Astronomy
&
Astrophysics

Bayesian parameter constraints for neutron star masses and radii using X-ray timing observations of accretion-powered millisecond pulsars

T. Salmi[1], J. Nättilä[2], and J. Poutanen[1,2,3]

[1] Tuorla Observatory, Department of Physics and Astronomy, 20014 University of Turku, Finland
  e-mail: thjsal@utu.fi
[2] Nordita, KTH Royal Institute of Technology and Stockholm University, Roslagstullsbacken 23, 10691 Stockholm, Sweden
[3] Space Research Institute of the Russian Academy of Sciences, Profsoyuznaya str. 84/32, 117997 Moscow, Russia

ABSTRACT

We present a Bayesian method to constrain the masses and radii of neutron stars (NSs) using the information encoded in the X-ray pulse profiles of accreting millisecond pulsars. We model the shape of the pulses using "oblate Schwarzschild" approximation, which takes into account the deformed shape of the star together with the special and general relativistic corrections to the photon trajectories and angles. The spectrum of the radiation is obtained from an empirical model of Comptonization in a hot slab in which a fraction of seed black-body photons is scattered into a power-law component. By using an affine-invariant Markov chain Monte Carlo ensemble sampling method, we obtain posterior probability distributions for the different model parameters, especially for the mass and the radius. To test the robustness of our method, we first analysed self-generated synthetic data with known model parameters. Similar analysis was then applied for the observations of SAX J1808.4−3658 by the *Rossi* X-ray Timing Explorer (RXTE). The results show that our method can reproduce the model parameters of the synthetic data, and that accurate constraints for the radius can be obtained using the RXTE pulse profile observations if the mass is a priori known. For a mass in the range 1.5–1.8 $M_\odot$, the radius of the NS in SAX J1808.4−3658 is constrained between 9 and 13 km. If the mass is accurately known, the radius can be determined with an accuracy of 5% (68% credibility). For example, for the mass of 1.7 $M_\odot$ the equatorial radius is $R_{eq} = 11.9^{+0.5}_{-0.4}$ km. Finally, we show that further improvements can be obtained when the X-ray polarization data from the Imaging X-ray Polarimeter Explorer will become available.

Key words. pulsars: individual: SAX J1808.4–3658 – stars: neutron – X-rays: binaries – X-rays: stars

1. Introduction

Neutron stars (NSs) are the densest directly observable objects in the Universe. The matter inside a NS is at supranuclear densities. The pressure–density–temperature relation of such dense matter is described by the equation of state. There is one-to-one mapping relating the equation of state to the mass–radius dependence of the NS (see, for example, Lindblom 1992; Lattimer 2012). Thus, measurements of the NS masses and radii from observations can in principle help us to constrain microphysical properties of high-density matter. One way to obtain information on these parameters is to study pulse profiles produced by hotspots on the surface of rapidly spinning NSs (Watts et al. 2016), in particular, in accreting millisecond pulsars (AMPs).

In accreting millisecond pulsars, the matter from a low-mass companion star accretes onto the magnetic poles of a rapidly rotating NS, forming two hotspots on the stellar surface. Because the hotspots are misplaced from the rotational axis of the NS, the X-ray radiation observed from these spots pulsates coherently at the spinning frequency of the NS. The pulse profiles carry information about the mass and radius of a NS because the light bending and thus pulse shape depends strongly on the compactness of the star (Poutanen 2008). However, many other physical parameters and observing angles affect both the light curves and the spectra of these sources.

To model the pulse profiles from the hotspots of NSs, the necessary formalism was first developed by Pechenick et al. (1983) and Page (1995). For rapid rotation the formalism was extended by Miller & Lamb (1998), Weinberg et al. (2001), and Poutanen & Gierliński (2003). The transformation of polarization was studied by Viironen & Poutanen (2004). A simple approximation for light bending was given in Beloborodov (2002). Other useful analytical results were obtained by Poutanen & Beloborodov (2006). How deviations from the Schwarzschild metric affect the profiles have been studied by Braje et al. (2000; using the Kerr metric), who corrected previous errors by Chen & Shaham (1989) and showed that the effects are rather small. The effects of oblateness of the star due to rapid rotation were studied by Morsink et al. (2007) using Schwarzschild metric and by Cadeau et al. (2007) using a numerically-generated metric for rapidly rotating NSs in general relativity. A ray-tracing algorithm using Hartle–Thorne metric was examined by Bauböck et al. (2012). In addition, Nättilä & Pihajoki (2018) showed that the special relativistic rotational effects emerge directly from a general relativistic treatment of a rotating compact object.

In order to infer the stellar mass and radius, with help of the forthcoming observations, the pulse profiles were investigated by Psaltis et al. (2014) using approximate relations between the modelled and observed parameters. Previously, mass and radius constraints using pulse profile modelling for thermonuclear

burst oscillations and Bayesian analysis were studied by Lo et al. (2013) and Miller & Lamb (2015). Mass–radius constraints were also studied by Stevens et al. (2016) using evolutionary optimization algorithm. They as well considered the thermonuclear burst oscillations and used the synthetic data, however, keeping spot size and temperature fixed.

In this paper, our aim is to get new information of mass–radius relation using the energy-resolved pulse profiles of accretion-powered millisecond pulsations, for which we have plenty of already existing data (see e.g. Gierliński et al. 2002; Poutanen & Gierliński 2003; Gierliński & Poutanen 2005; Leahy et al. 2008; Morsink & Leahy 2011). On the other hand, we also want to show how accurately NS parameters could be constrained with help of the upcoming polarimeters and X-ray missions like the Imaging X-ray Polarimeter Explorer (IXPE; Weisskopf et al. 2016) and the enhanced X-ray Timing and Polarimetry mission (eXTP; Zhang et al. 2016). We do the full Bayesian analysis with both synthetic data and observations of SAX J1808.4−3658, and fit the data simultaneously in phase and energy dimensions. Like Miller & Lamb (2015), we sample the parameter space for synthetic data using a Markov chain Monte Carlo (MCMC) method. However, we have more free parameters, and instead of maximizing, we also marginalize the likelihoods over the distance and spot temperature. In particular, we now have the angular dependence of the emitted radiation (beaming) as a free parameter.

The remainder of this paper is structured as follows. In Sect. 2, we describe the methods we have used to fit the pulse profiles and spectra of accreting millisecond pulsars. In Sect. 3 we present our synthetic data as well as the original data for SAX J1808.4−3658 that we have used in our analysis. The results of the modelling are described in Sect. 4. We discuss the results in Sect. 5 and summarize in Sect. 6.

2. Methods

2.1. Pulse shape modelling

Our pulse shape modelling is based on the model introduced in Poutanen & Beloborodov (2006), which takes the special and general relativistic effects into account by using the Schwarzschild-Doppler approximation. The effects of general relativity (gravitational light bending) are modelled as though the star is not rotating describing the exterior metric with a Schwarzschild solution. On the other hand, rotational effects have been approximated by using special relativistic Doppler transformations and angle aberrations as though the star is a rotating object with no gravitational field.

In Schwarzschild geometry we know the exact relation between the photon emission angle α and the deflection angle ψ. The angle α is measured between the radial direction $\boldsymbol{r}$ and the initial direction of photon $\boldsymbol{k_0}$ (see Fig. 1), and the angle ψ between $\boldsymbol{r}$ and observer $\boldsymbol{k}$ (see Fig. 1). When $\alpha < \pi/2$, it is given by (e.g. Misner et al. 1973; Pechenick et al. 1983; Beloborodov 2002)

$$\psi_{\mathrm{p}}(R, \alpha) = \int_R^\infty \frac{\mathrm{d}r}{r^2} \left[\frac{1}{b^2} - \frac{1}{r^2}\left(1 - \frac{r_{\mathrm{S}}}{r}\right) \right]^{-1/2}, \qquad (1)$$

where b is the impact parameter,

$$b = \frac{R}{\sqrt{1-u}} \sin\alpha, \qquad (2)$$

$u \equiv r_{\mathrm{S}}/R$, $r_{\mathrm{S}} = 2GM/c^2$ is the Schwarzschild radius, G is the gravitational constant, c is the speed of light, R is the radius of
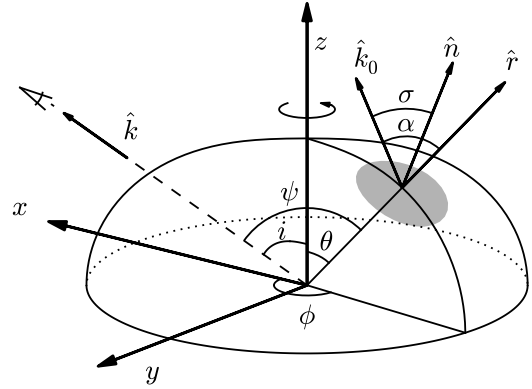

Fig. 1. Geometry of the problem. The Cartesian (x, y, z) and spherical (with θ and ϕ) coordinate systems are shown. The star has an oblate shape, because it is rotating rapidly around the z-axis. Therefore, the radial vector $\hat{r}$ and the surface normal $\hat{n}$ differ from each other. The photon from the hot spot is initially emitted towards $\hat{k}_0$ but due to the light bending it reaches the observer in direction of $\hat{k}$.

the star at the point which emits the photon, and M is the mass of the NS. A numerical method to evaluate integral (1) is presented in Appendix A. For an oblate star, there is a possibility to have photon trajectories with decreasing radial coordinate (i.e. photons emitted towards the star) corresponding to $\alpha > \pi/2$. In this case, the relation between α and ψ takes the form

$$\psi(R, \alpha) = 2\psi_{\max} - \psi_{\mathrm{p}}(R, \pi - \alpha), \qquad (3)$$

where $\psi_{\max} = \psi_{\mathrm{p}}(p, \alpha = \pi/2)$ and p is the distance of closest approach, given by

$$p = -\frac{2}{\sqrt{3}} b \cos\left([\arccos\left(3\sqrt{3} r_{\mathrm{S}}/2b\right) + 2\pi]/3 \right). \qquad (4)$$

The original model of Poutanen & Beloborodov (2006) has been expanded to take into account the geometrical effects of the oblate shape of the NS as described by Morsink et al. (2007). The oblate shape of the star is obtained by using a function from AlGendy & Morsink (2014), which gives the radius R of the star as a function of co-latitude. The geometry of the system with the star, hot spot, and observer in the direction $\boldsymbol{k}$ is presented in Fig. 1. Here θ is the co-latitude of the spot, i is the observer inclination, and ϕ is the spot phase angle. Due to the oblateness, the local surface normal $\boldsymbol{n}$ and radius vector $\boldsymbol{r}$ are not always pointing to the same direction. Therefore, to check the visibility condition, we need to consider the emission angle σ relative to $\boldsymbol{n}$, instead of α. Only photons with emission angle $\sigma \leq \pi/2$ can reach the observer. However, the emitted photon might not be visible, even though $\cos\sigma > 0$, if the photon hits the NS surface at any later phase of its trajectory. We tested this requirement empirically by studying different photons emitted almost parallel to the surface from various emission locations. Photons were propagated from the surface and their radial location was tested to see if the trajectory intersected with the oblate surface. These tests show that (at least for all the cases we considered) a photon initially satisfying the requirement $\cos\sigma > 0$ will always reach the distant observer.

Hereafter, all the primed quantities are measured in the co-rotating frame of the spot. The infinitesimal spot area measured in the frame co-moving with the spot may be calculated as (Morsink et al. 2007)

$$\mathrm{d}S'(\theta) = \gamma R^2(\theta)[1 + f^2(\theta)]^{1/2} \sin\theta \, \mathrm{d}\theta \mathrm{d}\phi, \qquad (5)$$

where the Lorentz factor $\gamma = 1/\sqrt{1-\beta^2}$, β is the spot velocity in units of c,

$$f(\theta) = \frac{1+z}{R}\frac{dR}{d\theta}, \tag{6}$$

and $1 + z = 1/\sqrt{1-u}$ is the gravitational redshift. The factor $[1 + f^2(\theta)]^{1/2}$ takes into account the oblateness of the spot surface. The Lorentz factor γ originates from considering a differential area element in rotating coordinates. It results from the fact, that the surface element dS' here is defined based on simultaneous measurements of co-moving observers, instead of using simultaneous measurements of local static observers (Nättilä & Pihajoki 2018; Lo et al. 2018).

Following derivation presented by Poutanen & Beloborodov (2006), we can get the observed spectral flux

$$dF_E = (1-u)^{1/2}\delta^4 I'_{E'}(\sigma')\cos\sigma\frac{d\cos\alpha}{d\cos\psi}\frac{dS'}{D^2}, \tag{7}$$

where D is the distance to the star, $I'_{E'}$ is the spectral intensity and σ' is the emission angle relative to the normal as measured in the co-moving frame of the spot,

$$\delta = \frac{1}{\gamma(1-\beta\cos\xi)} \tag{8}$$

is a Doppler factor, and ξ is the angle between the spot velocity and initial direction of the photon. We note that the only difference from Eq. (A15) given in Poutanen & Beloborodov (2006) is the change of α' to σ' in the argument of the intensity and in $\cos\sigma$, which accounts for the projection of the spot area on the observer sky.

The flux in expression (7) depends on the pulsar phase ϕ (because e.g. the Doppler factor δ depends on it). This flux actually corresponds to an observed phase ϕ_{obs}, which is different from ϕ due to light travel delays. The time delay is caused by different travel times of emitted photons to the observer, depending on the position of the emitting point. A photon following the trajectory with an impact parameter b (and $\alpha < \pi/2$) is lagging a photon originating at the same radius R with impact parameter $b = 0$ by (Pechenick et al. 1983)

$$c\Delta t_p(R, \alpha) = \int_R^\infty \frac{dr}{1-r_S/r}\left\{\left[1 - \frac{b^2}{r^2}\left(1 - \frac{r_S}{r}\right)\right]^{-1/2} - 1\right\}. \tag{9}$$

Because we need to account for the stellar oblateness, we have to compute all the delays relative to the photons emitted at some reference radius r_{ref} (which can be chosen arbitrarily as long as it exceeds r_S) and $b = 0$ (i.e. $\alpha = 0$). Photons with $b = 0$ emitted at radius R will be lagging photon emitted at r_{ref} by

$$c\delta t(R, r_{\text{ref}}) = r_{\text{ref}} - R + r_S\ln\left(\frac{r_{\text{ref}} - r_S}{R - r_S}\right). \tag{10}$$

Thus the total delay of photons emitted at R and angle α (with corresponding impact parameter given by Eq. (2)) relative to the photons emitted at r_{ref} with $b = 0$ is

$$\Delta t(R, \alpha) = \Delta t_p(R, \alpha) + \delta t(R, r_{\text{ref}}). \tag{11}$$

In our model, we use the equatorial radius as a reference radius ($r_{\text{ref}} = R_{\text{eq}}$). A numerical method to evaluate integral (9) is presented in Appendix A.

In the case of $\alpha > \pi/2$, the corresponding delay can be calculated using the symmetry around distance p of the closest approach as

$$\begin{aligned}\Delta t(R, \alpha) &= 2\Delta t(p, \pi/2) - \Delta t(R, \pi - \alpha)\\ &= 2\Delta t_p(p, \pi/2) - \Delta t_p(R, \pi - \alpha)\\ &\quad + 2\delta t(p, r_{\text{ref}}) - \delta t(R, r_{\text{ref}}).\end{aligned} \tag{12}$$

Otherwise, the computation of the light curve follows Poutanen & Beloborodov (2006) with corrections for oblateness as described in Morsink et al. (2007). One notable difference is, however, that we are considering finite-sized spots which may be also large in size. In order to compute the flux from a spot with finite size, we need to integrate over the spot surface, in other words we integrate Eq. (7) with dS' given by Eq. (5) over the solid angle. It is done by splitting the spot into a number of small sub-spots and computing the fluxes to each sub-spot separately. For a circular spot around the magnetic pole, instead of using θ and ϕ are variables, it is easier to integrate over the solid angle using Gaussian quadrature in the cosine of the magnetic co-latitude (i.e. the angle measured from the magnetic pole) and using trapezoidal rule for integration over the corresponding azimuth. For each phase bin and sub-spot we need to evaluate light bending and time delays separately, because each sub-spot is located at a different radius due to the oblateness.

We have also made some physical assumptions to simplify the model. We assume a spot which has a constant angular radius ρ, which is the angle between the centre and the edge of the spot measured from the centre of the star. In this paper we also model the pulse profiles using only one hot spot (see e.g. Ibragimov & Poutanen 2009). In addition, we ignore the radiative transfer effects related to photon propagation through the accretion stream. See Sect. 5 for discussion how these assumptions may affect the results.

2.2. Spectrum of the radiation

Next we present how the energy and angular spectra of radiation are modelled, or how the intensity $I'_{E'}(\sigma')$ depends on energy E' and emission angle σ'. The material accreted onto a NS is abruptly decelerated close to the surface releasing its kinetic energy and forming a hotspot. The energy spectrum of AMPs above 3 keV can be well represented as a sum of the black-body-like component at lower energies and a power-law-like component extending to ~100 keV and likely produced by thermal Comptonization (e.g. Poutanen & Gierliński 2003; Gierliński & Poutanen 2005; Falanga 2008). We can associate the black-body component with the emission from the heated NS surface and the Comptonization components with the accretion shock or very surface layers heated to 50–100 keV by bombarding particles. The seed photons for Comptonization are likely coming from the cooler layers immediately under the shocked plasma. In this work we thus use a two-component model for the spectrum.

For each point inside the spot, we represent the intensity as a sum of the black-body (I'_{bb}) and Comptonization (I'_{comp}) parts. Following Poutanen & Gierliński (2003), we assume that the intensities of both black-body and Comptonization components can be expressed as products of functions that depend either only on energy E' or only on emission angle σ'. For the black-body component we use the Planck law for the specific intensity, in other words we assume the angular pattern to be isotropic, which means that $I'_{\text{bb}}(\sigma') = I'_{\text{bb}}$.

2.2.1. Radiation from hot spots

In this paper we approximate the spectrum with Comptonization model SIMPL introduced by Steiner et al. (2009) and implemented into XSPEC (Arnaud 1996). This model is based on the solution of the Kompaneets equation describing Compton scattering in non-relativistic plasmas (Sunyaev & Titarchuk 1980). In this model a fraction of photons of a seed black-body spectrum is Compton scattered into a power-law-like component with photon spectral index Γ:

$$I'_{E'} \propto E'^{-(\Gamma-1)}. \tag{13}$$

The parameters of the SIMPL model are Γ and the scattered fraction X_{sc} of black-body photons. In addition, we take into account only the up-scattered photons. The model SIMPL-2, designed for non-relativistic thermal Comptonization, would take into account also the down-scattered photons. However, since the difference between the models is small, we use the more simple model SIMPL-1 introduced by Steiner et al. (2009). Later we discuss, how the results would change if SIMPL-2 was used (see Sect. 5). The input black-body spectrum is convolved with the Green's function describing the scattering (see Eqs. (1) and (4) in Steiner et al. 2009) into a Comptonized power-law-like spectrum.

For a given temperature T' of the spot, we compute a seed black-body spectrum $I'_{bb}(E')$ at a grid of photon energies logarithmically spaced from 0.1 to 100 keV with a step $\Delta \log E' = 0.02$. Using then SIMPL, we obtain also a table of intensities for the Comptonized spectrum $I'_{comp}(E')$. For each emitting point, we compute the intensity as

$$I'_{E'}(\sigma') = (1 - X_{sc})I'_{bb}(E') + f(\sigma')I'_{comp}(E'), \tag{14}$$

where the black body is assumed to be isotropic and the angular dependence of the Comptonized emission is given by the function

$$f(\sigma') = I_0(1 + h \cos \sigma'), \tag{15}$$

with h being the (free) anisotropy parameter and I_0 being the normalization, such that $2 \int_0^{\pi/2} \cos \sigma' f(\sigma') \sin \sigma' d\sigma' = 1$.

2.2.2. Disc reflection

In order to fit the spectrum of SAX J1808.4−3658, we need also take into account the reflection of the photons from the accretion disc. In addition to the Compton reflection continuum that produces spectral hardening above ∼10 keV, a fluorescent Kα iron line close to 6.4 keV is emitted as observed by Gierliński et al. (2002). To describe reflection, we use the XSPEC model XILCONV, which is a convolution model combining reflection model XILLVER from an ionized disc (García et al. 2013) with the Compton reflection code by Magdziarz & Zdziarski (1995) to give a more correct representation of Compton recoil effect. It is a modified version of the RFXCONV model described in Kolehmainen et al. (2011).

We use the XILCONV algorithm to re-compute all the energy spectra separately in each of our modelled phase bins. The original phase-resolved spectra are given as input. We fix most of the additional parameters of the reflection model at the best-fits of the observed phase-average spectrum of SAX J1808.4−3658 with the model PHABS×(XILCONV(SIMPL(BBODYRAD))). We assume that the redshift parameter is zero, because the accretion disc is not expected to be very close to the compact object. We also assume that the exponential cut-off energy is 300 keV. From the

XSPEC fits we get that the iron abundance is ten times higher than the solar iron abundance, and that the ionization parameter $\log \xi \approx 3.2$. The relative reflection normalization X_{refl} is still kept as a free parameter, as well as the inclination angle i (same angle as in the pulse profile modelling).

2.2.3. Response convolution

In this work, the instrumental response of the detector has been taken into account by convolving all modelled physical fluxes with response matrix of the Proportional Counter Array (PCA) at the *Rossi* X-ray Timing Explorer (RXTE) satellite. This is done by integrating the product of modelled photon number flux $N(E)$ and response matrix $R(n, E)$ over the energies E to get modelled counts $C_{model}(n)$ in each observed energy channel n. The integration is performed from the lowest response energies up to $E_{max} = 60$ keV (above which the source flux would contribute negligibly to the counts observed in the RXTE/PCA band below ∼20 keV) using equation

$$C_{model}(n) = T_{obs} \int_0^{E_{max}} N(E)R(n, E) dE, \tag{16}$$

where T_{obs} is the observing time. The values of photon flux $N(E)$ are computed in 50 logarithmic energy bins from 1 keV to E_{max}. The integration is performed using denser energy grid of the response matrix. The corresponding values of $\log N(E)$ are then linearly interpolated to the desired $\log E$.

In addition, the detected instrumental background of SAX J1808.4−3658 is kept fixed and it is added to the modelled counts. These predicted counts are then fitted with the observed counts as explained in the following sections. However, only the 24 observational energy channels between 3 and 18 keV have been used. This is based on the best sensitivity of proportional counter units of RXTE and the exclusion of the highest energies where instrumental background dominates for SAX J1808.4−3658 (Hartman et al. 2008).

Interstellar photon absorption affecting at the lowest energies has been taken into account by using neutral hydrogen column density for interstellar absorption N_H as one of the free parameters in our model. The computed amount of absorption is based on the PHABS-function in XSPEC. In practice, this has very little effect because interstellar absorption is severely affecting the spectra only below 1 keV. Hence, instead of using the typically more accurate TBABS we can simplify the treatment by adopting the PHABS model. The difference between these models was investigated by fitting the phase-averaged spectrum of SAX J1808.4−3658 with XSPEC using both the model PHABS×(XILCONV(SIMPL(BBODYRAD))) and the model TBABS×(XILCONV(SIMPL(BBODYRAD))). Our fits show that the difference between the fitted parameters with the two models are at most at 1% level, and the difference in the fitted flux is almost indistinguishable, even at the lowest energy channels of 3 keV.

2.3. Bayesian modelling

In order to constrain parameters of NSs we use Bayesian inference combined with the pulse shape and spectral model presented in Sects. 2.1 and 2.2. We fit the model to the observed or synthetic data and use Markov chain Monte Carlo (MCMC) methods to integrate over the parameter space to find the most probable values for the parameters of the model. As a first step, we use synthetic data in order to test our methods and tools. The aim of this work is to determine the credible regions for all the

parameters in the pulse profile model, given the synthetic and observed waveforms. In particular, the credible regions for the mass and radius are obtained by marginalizing over the uncertainties in other parameters.

We are interested in the probability of the parameters **y** of the waveform model when the observed waveform is known. The probability distribution of the parameters given the data is $p(\mathbf{y}|\mathcal{D})$, where $\mathcal{D}$ is the energy- and phase-resolved waveform data. According to the Bayes' theorem this (posterior) probability distribution can be obtained from the likelihood of the data, given the parameter values as (see e.g. Grinstead & Snell 1997)

$$p(\mathbf{y}|\mathcal{D}) \propto p(\mathcal{D}|\mathbf{y})p(\mathbf{y}), \tag{17}$$

where $p(\mathcal{D}|\mathbf{y})$ is the likelihood or the probability distribution of the data given the parameters. The factor $p(\mathbf{y})$ is the prior probability distribution of the parameter values. As a first approximation we use uniform priors for most of our parameters (see discussion of the priors in Sect. 5). The constant of proportionality is the inverse of the normalization factor, but it is irrelevant when estimating the values of the parameters in a given model.

2.3.1. Probabilities

The likelihood of the data for each sample $p(\mathcal{D}|\mathbf{y})$ is calculated by fitting the data to the pulse profile model. To be more specific, we assume that the probability density of the data counts C_{data} given the modelled counts C_{model} is normally distributed around C_{model} with the square root of the modelled counts as standard deviation similar to Pearson's chi-squared test (Cash 1979). Adding also intrinsic scatter of the model and the calibration error of the instrument, we have

$$p(C_{\mathrm{data}}|C_{\mathrm{model}}) = \frac{\exp\left(\frac{-(C_{\mathrm{data}}-C_{\mathrm{model}})^2}{2(C_{\mathrm{model}}+\sigma_{\mathrm{i}}^2+\sigma_{\mathrm{c}}^2)}\right)}{\sqrt{2\pi(C_{\mathrm{model}}+\sigma_{\mathrm{i}}^2+\sigma_{\mathrm{c}}^2)}}, \tag{18}$$

where the intrinsic scatter σ_{i} includes the error of the model in describing the data, and $\sigma_{\mathrm{c}} = 0.005 \times C_{\mathrm{model}}$ includes the calibration error of the detector (see Sect. 3). In other words, intrinsic scatter is the measure of the systematic errors coming from the choice of the model. It is a free parameter in the model: if the data are not fully described by the model, it leads to an increase both in σ_{i} and in the credible regions of other parameters. Normalization is needed because it will not cancel out when calculating likelihood ratios (because of the intrinsic scatter and calibration error). In an optimal case, σ_{i}^2 is very small compared to the count noise term $\sqrt{C_{\mathrm{model}}}$. The total probability density of the data given the sample is the product of $p(C_{\mathrm{data}}|C_{\mathrm{model}})$ from each phase and energy bin including at least 20 observed photons (or the sum in case of log-probabilities).

We have also made tests to see that the similar-looking posterior probabilities for pulse profile parameters are also obtained by using Poisson probabilities:

$$p(C_{\mathrm{data}}|C_{\mathrm{model}}) = \exp(-C_{\mathrm{model}})\frac{C_{\mathrm{model}}^{C_{\mathrm{data}}}}{C_{\mathrm{data}}!}. \tag{19}$$

The phase shift is treated as a nuisance parameter in our modelling. We calculate the probability densities, the products of $p(C_{\mathrm{data}}|C_{\mathrm{model}})$ over phase and energy bins, using different phase shifts. A bisection method is used to find the phase shift which has the highest probability. As final probability $p(\mathcal{D}|\mathbf{y})$ we use the solution with the most probable phase shift. Instead of maximizing the likelihood, we could have also marginalized likelihoods over all phase shifts, but the difference between these two

methods was found to be marginal. However, marginalization requires a denser and more evenly distributed phase shift grid than what is possible by using only bisection method (at least in case of very flat light curves).

2.3.2. Sampling methods

To get constraints for our model parameters, we make samples from the posterior distributions using an affine invariant ensemble sampler, which is a MCMC method described in Goodman & Weare (2010). The algorithm has a similar structure to the normal Metropolis scheme and still uses a proposal and either accepts or rejects a step. But instead of evolving only one sample value, it evolves an ensemble of sample values, called walkers, together. On each iteration the algorithm generates a new sample for every walker using the current positions of all the other walkers in the ensemble. The affine invariant trial move, that we use here, is the so-called stretch move. In this method each walker is moved using only one randomly selected complementary walker.

When using the ensemble sampler method to data with significantly less than 1% errors (and using more than a few free parameters), we encountered problems with walkers either getting stuck into local optima or being unable to move efficiently enough due to the non-linear parameter degeneracies. The convergence of the posterior probability distributions was also very slow, even though we used multiple independent ensembles.

To solve these issues, we made few variable transformations (see Sect. 2.3.3 for more detailed discussion). Secondly, we used more precise starting limits for the priors in parameters that could be fitted separately with xspec (spectral parameters) or could be approximated from other studies (size of the spot and distance). The recommended technique is to start walkers in a small sphere around the a priori preferred position. However, for the radius, mass, observer inclination, and the magnetic inclination, we allowed a larger range of initial positions. The exact starting limits ranged from 0.01% to 20% of the corresponding widths of the final priors shown in Figs. 2, 5, and 6 (with the exceptions mentioned in Sect. 4). The starting points for every walker were drawn randomly from this ball, which was surrounding our initial guess. For each ensemble, the initial guess was a point obtained by perturbing the correct solution of synthetic data by 4% of the width of the final priors in each dimension (see Sect. 3.2 for the correct solution for the synthetic data).

Finally, we also applied a clustering method described in Hou et al. (2012). At certain moments during the burn-in phase of the sampling, the walkers with worst-fitting parameters were drawn to the vicinity of the best-fitting walkers of the same ensemble, if the gap between log-likelihoods of two adjacent walkers (ordered according to decreasing likelihood) became too high compared to differences of other adjacent walkers. We also tried to use the simulated annealing method[1] described again in Hou et al. (2012), but found no significant help from that.

---

[1] In this method an "artificial temperature" is introduced to create a modified posterior distribution allowing an exploration of parameter space without getting stuck into the local optima. The temperature is decreased slowly until the true posterior distribution is obtained. However, in our case we did not find any cooling schedule that would be computationally short enough, and would have significantly helped to reduce the multiple solutions found by different ensembles with different starting positions.
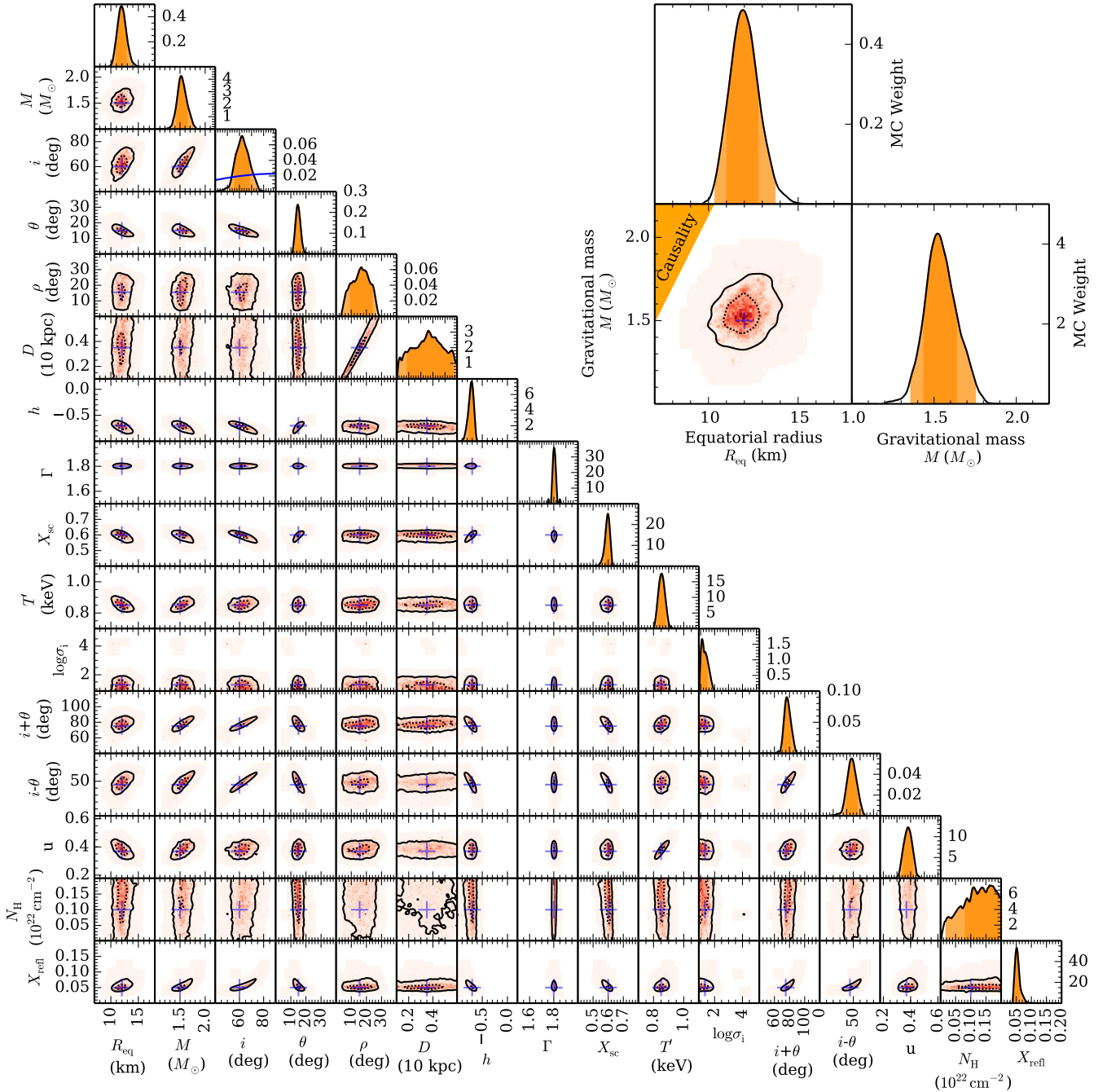
Fig. 2. Posterior probability distributions for the MCMC runs with the synthetic data. The dark orange colour shows a 68% and the light orange colour a 95% highest posterior density credible interval. In the 2D posterior distributions the solid contour shows a 95% and the dashed contour a 68% highest posterior density credible region. The blue crosses show the correct solution. The $\sin i$ prior in i is shown with a blue line. The compactness parameter u is defined here as $u \equiv r_S/R_{eq}$. The inset in the upper right corner shows the mass–radius posterior distribution in more detail.

In addition, we have checked our results using a multi-modal nested sampling algorithm, implemented in MultiNest (Feroz et al. 2009)[2], instead of the ensemble sampler. It is a Monte Carlo method targeted at the efficient calculation of Bayesian evidence for a model, but it also produces posterior samples as a by-product. The method should have improved efficiency especially for posterior distributions that may contain multiple probability maxima and pronounced degeneracies in high dimensions, that are multi-modal distributions. To compute the posterior probability distributions, we used the PyMulti-Nest[3] package of Python.

2.3.3. Variable transformations

In all of our models, we have 13 free parameters (in addition to the phase shift). The parameters are listed in Table 1. To improve the performance of our method, we have used several parameter transformations given as $i + \theta$, $i - \theta$, M/R_{eq}, $(R_{eq}\rho(1 + z_{eq})/D)^2$,

---

[2] https://ccpforge.cse.rl.ac.uk/gf/project/multinest/

[3] https://github.com/JohannesBuchner/PyMultiNest

Table 1. Parameters of the synthetic data.

Parameter	Value
Equatorial radius R_{eq}	12.0 km
Mass M	1.5 $M_\odot$
Inclination i	60°
Spot co-latitude θ	15°
Spot angular radius ρ	15.5°
Distance D	3.5 kpc
Temperature T'	0.85 keV
Beaming parameter h	−0.7
Scattered photon fraction X_{sc}	0.6
Photon spectral index Γ	1.8
Intrinsic scatter σ_i	0.0
Hydrogen column density N_H	10^{21} cm^{-2}
Relative reflection from accretion disc X_{refl}	0.05

$(T'/(1 + z_{eq}))^4$, $X_{refl} \cos i$, and $\log \sigma_i$ as our parameters instead of i, θ, R_{eq}, ρ, T', X_{refl}, and σ_i, when sampling the parameter space (see discussion in Sect. 5.1). Here z_{eq} is the gravitational redshift at $R = R_{eq}$.

The choice of sampling $i + \theta$ and $i - \theta$ is easy to explain, because the pulse profile is nearly degenerate to switching i and θ (see e.g. Viironen & Poutanen 2004, and Sect. 5.1) and depends only on $x \equiv \sin i \sin \theta$ and $y \equiv \cos i \cos \theta$. Therefore, it is easier to get constraints for x and y, or, alternatively, for $y + x$ and $y - x$. Because $i - \theta = \pm \arccos(x + y)$ and $i + \theta = \arccos(y - x)$, we get better constraints also for $i + \theta$. For $i - \theta$ we expect still a bimodal distribution, if there is no prior information of these angles. The inclination has been constrained to $i = 36° - 67°$ using optical observations (Deloye et al. 2008). The X-ray analysis of the 2002 outburst (Ibragimov & Poutanen 2009) led to a similar constraint of $i = 50° - 70°$. Modelling the broadened iron line with observations from *Suzaku* and *XMM-Newton*, Cackett et al. (2009) obtained $i = 51° - 63°$ with 90% confidence. In any case, due to the lack of X-ray eclipses, the inclination should be below $82° - 84°$, depending on the assumed NS mass (Chakrabarty & Morgan 1998). Using slightly more conservative values, we first limit the inclination still between 40° and 90° in our model, in order to be as model-independent as possible. Later we test, how the tighter constraints on inclination, will affect the results (see Sect. 5.3).

The choices of sampling mass–radius-ratio M/R_{eq}, normalization parameter $(R_{eq}\rho(1 + z)/D)^2$, the observed temperature in the fourth power $(T'/(1 + z))^4$, and the angle-corrected relative reflection $X_{refl} \cos i$ are based on our expectations of which variables most directly affect the observed pulse profiles (see Sect. 5.1). These parameter transformations are also aimed to help sampling with for example the coupled R_{eq} and ρ. In the case of σ_i, we sample $\log \sigma_i$, because the intrinsic scatter is a scale parameter of the model. Even though the original non-transformed parameters were not sampled, we show their posterior distributions in Sect. 4.

3. Data

3.1. SAX J1808.4−3658

The primary data used in this article is the phase-resolved energy spectrum of SAX J1808.4−3658 observed during its outburst in 1998 by RXTE. The energy-dependent pulse profiles are constructed in the same way as in Poutanen & Gierliński (2003).

We use the data binned in 16 phase bins and use the 24 energy channels which extend between 3 keV and 18 keV, as justified in Sect. 2.2.3. The data are obtained by combining observations of all the 5 proportional counters of the PCA/RXTE and observations between 1998 April 11 and 29 (identical to that used by Poutanen & Gierliński 2003). During this period the pulse shape stayed most of the time almost constant (see Fig. 3 in Hartman et al. 2008). Therefore, we do not expect large changes in best-fit mass and radius when using a shorter interval. This was also noted by Morsink & Leahy (2011). The exposure time of each phase bin is approximately 10.5 h, and the total number of detected counts is close to 4×10^7. This means that the mean statistical uncertainty in our 384 phase-energy-bins is approximately 0.3%. The systematic calibration error of PCA is about 0.5% (Jahoda et al. 2006). Although, the response files we use were generated in 2003 with a possibly slightly higher calibration error, this not critical, because the underestimated calibration error should be captured by the intrinsic scatter σ_i term when fitting the data (see Sect. 2.3.1).

3.2. Synthetic data

In order to test the method and measure its accuracy, we also apply the fitting framework for the synthetic data, which are generated using our model, and thus we know the result. The synthetic data in this work are designed to resemble the actual observations of SAX J1808.4−3658 as closely as possible. The amplitude of the modulation in the normalized number fluxes has been set close to the same value in the synthetic data as in the observations. We also use the same energy intervals and the same number of phase bins as in the actual data. In addition, the total number of counts is the same as in the observed data (see Sect. 3.1).

The synthetic data have been generated using our pulse profile model with parameters shown in Table 1. In addition, we have fixed the rotational frequency of the star at 401 Hz, according to the observations of SAX J1808.4−3658 (Wijnands & van der Klis 1998). We have also performed an extra phase shift to our synthetic pulse profile so that the best-fitting phase shift is not immediately found, which is also the case for the observed data.

When creating the data, we have used a higher phase resolution (500 phase bins) in the model than when actually fitting the data (128 phase bins). In both cases, the physical fluxes are calculated with 50 energy bins between 1 and 60 keV. The synthetic pulse profiles are interpolated and integrated to 16 phase bins (similar to the data of SAX J1808.4−3658) and convolved with the response matrix, in order to get light curves in the 24 PCA energy channels between 3 and 18 keV. In order to include the noise, we have Poisson sampled the detected counts at each phase point and energy channel. Thus, a relatively larger noise appears at the highest energy channels where the count rates are lowest. The 0.5% calibration error of the PCA instrument is included in to the fit because of the systematic uncertainties in the response matrix (see Sect. 2.3.1).

4. Results

4.1. Synthetic data

We begin our analysis with synthetic data created as described in the previous section. In the first model we keep all our parameters free and use the least constraining limits for prior
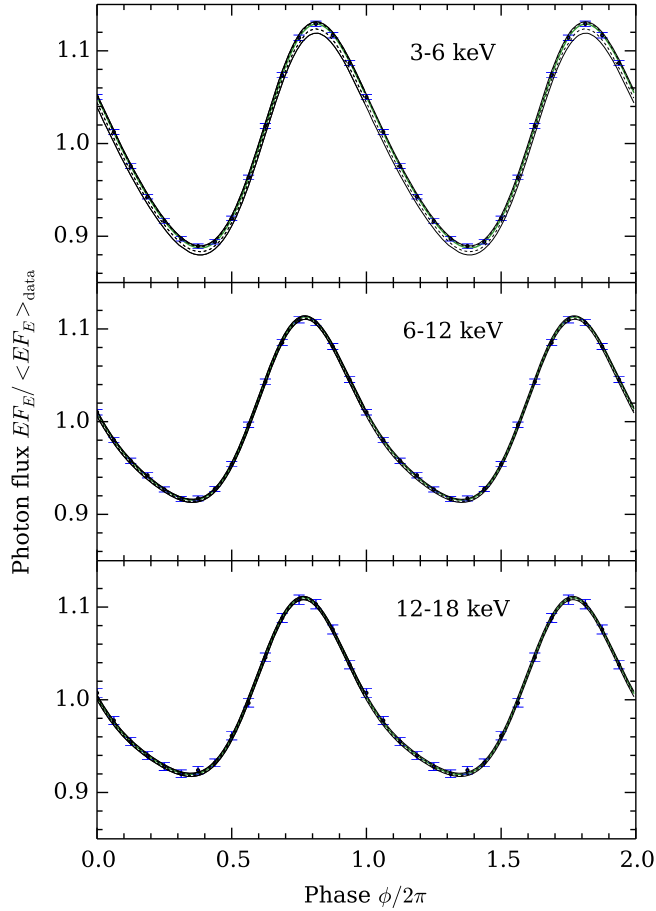
Fig. 3. Normalized pulse profiles integrated to 3 energy bins for the synthetic data. The solid black contour shows a 95% and the dashed black contour a 68% highest posterior density credible region. The green solid line shows the best-fit solution. The observed data converted to the physical units using the best-fit model are shown with blue circles with the error bars according to the Poisson noise. The 0.5% calibration error of the detector, used in the fitting procedure, is not included to the error bars.

distributions. The uniform prior probability is set for $\cos i$ instead of i, because we assume the observers to be uniformly distributed on a sphere around the star. Because $d\cos i \propto \sin i\, di$, we have $\sin i$ prior probability in i. Otherwise, we use uniform prior distributions for our sampling parameters. Of course, since the sampling parameters differ from the model parameters (see Sect. 2.3.3), the prior probabilities in R_{eq}, i, θ, ρ, T', and X_{refl} are multi-dimensional and non-uniform (and therefore not shown in the figures of this article).

The resulting posterior distributions are shown in Fig. 2. The intervals shown in this figure are also the boundaries of the prior distributions we have used, excluding the upper limit $h = 1.0$ for the beaming parameter and $\theta = 90°$ for the magnetic obliquity. Also, M/R_{eq}, $i - \theta$, and $i + \theta$ are not sampled in the range shown in Fig. 2, but in a range allowing all M, R_{eq}, i, and θ be inside their prior, but rejecting the step if any of the parameters would be outside of it. In addition, we have estimated a more strict upper limit for M/R_{eq} from the causality condition $u \equiv 2GM/R_{\mathrm{eq}}c^2 \lesssim 0.96 \times 2/3$ (Lattimer & Prakash 2004). The factor 0.96 is added to ensure that the causality condition is valid also at the pole of the oblate NS. We also limit the spot temperature in the observer frame T to be in the range $0.6-0.7\,\mathrm{keV}$, which is obtained by spectral fitting with xspec.
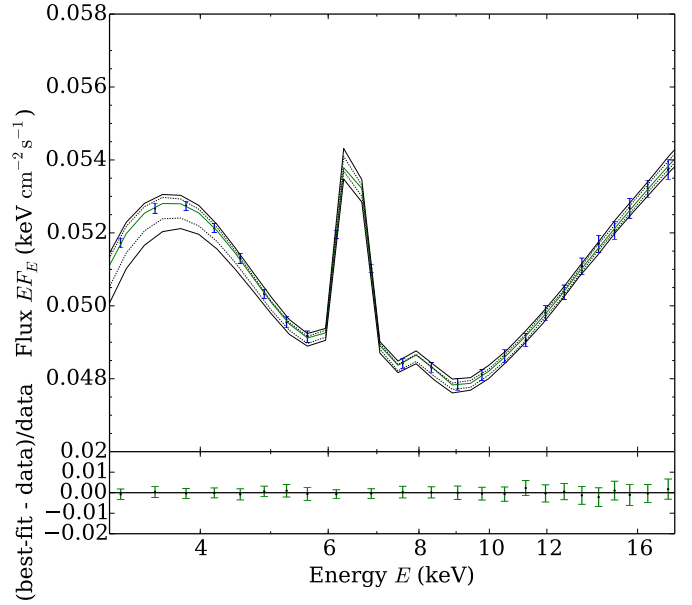
Fig. 4. Time-averaged spectrum for the synthetic data. The *lower panel* shows the residuals. The contours and the symbols are the same as in Fig. 3. The 0.5% calibration error of the detector, used in the fitting procedure, is not included to the error bars.

The contours for pulse profiles (integrated to three energy intervals for easier visualization purposes only) and for the spectrum are shown in Figs. 3 and 4. These figures demonstrate that very good fits are found both in time and energy dimension. The best-fit solution presented has χ^2/d.o.f. $= 119.1/(384 - 14) \approx 0.32$ (for 14 free parameters including the phase shift). The reduced chi-squared value is very low, because we have fitted the data using calibration error, even though there is no such error in the observed synthetic counts. Without the calibration error we would have χ^2/d.o.f. $= 375.6/370 \approx 1.02$. Also, the probability distribution around the best-fit is very tightly concentrated around the correct solution.

From Fig. 2 we see that the posterior probability distributions for each parameter are more or less close to the original input values of the synthetic data. As expected, the parameter $i - \theta$ is more difficult to constrain than $i + \theta$. But from two solutions with opposite signs, we find the correct one, since the solution with switched i and θ is, fortunately, below our lower limit for inclination, which is $40°$. However, the inclination angle i is slightly biased towards higher angles. On the other hand, θ is better constrained than i.

Most of the parameters of our model seem to be very well constrained. The relative reflection X_{refl}, the scattered fraction of photons X_{sc}, the spot temperature T', the photon spectral index Γ, and the beaming parameter h are the most tightly constrained. The angular radius of the spot ρ has a relatively broad probability distribution, however, not as broad as the distance to the star D. The distribution for the hydrogen column density N_{H} is also very broad, and it is slightly biased towards higher values. The uncertainty is, however, expected because N_{H} will only affect the lowest energy channels.

The most interesting parameters regarding this work are, of course, the mass M and the radius R_{eq}. The results in Fig. 2 show that even with using only broad priors and high level of flexibility in our model, we are able to get meaningful constraints for masses and radii. The correct value for the synthetic data is located inside the 68% credible region. The probability

Table 2. Most probable values and 68% and 95% credible limits for all 3 different simulations.

Quantity	95% lower limit	68% lower limit	Most probable value	68% upper limit	95% upper limit
			Synthetic data		
R_{eq} (km)	10.4	11.1	12.0	12.8	13.7
M ($M_\odot$)	1.36	1.44	1.51	1.64	1.75
i (deg)	52	56	61	68	75
θ (deg)	12	13	15	16	18
ρ (deg)	4.8	9.7	16	24	26
D (kpc)	1.1	1.9	3.6	5.6	5.9
h	−0.84	−0.77	−0.70	−0.66	−0.63
T' (keV)	0.81	0.83	0.85	0.87	0.90
X_{sc}	0.56	0.58	0.60	0.61	0.62
Γ	1.794	1.798	1.803	1.807	1.812
$\log \sigma_i$	0.90	0.92	1.05	1.42	1.75
N_H (10^{22} cm^{-2})	0.019	0.083	0.14	0.20	0.20
X_{refl}	0.040	0.044	0.051	0.059	0.076
			Synthetic data with more precise priors for i and θ		
R_{eq} (km)	10.6	11.0	11.9	12.4	13.3
M ($M_\odot$)	1.38	1.46	1.53	1.57	1.61
i (deg)	57.1	58.8	62.6	62.9	62.9
θ (deg)	13.8	14.6	15.2	16.5	16.9
ρ (deg)	4.7	9.2	18	23	26
D (kpc)	1.1	1.9	3.7	5.3	5.9
h	−0.75	−0.72	−0.70	−0.68	−0.67
T' (keV)	0.81	0.83	0.86	0.87	0.89
X_{sc}	0.59	0.59	0.60	0.61	0.61
Γ	1.794	1.798	1.803	1.808	1.812
$\log \sigma_i$	0.90	0.93	1.1	1.4	1.7
N_H (10^{22} cm^{-2})	0.025	0.087	0.14	0.20	0.20
X_{refl}	0.046	0.048	0.050	0.052	0.055
			Observed data of SAX J1808.4−3658		
R_{eq} (km)	4.80	5.13	5.77	6.71	10.5
M ($M_\odot$)	1.01	1.07	1.13	1.26	1.54
i (deg)	71	79	82	86	87
θ (deg)	9.5	14	17	19	21
ρ (deg)	8.4	10	12	26	36
D (kpc)	1.0	1.1	2.5	2.8	3.8
h	−0.89	−0.75	−0.69	−0.65	−0.62
T' (keV)	0.79	0.96	1.0	1.0	1.1
X_{sc}	0.56	0.60	0.62	0.63	0.65
Γ	1.798	1.803	1.807	1.812	1.816
$\log(\sigma_i)$	1.65	2.12	2.30	2.44	2.79
N_H (10^{22} cm^{-2})	0.0021	0.0027	0.0049	0.042	0.11
X_{refl}	0.027	0.052	0.087	0.12	0.20

Notes. The quantities shown in the table are equatorial radius R_{eq}, mass M, inclination i, spot co-latitude θ, spot angular size ρ, distance D, beaming parameter h, temperature of the spot T', scattered photon fraction X_{sc}, photon spectral index Γ, intrinsic scatter $\log \sigma_i$, hydrogen column density N_H, and the reflection amplitude X_{refl}.

distribution for the mass seems to slightly favour too high masses. We do not see any significant biases in the radius. All calculated credible limits and the most probable values are given in Table 2. The limits are calculated using the highest posterior probability credible intervals. For the radius we find the 68% (95%) limits and the most probable value as $R_{eq} = 12.0^{+0.8\ (1.7)}_{-0.9\ (1.6)}$ km, and for the mass $M = 1.51^{+0.13\ (0.24)}_{-0.07\ (0.15)}\ M_\odot$.

The posterior probability distribution for intrinsic scatter σ_i shows that the model describes the synthetic data well. The mean $\log \sigma_i \approx 1$ of the posterior translates to an error of 10–100 counts in each phase bin. This can be compared to the level of Poisson noise in the data, which is approximately 300

counts on average in each phase bin. Therefore, $\log \sigma_i \approx 1$ is effectively the same as $\sigma_i \approx 0$, as expected for synthetic data.

4.2. Synthetic data with additional polarimetric constraints

In this section we show what is the effect of having prior information of observer inclination i and spot co-latitude θ. This information could be obtained using the polarization measurements from, for example, the upcoming IXPE mission. The swing of the polarization angle with phase can be used to determine both i and θ (Viironen & Poutanen 2004). Our
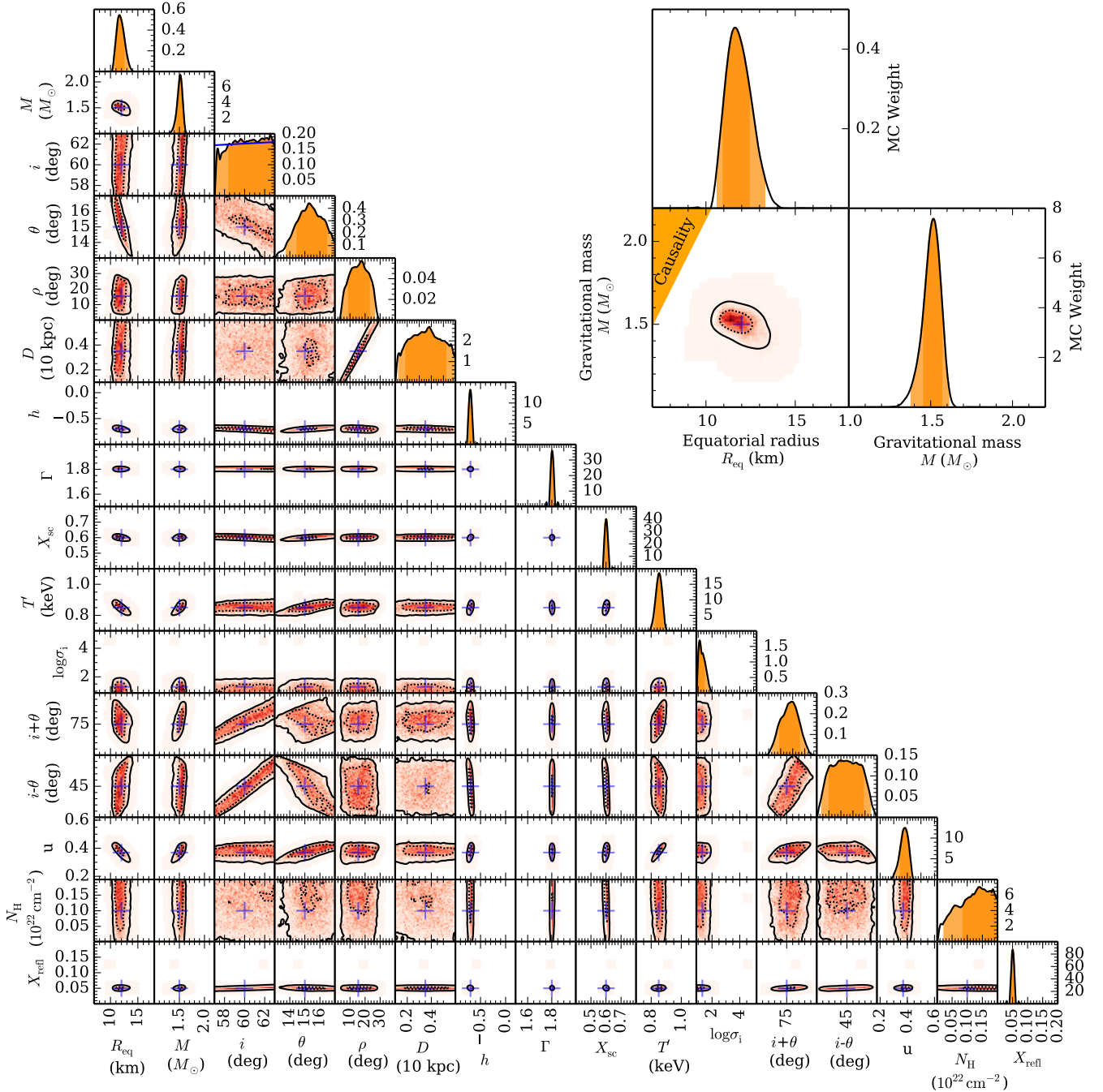
Fig. 5. Posterior probability distributions for the MCMC runs with the synthetic data, where angles i and θ are better constrained using the data possibly coming from IXPE. The colours, contours, and other symbols are the same as in Fig. 2.

synthetic data are exactly the same as in Sect. 4.1. However, we assume that the polarization data allow constraining inclination between $57°$ and $63°$ and the spot co-latitude between $13°$ and $17°$ (i.e. around the assumed parameters, see Table 1), restricting therefore the prior distributions. The boundaries of the prior distributions for other parameters are the same as in Sect. 4.1. The results are shown in Fig. 5. We emphasize here that the values used here might not be exactly correct but they do provide a rough estimate of how the possible polarization measurements might improve our inference.

From the resulting posterior distribution, we see that mass and radius, among all the other parameters, are only slightly

better constrained when compared to Fig. 2. The biases in i and $N_{\rm H}$ point to the same direction as previously. The resulting credibility limits can be again found from Table 2. In this case, we get a radius $R_{\rm eq} = 11.9^{+0.5\ (1.4)}_{-0.9\ (1.3)}$ km, and for the mass $M = 1.53^{+0.04\ (0.08)}_{-0.07\ (0.15)}\ M_{\odot}$. Compared to the previous model, additional constraints for the geometrical angles improved the accuracy clearly more in mass than in the radius. The accuracy in determination of mass, with 68% credibility level, increased from 7% to 4%, when the corresponding accuracy for radius increased from 8% to 6%. For the 95% credibility level, the accuracy in mass improved from 13% to 8% and in the radius from 14% to 12%.
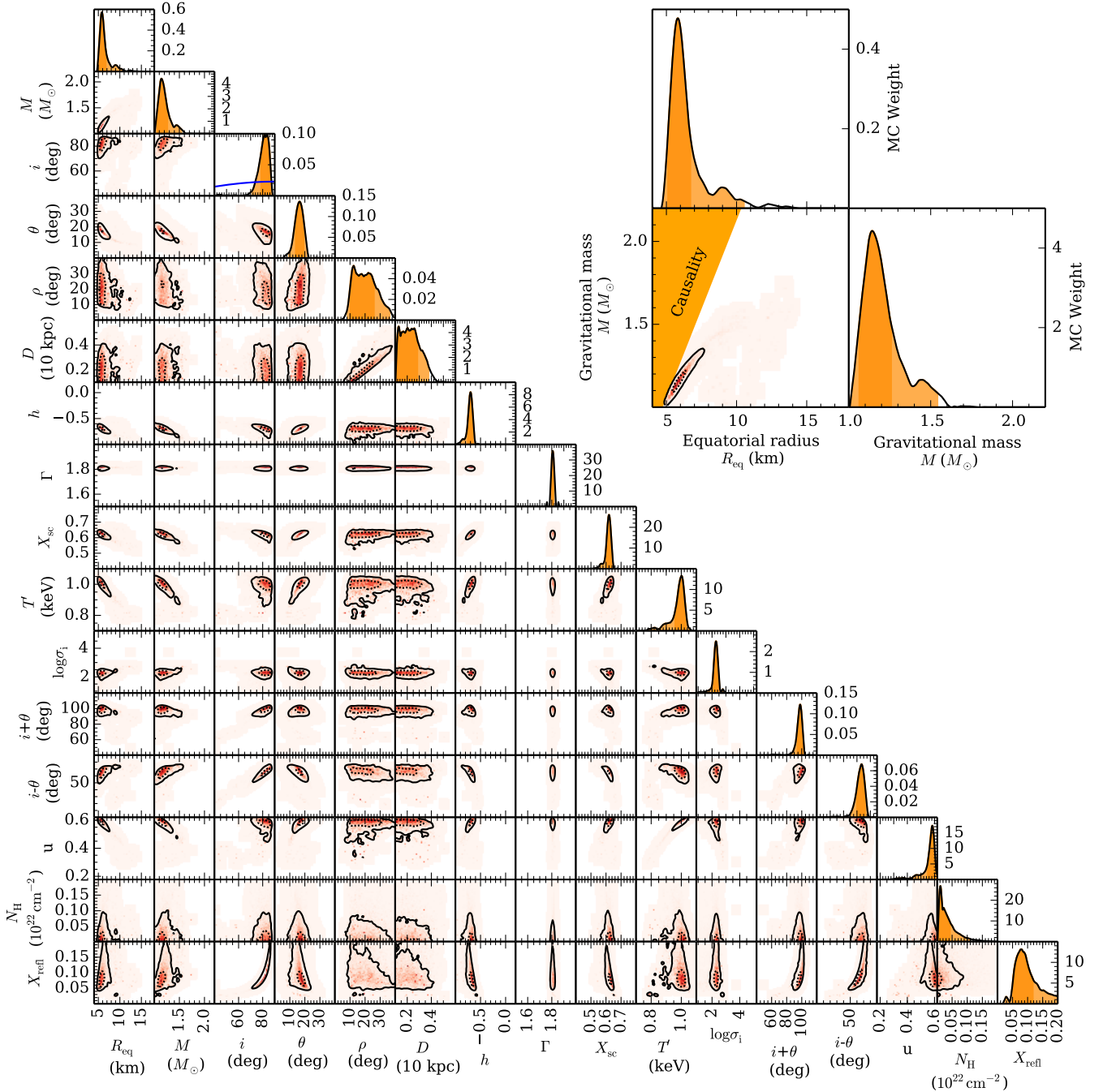
Fig. 6. Posterior probability distributions for the MCMC runs with SAX J1808.4−3658 data. The colours, contours, and other symbols are the same as in Fig. 2. The posterior shows that letting the mass be a free parameter results in unrealistic radius estimates due to possible biases in other parameters when the model does not describe data perfectly. More realistic parameter constraints with a fixed mass grid are shown in Fig. 9.

4.3. SAX J1808.4−3658 with RXTE

We now analyse the actual data of SAX J1808.4−3658 from the 1998 outburst obtained with the RXTE. We use the same model as in Sect. 4.1 to fit the data. The resulting posterior probability distributions are shown in Fig. 6. The limits of the priors are the same as in Sect. 4.1, except the lower limit for the radius is decreased to 4 km.

The contours for the pulse profiles and for the spectrum are again shown in Figs. 7 and 8. From these figures we see that the reasonably good fits are found both in time and in energy dimensions. The best-fit solution presented has $\chi^2/\mathrm{d.o.f.} = 420.6/370 \approx 1.14$ where we have taken into account the 0.5% calibration error of the PCA. Also, the probability distribution around the best-fit is quite tightly concentrated around the observed data. However, at the higher energy channels the fits are worse and they are not fully able to produce the observed smaller pulse amplitude (see the energy band 12–18 keV in Fig. 7).

From the posterior distributions we can see that many of the parameters of our model are quite well constrained. For example, the photon spectral index Γ and the scattered fraction of photons X_{sc} are reasonably well-constrained close to the values they should have according to the independently performed spectral fits with XSPEC ($\Gamma = 1.82^{+0.04}_{-0.03}$ and $X_{\mathrm{sc}} = 0.60^{+0.02}_{-0.05}$ with 90% confidence ranges). The beaming parameter $h = -0.69^{+0.04\,(0.07)}_{-0.06\,(0.20)}$ is
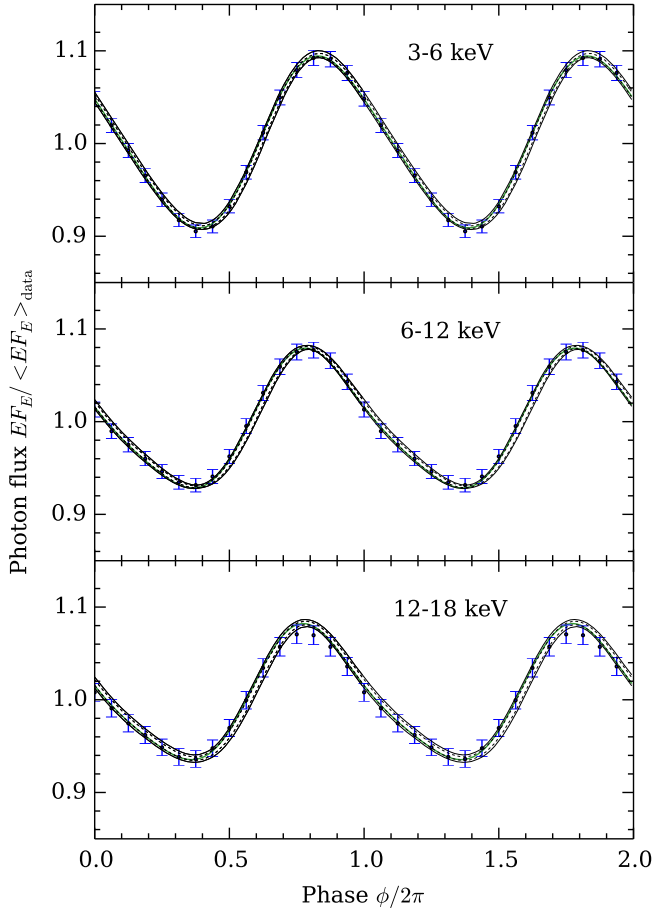
Fig. 7. Normalized pulse profiles integrated to 3 energy bins for SAX J1808.4−3658 data. The contours and the symbols are the same as in Fig. 3. The 0.5% calibration error of the detector is included to the error bars.

constrained close to the results of Poutanen & Gierliński (2003), and the reflection amplitude $X_{refl} = 0.09^{+0.03\ (0.11)}_{-0.04\ (0.06)}$ is consistent with the results of Gierliński et al. (2002). The posterior of the hydrogen column density N_H peaks at a value factor of twenty lower than expected. However, this is not critical because it just implies that the data do not contain enough information to put any stringent constraints on it. Later we confirm this by fixing the value of N_H to 0.14×10^{22} cm^{-2} given by Pinto et al. (2013) (see Sect. 5.3).

The angular size of the spot ρ has a relatively broad probability distribution, like in the case of the synthetic data, but it reaches slightly higher values (up to about 30°). For the distance we get $D = 2.5^{+0.3\ (1.3)}_{-1.4\ (1.5)}$ kpc, where the most probable distances are slightly smaller than the most probable distance of 3.5 kpc given by Galloway & Cumming (2006). The source inclination is constrained to lie above 71°.

The credibility limits for parameters, when fitting the observed data, are also shown in Table 2. In case of radius R_{eq}, our original choices to the limits of the priors would have strongly constrained the results. In this case, we get a radius $R_{eq} = 5.8^{+0.9\ (4.7)}_{-0.6\ (1.0)}$ km, and for the mass $M = 1.13^{+0.13\ (0.41)}_{-0.06\ (0.12)}\,M_{\odot}$. These results favour smaller NSs than generally accepted. However, this could occur, for example, due to any biased estimates in the other model parameters (see Sect. 5 for more detailed discussion).
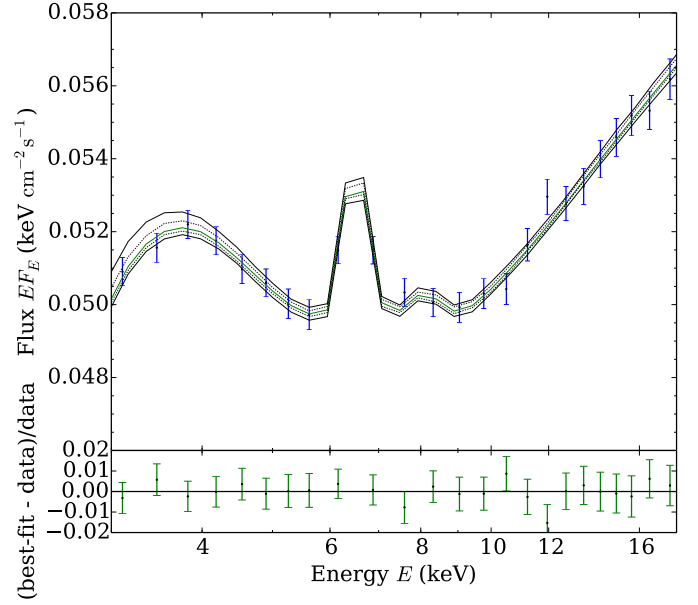
Fig. 8. Time-averaged spectrum for SAX J1808.4−3658. The contours and the symbols are the same as in Fig. 3. The 0.5% calibration error of the detector is included to the error bars.

The posterior probability distribution for intrinsic scatter with the mean $\log \sigma_i \approx 2.3$ shows that the model does not describe the data as well as the synthetic data. This corresponds an absolute error of about 200 counts or a relative error of 0.2% in each phase-energy bin. However, this error is still smaller than the Poisson noise level in the data (see Sect. 3).

4.4. SAX J1808.4−3658 with fixed M

To get more reasonable estimates for the NS radius, we have also computed the posterior distributions for all the parameters with a fixed grid of NS masses from 1.4 to 2.2 $M_{\odot}$ with the step of 0.1 $M_{\odot}$. These are then combined to obtain 2D-posterior distributions for every parameter against mass. The assumption is that each of the mass bins has an equal probability. This way it is possible to see what would be the constraints, especially for the radius, if the requirement for the free mass is relaxed and we assume it is to be known instead.

The results for this analysis are shown in Fig. 9. From the posterior distributions of the intrinsic scatter, we see that all the mass bins give more or less similar fits. Most of the parameters depend on the mass only slightly. This then indeed implies that the data do not contain enough information to convincingly constrain the radius and the mass simultaneously. However, some correlation with the mass exists especially in the radius, inclination, distance, temperature, and relative reflection amplitude. This time we can get reasonable estimates for the radius; for example, for a NS mass of 1.7 $M_{\odot}$ we get $R_{eq} = 11.9^{+0.5\ (1.0)}_{-0.4\ (1.0)}$ km for the equatorial radius. Furthermore, at this mass we have the most probable distance to the source closest to distance of 3.5 kpc given by Galloway & Cumming (2006).

5. Discussion

The results presented in Sect. 4 demonstrate that obtaining reliable radius constraints is difficult without having a priori knowledge of the NS mass. Our method works well with the synthetic
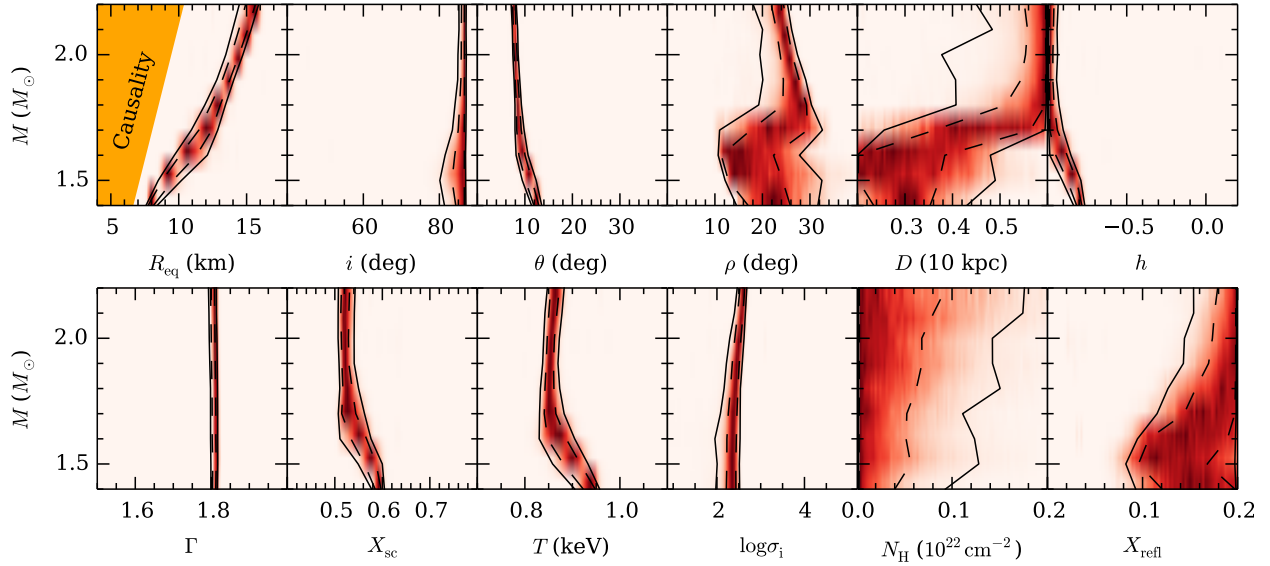
Fig. 9. Posterior probability distributions for the MCMC runs with SAX J1808.4−3658 data, where a fixed mass grid has been used. One-dimensional posterior histograms are presented at the top of each other to make approximate two-dimensional histograms with respect to mass. The colours and contours are the same as in Fig. 2.

data, but in case of SAX J1808.4−3658 our model describes the data less accurately, which might lead to larger biases in parameter estimates. Although, using a fixed grid of NS masses, we obtained more realistic results for the radius, in addition to more likely values for the distance of the star.

When using synthetic data and having additional information on the observer inclination and spot co-latitude, we get only slightly better constraints. However, it is not yet exactly known how much the upcoming polarization measurements could limit the possible values of these angles. In our case, for the synthetic data, constraining i in the interval $[57°, 63°]$ and θ in $[13°, 17°]$, instead of $[40°, 90°]$ and $[0°, 90°]$, respectively, resulted in a change of the 68% credibility interval for the radius from 11.1–12.8 to 11.0–12.4 km and the interval for the mass from 1.44–1.64 to 1.46–1.57 $M_\odot$. Similarly, the 95% credibility intervals have changed from 10.4–13.7 to 10.6–13.3 km and from 1.36–1.75 to 1.38–1.61 $M_\odot$. In addition, the most probable posterior value for the mass was found to be closer to the correct solution.

The more significant accuracy improvement on the mass, rather than on the radius, results from the stronger correlation between mass and inclination, than between radius and inclination. This is also seen in the posterior probability distributions in Figs. 2 and 5. From the latter figure, we see that the equatorial radius R_{eq} is more correlated with the spot co-latitude θ, as expected for the oblate star. However, our new limits for i and θ bind the original posterior more for i than for θ.

The constraints in mass and radius appeared to be very good in both cases where we fitted our synthetic data. However, as expected, the obtained constraints were weaker in the case of the observed data. This shows that our model does not describe the observations perfectly. On the other hand, the most probable values obtained for the radius (from 5 to 11 km), are somewhat lower than usually predicted for NSs. However, these results are quite similar than what has previously been obtained by Poutanen & Gierliński (2003) and Leahy et al. (2008) for the same source SAX J1808.4−3658. Of course, the (equatorial) radius seems to depend strongly on the combination of many other parameters, including at least the mass, spot co-latitude, magnetic inclination, spot angular size, distance, beaming, and

the spot temperature. An error or bias in any of these parameters would give incorrect results in R_{eq}. For example, a higher mass than the best-fitted 1.13 $M_\odot$, would give higher radii, if the compactness M/R_{eq} remained the same. The results with a fixed mass grid, shown in Fig. 9, also support this and give reasonable radius estimates if the mass is known a priori.

The beaming parameter h turned out be better constrained than the mass and the radius, both with synthetic and real data. However, the linear $I'_{\text{non-th}} \propto I_0(1 + h \cos \sigma')$ model for the beaming might not be the optimal one, because the largest deviations between the model and the data in the pulse profile are found from the reduced pulse amplitude at highest energies (see Fig. 7). The beaming parameter h is expected to be larger in absolute value at higher energies where the non-thermal radiation dominates, and this should flatten the pulse profiles. The current model may not do that efficiently enough.

5.1. Analytical approximations

In order to understand the relations between the parameters of our model and the observed quantities, we next present approximate analytical expressions for the pulse amplitudes. These amplitudes are expected to be the best constrained quantities from the observations. Using these expressions, we may also better interpret our results.

First we would like to get an estimate for the normalization of the light curve. This can be done by using the bolometric version of Eq. (7), where instead of $(1 - u)^{1/2}\delta^4 I'_{E'}(\sigma')$ we have $(1-u)\delta^5 I'(\sigma')$. Assuming a slowly rotating ($\delta \approx 1$) spherical star ($R = R_{eq}$), with an always visible small spot ($dS' \approx \pi(\rho R_{eq})^2$) radiating black-body flux ($I'(\sigma') \propto T'^4$), the observed normalization, or the phase-independent part of the flux, scales as

$$\mathcal{N}_0 \propto \left(\frac{\rho R_{eq}(1 + z_{eq})}{D}\right)^2 \left(\frac{T'}{1 + z_{eq}}\right)^4, \qquad (20)$$

where the first factor is proportional to the solid angle the spot occupies on the sky and the second factor is related to the bolometric black-body flux, both in the frame of an observer. Because

linear correlations are the optimal relations between different variables when using ensemble sampler, we chose to sample the solid angle the spot is seen instead of the spot angular size ρ, and the bolometric black-body flux instead of temperature T' (see Sect. 2.3.3).

We can also get a useful approximation for the bolometric pulse amplitude. This time we use the same assumptions as when deriving $\mathcal{N}_0$, but we, in addition, allow the radiation to be anisotropic and allow a larger spot size. This is done by replacing I_0 with $I_0(1 + h \cos \alpha')$ in the derivation of oscillation amplitude in the paper by Poutanen & Gierliński (2003). In case of using the following notations:

$$U \equiv (1 - r_S/R_{eq}) \sin i \sin \theta, \tag{21}$$

$$Q \equiv r_S/R_{eq} + (1 - r_S/R_{eq}) \cos i \cos \theta, \tag{22}$$

$$P \equiv \frac{r_S}{R_{eq}} \tan^2 \frac{\rho}{2}, \tag{23}$$

we get for the flux (see also Poutanen & Beloborodov 2006):

$$F(\phi) \propto Q + hQ^2 + \frac{hU^2}{2} + hQP + P$$
$$+ U(1 + 2hQ + hP) \cos \phi + h \frac{U^2}{2} \cos 2\phi. \tag{24}$$

From this we see that the pulse amplitude of fundamental increases with a positive h, and decreases with a negative one.

Using Eq. (24), we get the relative pulse amplitude

$$\mathcal{A}_1 = \frac{U(1 + 2hQ + hP)}{Q + hQ^2 + hU^2/2 + hQP + P}, \tag{25}$$

where we ignored the harmonic, which has a smaller amplitude. Without beaming ($h = 0$) and with a small spot ($P = 0$), this expression reduces to the usual $\mathcal{A}_1 = U/Q$ (Beloborodov 2002; Poutanen & Gierliński 2003). This shows the problem with distinguishing i and θ, when fitting the data without independent knowledge of any of the two parameters. As explained in the Sect. 2.3.3, we may reduce the number of (almost) indistinguishable parameters by sampling $i + \theta$ and $i - \theta$, instead of i and θ.

In addition, Eqs. (25) and (20) show that ρ is involved both in the normalization and the pulse amplitude. Therefore, even our transformed spot size parameter (see Sect. 2.3.3) that have a linear relation to $\mathcal{N}_0$, is not linear with respect to $\mathcal{A}_1$ or the observing angles. Also, the beaming parameter h is coupled with ρ. The pulse amplitude $\mathcal{A}_1$ can be also used to predict the accuracy in measurement of NS compactness when the uncertainties in the other parameters of the expression are known and their values estimated (see, e.g. Özel et al. 2016).

The ratio between the amplitudes of second and first harmonic is again obtained from Eq. (24)

$$\mathcal{A}_2 = \frac{hU/2}{1 + 2hQ + hP}. \tag{26}$$

With a small spot, this reduces to $\mathcal{A}_2 \propto h \sin i \sin \theta$ (Poutanen & Beloborodov 2006). The beaming parameter h is thus strongly connected to the observing angles also through this expression. However, the effect of fast rotation (Doppler effect) causes slightly different amplitudes, and introduces a phase shift between the fundamental and the harmonic, skewing the pulse profile and making the relations between parameters even more complex.

5.2. Uncertainties and systematic errors

The possible systematic errors can emerge either from the sampling method or from the physical modelling of the energy-resolved pulse profiles. As described in Sect. 2.3.2, sampling the parameter space introduces difficulties when the model parameters are highly degenerate and have non-linear relations with each other. This issue has been addressed by making variable transformations in order to sample parameters that are easier to constrain observationally and that would have simpler relations with each other.

The method might still suffer from a very slow convergence. The posterior distributions were slowly changing even after tens of thousands of samples from individual ensembles, or more than a thousand accepted steps on average for an individual walker. The problems with convergence of affine invariant ensemble sampler in high dimensions have been noted by Huijser et al. (2015), and the problem is expected to be worse for strongly correlated and complex models. Our solution was to use very long runs with ensemble sampler, and confirm the results with the independent nested sampling method MULTINEST. In MULTINEST we did not use any particular initial position for sampling, which also reduced the risk of the results being dependent on the initial position of an ensemble.

In the case of fitting the SAX J1808.4−3658 data, a systematic error can arise from the assumptions made in physical modelling. As already discussed, the angular distribution of the radiation might need a more realistic model. Overall, to explain all the spectral features, an accurate and physical atmosphere model of the accreting NS is required. In future work, we plan to construct the model for the accretion shock and compute self-consistently the angular distribution of the escaping radiation. This first-principle model would significantly reduce the number of free parameters and would allow to get better mass-radius constraints.

The use of model SIMPL-1, which includes only the up-scattered photons, instead of SIMPL-2 including both up- and down-scattered photons, can be one source of error. For non-relativistic thermal Comptonization some part of the photons is also scattered to lower energies. The use of SIMPL-1 could thus result in a slightly too small scattering fraction when compensating the down-scattering effects. It could also explain the unexpectedly low values of hydrogen column density N_H, since both the lower N_H and the down-scattering result in a higher number of photons in the lowest energy bands. However, the mass and radius estimates are not very sensitive to a small change in these parameters. This was also confirmed by the fit of the SAX J1808.4−3658 data, where we switched to SIMPL-2, while keeping all parameters still free. The only major difference we found was in X_{sc}, which was increased from 0.6 to 0.7.

As mentioned in Sect. 2, we ignore the radiative transfer processes of photons through an accretion column. For a typical luminosity of 5×10^{36} erg s^{-1} (Gilfanov et al. 1998), the assumed accretion efficiency of 0.15 and the circular radius of the spot of 2 km, the Thomson optical depth through the column is about a unity (see Sect. 5.3 of Poutanen & Gierliński 2003). Thus, the column may significantly affect the observed pulse profiles by making a dip at the phase corresponding to the position of the spot in front of the observer. How this would affect the results is impossible to predict without simulations and the detailed model of the accretion shock.

In our analysis we assumed a circular spot. This approximation is more likely valid when the magnetic inclination is small and the inner radius of accretion disc is much larger

than the NS radius. However, for the disc close to the star and for large magnetic inclination, the spot becomes ring-like (because the matter accretes via field lines penetrating the disc within a narrow ring) or even crescent-like (Romanova et al. 2004; Kulkarni & Romanova 2005). A change in the spot shape may affect the pulse profile (see e.g. Poutanen et al. 2009; Kajava et al. 2011). Although, based on the simplicity of the pulse profiles, we expect this effect to be rather small.

As mentioned in Sect. 2, we calculate the pulse profiles using only one spot. This is justified, because we use the observations of SAX J1808.4−3658, where the accretion disc is expected to hide the second spot (Ibragimov & Poutanen 2009). The lack of a second maximum in the pulse profile is an evidence for this. However, there could still be a small contribution from the second spot. In future work and when modelling also other sources, the second spot, along with a model of light propagation to the observer through the truncated accretion disc, should be included.

5.3. Additional observational constraints

As mentioned in Sects. 2.3.3 and 4.3, there are independent constraints, for instance on the observer inclination i and on the interstellar hydrogen column density $N_{\rm H}$ for SAX J1808.4−3658. Using X-ray spectroscopy with *XMM-Newton* Pinto et al. (2013) found out that $N_{\rm H} = 1.40^{+0.03}_{-0.03} \times 10^{21}\,{\rm cm}^{-2}$ with 1σ errors. On the other hand, the inclination and the mass of the NS are connected and constrained via the mass function $f_{\rm M} = M\sin^3 i/(1 + q)^2$, where q is the binary mass ratio. Using optical observations of the companion star, during the 2008 outburst, Elebert et al. (2009) constrained the mass ratio of the system to $q = 0.044^{+0.005}_{-0.004}$ and the mass function to $f_{\rm M} = 0.44^{+0.16}_{-0.13}\,M_\odot$. Cornelisse et al. (2009) used the same dataset, but ended up to less realistic mass estimates. Later, Wang et al. (2013) got similar estimates as Elebert et al. (2009), using optical observations of the companion during quiescence.

We have studied how these constraints affect our results by making an additional analysis to the SAX J1808.4−3658 data with a fixed value of $N_{\rm H} = 0.14 \times 10^{22}\,{\rm cm}^{-2}$ and using a Gaussian 2-dimensional prior probability distribution $M\sin^3 i = 0.48^{+0.15}_{-0.15}\,M_\odot$ for the mass and the inclination. In case of the model where mass is a free parameter, using the new 2-dimensional prior we get smaller and more realistic inclinations, around 50°. However, they are still not small enough to make the prior prefer larger masses (and therefore radii) of the NS. The smaller inclination mainly affects the spot co-latitude θ, which now increased above 20° in order to still fit the observed pulse amplitude. Also, the relative reflection from the accretion disc $X_{\rm refl}$ becomes smaller and is now constrained below 0.05. When a fixed mass grid with the new priors was used, we were able to get inclinations significantly below 50°, as expected in case of higher masses. However, the resulting posteriors for radii did not change significantly.

To get the same mass function as Wang et al. (2013) or Elebert et al. (2009), we would need i smaller than 50° for a NS with a mass $1.4\,M_\odot$. Since the observed small pulse amplitude requires large enough separation between i and θ, this inclination is not very easy to produce without fixing the mass (assuming that we still exclude the smallest i and hence the case of almost equatorial spot). Of course, as seen in Sect. 5.1 and in Eq. (25), the relative pulse amplitude is also determined by the beaming parameter h and the spot size ρ. This might suggest that our spot size is too small, for example due to the possibly diffuse boundary of the spot, or that the beaming model needs to

be improved. In any case, the model is not perfect, since we have also neglected, for instance, absorption of light by the accretion column.

6. Summary

We have presented a method to constrain masses and radii of NSs using both spectral and X-ray timing information from AMPs. We have modelled the NS spectra and pulse profiles in case of two different synthetic data. In the first case, we assumed that the observer inclination and spot inclination angles are unknown, while in the second case we assumed that there exist constraints on both angles, for example, from the X-ray polarization measurements. The results showed that our method works in both cases, but the second case led to slightly tighter constraints on mass and radius. Accuracy improved from 13% to 8% in determination of the mass and from 14% to 12% in determination of the radius (with 95% credibilities).

We have also fitted our model to the already existing RXTE data on SAX J1808.4−3658. The obtained constraints for the mass and the radius are in good agreement with previous studies of this object, favouring relatively small NSs. Because the smallest masses and radii allowed by our model are not realistic for any modern equation of state, we have computed the probability distributions using also a fixed mass grid. The results showed, that the radius can be well constrained if the NS mass is known a priori. For example, a NS mass of $1.7\,M_\odot$ corresponds to the equatorial radius of $R_{\rm eq} = 11.9^{+0.5\ (1.0)}_{-0.4\ (1.0)}$ km. Getting reasonable constraints for the NS radius is thus possible using even the already existing data from RXTE.

References

AlGendy, M., & Morsink, S. M. 2014, ApJ, 791, 78
Arnaud, K. A. 1996, in Astronomical Data Analysis Software and Systems V, eds. G. H. Jacoby, & J. Barnes (San Francisco: ASP), ASP Conf. Ser., 101, 17
Bauböck, M., Psaltis, D., Özel, F., & Johannsen, T. 2012, ApJ, 753, 175
Beloborodov, A. M. 2002, ApJ, 566, L85
Braje, T. M., Romani, R. W., & Rauch, K. P. 2000, ApJ, 531, 447
Cackett, E. M., Altamirano, D., Patruno, A., et al. 2009, ApJ, 694, L21
Cadeau, C., Morsink, S. M., Leahy, D., & Campbell, S. S. 2007, ApJ, 654, 458
Cash, W. 1979, ApJ, 228, 939
Chakrabarty, D., & Morgan, E. H. 1998, Nature, 394, 346
Chen, K., & Shaham, J. 1989, ApJ, 339, 279
Cornelisse, R., D'Avanzo, P., Muñoz-Darias, T., et al. 2009, A&A, 495, L1
Deloye, C. J., Heinke, C. O., Taam, R. E., & Jonker, P. G. 2008, MNRAS, 391, 1619
Elebert, P., Reynolds, M. T., Callanan, P. J., et al. 2009, MNRAS, 395, 884
Falanga, M. 2008, in Cool Discs, Hot Flows: The Varying Faces of Accreting Compact Objects, ed. M. Axelsson, AIP Conf. Ser., 1054, 157
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Galloway, D. K., & Cumming, A. 2006, ApJ, 652, 559
García, J., Dauser, T., Reynolds, C. S., et al. 2013, ApJ, 768, 146
Gierliński, M., & Poutanen, J. 2005, MNRAS, 359, 1261
Gierliński, M., Done, C., & Barret, D. 2002, MNRAS, 331, 141
Gilfanov, M., Revnivtsev, M., Sunyaev, R., & Churazov, E. 1998, A&A, 338, L83
Goodman, J., & Weare, J. 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Grinstead, C. M., & Snell, J. L. 1997, Introduction to Probability, 2nd edn. (Providence, RI: American Mathematical Society)
Hartman, J. M., Patruno, A., Chakrabarty, D., et al. 2008, ApJ, 675, 1468

Hou, F., Goodman, J., Hogg, D. W., Weare, J., & Schwab, C. 2012, ApJ, 745, 198
Huijser, D., Goodman, J., & Brewer, B. J. 2015, ArXiv e-prints [arXiv:1509.02230]
Ibragimov, A., & Poutanen, J. 2009, MNRAS, 400, 492
Jahoda, K., Markwardt, C. B., Radeva, Y., et al. 2006, ApJS, 163, 401
Kajava, J. J. E., Ibragimov, A., Annala, M., Patruno, A., & Poutanen, J. 2011, MNRAS, 417, 1454
Kolehmainen, M., Done, C., & Díaz Trigo, M. 2011, MNRAS, 416, 311
Kulkarni, A. K., & Romanova, M. M. 2005, ApJ, 633, 349
Lattimer, J. M. 2012, ARNPS, 62, 485
Lattimer, J. M., & Prakash, M. 2004, Science, 304, 536
Leahy, D. A., Morsink, S. M., & Cadeau, C. 2008, ApJ, 672, 1119
Lindblom, L. 1992, ApJ, 398, 569
Lo, K. H., Miller, M. C., Bhattacharyya, S., & Lamb, F. K. 2013, ApJ, 776, 19
Lo, K. H., Miller, M. C., Bhattacharyya, S., & Lamb, F. K. 2018, ApJ, 854, 187
Magdziarz, P., & Zdziarski, A. A. 1995, MNRAS, 273, 837
Miller, M. C., & Lamb, F. K. 1998, ApJ, 499, L37
Miller, M. C., & Lamb, F. K. 2015, ApJ, 808, 31
Misner, C. W., Thorne, K. S., & Wheeler, J. A. 1973, Gravitation (San Francisco: W. H. Freeman and Co.)
Morsink, S. M., & Leahy, D. A. 2011, ApJ, 726, 56
Morsink, S. M., Leahy, D. A., Cadeau, C., & Braga, J. 2007, ApJ, 663, 1244
Nättilä, J., & Pihajoki, P. 2018, A&A, 615, A50
Özel, F., Psaltis, D., Arzoumanian, Z., Morsink, S., & Bauböck, M. 2016, ApJ, 832, 92
Page, D. 1995, ApJ, 442, 273
Pechenick, K. R., Ftaclas, C., & Cohen, J. M. 1983, ApJ, 274, 846
Pinto, C., Kaastra, J. S., Costantini, E., & de Vries, C. 2013, A&A, 551, A25
Poutanen, J. 2008, in A Decade of Accreting Millisecond X-ray Pulsars, eds. R. Wijnands, D. Altamirano, P. Soleri, et al., AIP Conf. Ser., 1068, 77
Poutanen, J., & Beloborodov, A. M. 2006, MNRAS, 373, 836
Poutanen, J., & Gierliński, M. 2003, MNRAS, 343, 1301
Poutanen, J., Ibragimov, A., & Annala, M. 2009, ApJ, 706, L129
Psaltis, D., Özel, F., & Chakrabarty, D. 2014, ApJ, 787, 136
Romanova, M. M., Ustyugova, G. V., Koldoba, A. V., & Lovelace, R. V. E. 2004, ApJ, 610, 920
Steiner, J. F., Narayan, R., McClintock, J. E., & Ebisawa, K. 2009, PASP, 121, 1279
Stevens, A. L., Fiege, J. D., Leahy, D. A., & Morsink, S. M. 2016, ApJ, 833, 244
Sunyaev, R. A., & Titarchuk, L. G. 1980, A&A, 86, 121
Viironen, K., & Poutanen, J. 2004, A&A, 426, 985
Wang, Z., Breton, R. P., Heinke, C. O., Deloye, C. J., & Zhong, J. 2013, ApJ, 765, 151
Watts, A. L., Andersson, N., Chakrabarty, D., et al. 2016, Rev. Mod. Phys., 88, 021001
Weinberg, N., Miller, M. C., & Lamb, D. Q. 2001, ApJ, 546, 1098
Weisskopf, M. C., Ramsey, B., O'Dell, S., et al. 2016, in Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray, Proc. SPIE, 9905, 990517
Wijnands, R., & van der Klis, M. 1998, Nature, 394, 344
Zhang, S. N., Feroci, M., Santangelo, A., et al. 2016, in Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray, Proc. SPIE, 9905, 99051Q

Appendix A: Numerical solutions for light bending and time delay integrals

Numerical calculations using directly integral (1) are not possible. Making substitution $y = R/r$, the integral becomes

$$\psi_{\rm p}(R, \alpha) = \sin \alpha \int_0^1 {\rm d}y \left[1 - u - y^2 \sin^2\alpha \, (1 - yu) \right]^{-1/2} . \quad (A.1)$$

Applying quadrature formulae to this form of the integral leads to dramatic errors when $\sin \alpha \sim 1$, because of the divergent behaviour of the integrand at $y = 1$. However, simple variable substitution $x = \sqrt{1 - y}$ removes the divergence:

$$\psi_{\rm p}(R, \alpha) = 2 \frac{\sin \alpha}{\sqrt{1 - u}} \int_0^1 \left(\cos^2\alpha + x^2 q \right)^{-1/2} x \, {\rm d}x, \quad (A.2)$$

where

$$q = \left[2 - x^2 - u(1 - x^2)^2/(1 - u) \right] \sin^2\alpha. \quad (A.3)$$

Notice that now for $\cos \alpha = 0$, the integrand does not diverge at $x = 0$ and the integral can be computed without problems using, for example, Simpson quadratures.

Also, direct calculations using expression (9) are not possible. Making substitution $y = R/r$, the integral becomes

$$\Delta t_{\rm p} = \frac{R}{c} \int_0^1 \frac{{\rm d}y}{y^2(1 - yu)} \left\{ \left[1 - \sin^2\alpha \, y^2(1 - yu)/(1 - u) \right]^{-1/2} - 1 \right\}. \quad (A.4)$$

This form has seemingly two divergences at $y = 0$ and for $\sin \alpha = 1$ at $y = 1$. The first divergence is removed by multiplying and dividing by the expression similar to that in the curly brackets but with the +1 in the end. This gives

$$\Delta t_{\rm p} = \frac{R}{c} \frac{\sin^2\alpha}{1 - u} \int_0^1 \frac{{\rm d}y}{\sqrt{1 - w} \left(1 + \sqrt{1 - w} \right)}, \quad (A.5)$$

where

$$w = \frac{y^2(1 - yu)}{1 - u} \sin^2 \alpha. \quad (A.6)$$

The divergence at $y = 1$ (for $\sin \alpha = 1$) is of the $1/\sqrt{1 - y}$-type and therefore is trivially removed by variable substitution $x = \sqrt{1 - y}$. This transforms the integral to

$$\Delta t_{\rm p} = 2 \frac{R}{c} \frac{\sin^2\alpha}{1 - u} \int_0^1 \frac{x \, {\rm d}x}{\sqrt{x^2 q + \cos^2\alpha} \left(1 + \sqrt{x^2 q + \cos^2\alpha} \right)}, \quad (A.7)$$

where q is given by Eq. (A.3). Now for $\cos \alpha = 0$, the integrand does not diverge at all, but becomes $1/(qx + \sqrt{q})$ and the integral can be computed using e.g. Simpson quadratures.