>
<td>12</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>$\phi_{\text{exp}}$ (cycles)</td>
<td>0.3</td>
<td>0.6</td>
<td>0.9</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>$\theta_{\text{exp}}$ (radian)</td>
<td>20</td>
<td>47</td>
<td>42</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>$\xi_{\text{exp}}$ (radian)</td>
<td>0.75</td>
<td>0.5</td>
<td>0.1</td>
<td>$\sim$</td>
<td>$\sim$</td>
</tr>
<tr>
<td>$\log [T_{\text{exp}}(K)/1K]$</td>
<td>$\sim$</td>
<td>$\sim$</td>
<td>$\sim$</td>
<td>$\sim$</td>
<td>$\sim$</td>
</tr>
<tr>
<td>$\log [T_{\text{ext}}(K)/1K]$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$N_H$ ($10^{23}$ cm$^{-2}$)</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

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