An Investigation of Systematic Effects from Background Priors on PSR J0740+6620 Radius Estimates using Synthetic NICER and XMM-Newton Data

Isiah M. Holt [ID],[1,2] M. Coleman Miller [ID],[1,3] Alexander J. Dittmann [ID],[4] and Frederick K. Lamb [ID][5,6]

[1]Department of Astronomy, University of Maryland, College Park, MD 20742-2421, USA
[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[3]Joint Space-Science Institute, University of Maryland, College Park, MD 20742-2421, USA
[4]Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, USA*
[5]Illinois Center for Advanced Studies of the Universe and Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801-3080, USA
[6]Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801-3074, USA

ABSTRACT

Accurate and precise measurements of neutron star radii provide invaluable information about the cold, dense matter in neutron star cores. Analyses of synthetic X-ray pulse waveform data similar to the data obtained from non-accreting neutron stars using the Neutron star Interior Composition Explorer (NICER) have indicated that mass and radius estimates made using such data are robust against some systematic errors that may be made when modeling these data, such as errors in the assumed pattern of the thermal X-ray emission from the surface of these stars. A potentially important but so far unexplored source of systematic error is misparameterization of unmodulated background components, which can bias the inferred radius, particularly when data from different telescopes are used in the analysis. In this study, we investigate the effects of the background model on radius estimates by jointly analyzing synthetic NICER and XMM-Newton data, using the $\sim 2.1\ M_\odot$ pulsar PSR J0740+6620 as a prototypical example. Our analysis shows that even if the background assumed in the model underestimates the actual background by a factor of more than five, the resulting shift of the radius posterior from the true value of the radius corresponds to only $\sim 1\sigma$. In all the cases we examined, the Bayesian evidence for the correct background model is greater than for the incorrect background model. These results add to the evidence that analyses of NICER-like data provide accurate measurements of neutron star radii when the statistical sampling is thorough and the model fits the data well.

1. INTRODUCTION

Neutron star cores have a combination of high density, low temperature, and large neutron-proton asymmetry that cannot be reproduced in laboratories, and their properties cannot currently be predicted precisely using quantum chromodynamics (due to the fermion sign problem; see, e.g., Troyer & Wiese 2005; Hsu & Reeb 2010). In contrast, observations of neutron stars provide constraints on the properties of cold dense matter at densities up to a few times nuclear saturation density ($n_s \approx 0.16$ fm^{-3}), the highest density that occurs in

nuclei when there is no confining pressure. Astrophysical observations of neutron stars are therefore crucial for understanding the nature of matter at supranuclear densities.

Most of the macroscopic properties of a neutron star, such as its mass, radius, and tidal deformability, depend only on the equation of state (EOS) of the matter in its interior, i.e., the pressure as a function of the density, temperature, and other properties. In all but newborn neutron stars, the thermal energies of the particles in this matter are much less than their Fermi energies. As a result, when this matter is in microscopic equilibrium, which is usually assumed, the pressure depends only on the energy density ϵ, that is, $P = P(\epsilon)$, where P is the pressure.

Email: imholt@umd.edu
* NASA Einstein Fellow

Over the last $\sim$ 15 years, numerous astronomical observations have helped constrain the EOS of cold, high-density matter. These include measurements of high masses for a few neutron stars such as PSR J1614$-$2230 (M $= 1.97 \pm 0.04$ M$_\odot$, Demorest et al. 2010), PSR J0348+0432 (M $= 1.806 \pm .037$ M$_\odot$, Saffer et al. 2024), PSR J0740+6620 (M $= 2.08 \pm 0.07$ M$_\odot$, Fonseca et al. 2021), and PSR J0952$-$0607 (M $= 2.35 \pm 0.17$ M$_\odot$, Romani et al. 2022); observational constraints on the tidal deformability of neutron stars inferred from observations of the gravitational wave event GW170817 (see Abbott et al. 2016 and De et al. 2018); and estimates of the radii of several neutron stars using various approaches to analyzing thermal X-ray emission from their surfaces.

Early attempts to measure the radii of neutron stars in quiescent low-mass X-ray binary systems (Guillot et al. 2011, 2013; Lattimer & Steiner 2014; Kim et al. 2021; Bogdanov et al. 2016; Marino et al. 2018) and/or thermonuclear X-ray burst sources (see e.g., van Paradijs 1979; Majczyna & Madej 2006; Steiner et al. 2010; Güver et al. 2010; Lo et al. 2013; Özel et al. 2016) used integrated flux and spectral observations. However, radius estimates made using these methods are subject to significant systematic errors: though the models used in these analyses fit the data well (with some exceptions; see Steiner et al. 2010; Kajava et al. 2014), making different assumptions about the composition of a neutron star's atmosphere can change its inferred radius by $\sim 50\%$, far more than the apparent uncertainty of the measurement (Miller 2013). The usual assumption that the neutron star surface radiates uniformly can also introduce large systematic errors. By relaxing this assumption Nättilä et al. (2017) obtained radii $\sim 10\% - 20\%$ larger than the radii they inferred using a model that assumed the surface temperature is uniform.

These difficulties motivated searches for alternative methods to determine the masses and radii of neutron stars, which led to the proposal to determine the masses and radii of certain rotating neutron stars by accurately measuring the periodic pulse waveforms of the thermal X-ray emission from their surfaces and then fitting theoretical models to them (Lo et al. 2013; Miller & Lamb 2015, 2016). Analyses of synthetic pulse waveform data demonstrated that this method could be used to estimate the masses and radii of these stars with interesting precisions, and furthermore, that the resulting mass and radius estimates are not significantly biased by certain types of systematic errors, such as using incorrect shapes and temperature distributions for the X-ray emitting regions on the stellar surface, provided that the fit to the pulse pulse waveform data is statistically acceptable (again see Lo et al. 2013; Miller & Lamb 2015, 2016).

The Neutron Star Interior Composition Explorer (NICER; Gendreau et al. 2012), with its ability to measure the X-ray pulse waveforms of some non-accreting neutron stars with high temporal and spectral resolution, made it possible to use the newly proposed pulse waveform fitting method to determine the masses and radii of these neutron stars with unprecedented precision. The rms uncertainties in the photon arrival times measured using NICER are less than 100 nanoseconds, much less than the minimum observed rotational periods of neutron stars. NICER made it possible to accurately measure the X-ray flux and spectrum of these rotating neutron stars as a function of the star's rotational phase rather than merely their time-averaged values.

As noted above, the first studies of the measurements of neutron star masses and radii that could be achieved using data of this quality (Lo et al. 2013; Miller & Lamb 2015, 2016) demonstrated that they are not significantly biased even if incorrect shapes and temperature distributions are assumed for the X-ray emitting regions on the stellar surface (which we hereafter call "spots"), provided the fit of the model pulse waveform to the observed pulse waveform data is statistically acceptable. In this work we expand on these initial studies by exploring the effect on measurements of the radius of PSR J0740+6620 caused by errors in the assumed X-ray background.

The waveform NICER observes from a rotating neutron star with X-ray emitting hot spots on its surface has modulated and unmodulated components. The hot spots generally produce a waveform with both types of components. In addition to counts from the hot spots, NICER also registers counts from other sources. The count rates from these other sources—such as nearby bright X-ray sources, the diffuse X-ray background from active galactic nuclei (AGN), charged particles interacting with the NICER detectors, and optical loading caused by light from the Sun—are not modulated at the neutron star's rotational frequency. The count rates produced by these other sources can therefore be treated collectively as a single, unmodulated background when analyzing the pulsar's waveform.

Correctly determining the relative sizes of these contributions is important for the following reason. When the observed pulse waveform is analyzed, more flux not from the hot spots implies less flux from the spots. Since the modulation amplitude in the NICER data is fixed, less flux from the spots implies that the fractional modulation of the flux contributed by the spots is greater. This usually requires a larger stellar radius, if the other

parameters of the pulsar remain unchanged (Miller 2016; Miller et al. 2021). Hence, incorrectly estimating the flux that is not coming from the hot spots could in principle bias the radius estimate. Including in the pulse waveform analysis any information about any component of the unmodulated background should improve the accuracy of the radius estimate, as long as the information and the method used to include it are reliable (see, e.g., Wolff et al. 2021).

PSR J0740+6620 is of special interest because it is a high-mass pulsar ($M \approx 2.1$ $M_\odot$; see Fonseca et al. 2021). The matter in its core is therefore substantially denser than the matter in the cores of ~ 1.4 $M_\odot$ neutron stars and hence measurement of its radius provides information about the EOS of neutron star matter at densities higher than do measurements of the radii of ~ 1.4 $M_\odot$ neutron stars. Unfortunately, the count rate measured by NICER when it is pointed at PSR J0740+6620 is dominated by sources other than the hot spots on its surface. Correctly estimating the unmodulated flux contributed by these other sources is therefore particularly important when analyzing the NICER observations of PSR J0740+6620.

In the case of PSR J0740+6620, important information about the X-ray flux that is not coming from the star, and therefore about the unmodulated background, is provided by the X-ray Multi-Mirror (XMM-Newton) blank-sky observations made over the lifetime of the mission (Carter & Read 2007). As discussed in Bogdanov et al. (2021); Miller et al. (2021); Riley et al. (2021); Salmi et al. (2022, 2024); Dittmann et al. (2024), knowing the XMM-Newton blank-sky background allows a more accurate estimate of the total X-ray flux from the pulsar, which helps to more accurately determine its radius. For other pulsars, this information could come from observations of the sky near the pulsar in question.

We used the XMM-Newton "blank-sky" background to estimate the background present in the XMM-Newton data on PSR J0740+6620 because the relatively short durations of the observations of PSR J0740+6620 by the XMM-Newton instruments (roughly 18.0 ks for MOS1, 18.7 ks for MOS2, and 6.8 ks for pn) yielded too few background counts to accurately estimate the background. XMM-Newton's low background count rate, high angular resolution, and wide field of view make it a good instrument for making blank-sky observations. The XMM-Newton "blank-sky" background therefore provides a reliable estimate of the blank-sky component of the background in XMM-Newton observations of PSR J0740+6620.

The XMM-Newton "blank-sky" background has been constructed using data obtained by observing apparently empty regions of the sky over ~ 20 years. These observations totaled ~ 1.5 Ms for the two European Photon Imaging Camera Metal-Oxide-Silicon detectors (MOS1 and MOS2) and ~ 450 ks for the pn CCD (pn) (Carter & Read 2007). These mission-averaged data files provide a standard estimate of the detector and diffuse X-ray backgrounds.

In this work we filtered the event files from the XMM-Newton blank-sky observations identically to the way we filtered the event files from the XMM-Newton observations of PSR J0740+6620. We then extracted the background counts and rescaled the number of counts to match the PSR J0740+6620 exposure times. We explored the implications of assuming different unmodulated backgrounds in the NICER and XMM-Newton data and different ways of including this background when estimating the radius of PSR J0740+6620. To do this, we generated synthetic NICER and XMM-Newton data by sampling the pulse waveform given by the model that Miller et al. (2021) found best fits the NICER data on PSR J0740+6620 and assumed two different Poisson-sampled unmodulated backgrounds for the synthetic XMM-Newton data.

In one analysis, we assumed that the only background in the XMM-Newton synthetic data is the blank-sky background measured by XMM-Newton. This corresponds to the assumption made in the analysis of the PSR J0740+6620 data by Miller et al. (2021). In a second analysis, we assumed a much larger background approximately equal to the sum of the pulsar's flux and the blank-sky background fluxes from PSR J0740+6620 measured using XMM-Newton. This larger background could be due, for example, to cosmic rays, X-ray bright stars, or AGN. We then analyzed both these synthetic data sets, first using a model that assumed an unmodulated XMM-Newton background approximately equal to the blank-sky background measured by XMM-Newton and then using a model that included the possibility of an extra unmodulated XMM-Newton background.

We found that even if the incorrect background model is assumed in the analysis, the resulting shift of the radius posterior from the true value of the radius corresponds to only $\sim 1\sigma$. In all the cases we examined, the Bayesian evidence for the model that assumes the correct XMM-Newton background is greater than the evidence for the models that assume an incorrect XMM-Newton background. This demonstrates that the Bayesian evidence can effectively discriminate between background models in otherwise ambiguous circumstances

In Section 2 we describe how we generated the synthetic NICER and XMM-Newton data we analyzed, including the emission from the neutron star surface and the background. Section 3 describes how we modeled the resulting synthetic data and evaluated our best-fit models and Section 4 presents our detailed results. Our conclusions are summarized in Section 5.

2. CONSTRUCTING THE SYNTHETIC DATA

Our goal in this paper is to better understand the effect on the accuracy of the inferred neutron star radius of errors made in modeling the XMM-Newton background. Because the inferred radius is obtained by jointly analyzing NICER and XMM-Newton data, we need to generate synthetic data for all the instruments that were used to collect these data.

In constructing the two synthetic PSR J0740+6620 data sets that we analyzed in this work, we assumed the same stellar rotational frequency, NICER exposure time, and XMM-Newton exposure times assumed by Miller et al. (2021). We also assumed that the periodic amplitude modulation in the synthetic data is produced entirely by two hot spots on the stellar surface with the properties found in the analysis by Miller et al. (2021). We note, however, that emission from a pulsar's magnetosphere and shock waves around it could, in principle, also contribute to the modulation of soft X-rays from the pulsar. Although most pulsars do not exhibit such emission, the emission from the brightest, non-accreting X-ray millisecond pulsar, PSR J0437−4715, does have a significant, although faint, modulated power-law component (Guillot et al. 2016). Hence the assumption that the periodically modulated soft X-ray flux is produced only by emission from hot spots on the stellar surface is not universally valid.

In Section 2.1 we describe how we constructed the synthetic NICER data, in Section 2.2 we explain how we generated the synthetic XMM-Newton data with a background equal to the blank-sky background, and in Section 2.3 we describe how we constructed the synthetic XMM-Newton data with a background substantially larger that the blank-sky background.

2.1. Synthetic NICER pulse waveform data

We created two synthetic NICER data sets independently for the two XMM-Newton backgrounds we considered. Each synthetic NICER data set was generated by Poisson sampling the pulse waveform model that Miller et al. (2021) found best fit the NICER and XMM data on PSR J0740+6620, when considered jointly (see Section 3.4 of Miller et al. 2019 for additional details). This model includes emission from two hot spots and an unmodulated background. The parameters in this model and their best-fit values are listed in Table 1.

Like the actual NICER data, these two synthetic NICER data sets were two-dimensional, containing both energy and pulse-phase information. We synthesized data for NICER pulse invariant (PI) channels 30 through 123 inclusive (providing 94 energy channels) and 32 equally-spaced rotational phases, for a total of 3008 phase-channel bins.

2.2. Data Set 1: Synthetic XMM-Newton data includes only the blank-sky background

One XMM-Newton synthetic data set we investigated assumed that the only XMM-Newton counts that did not come from the hot spots are "blank-sky" counts. The synthetic XMM-Newton background data for this case were generated by Poisson-sampling our best fits to the MOS1, MOS2, and pn blank-sky data. For the MOS1 and MOS2 detectors we used PI channels 20 through 99 inclusive while for the pn detector we used PI channels 57 through 299 inclusive, as in Miller et al. (2021). As noted previously, the XMM-Newton data do not have sufficient time resolution for these counts to be assigned to the finely spaced phase bins we used to analyze the NICER data. Hence no phases were assigned to the XMM-Newton counts. Therefore a model in which all the XMM-Newton counts are attributed to an unmodulated background with none attributed to the hot spots could, in principle, fit these XMM-Newton data perfectly. This emphasizes the importance of determining the actual background.

2.3. Data Set 2: Synthetic XMM-Newton data includes additional background

We also investigated a second XMM-Newton synthetic data set that assumed the number of XMM-Newton counts that did not come from the hot spots is substantially greater than the "blank-sky" counts. We constructed this second XMM-Newton synthetic data set by adding additional background counts to the synthetic blank-sky counts in the MOS1, MOS2, and pn detectors. We generated these additional counts by applying a 1st-order Savitzky-Golay filter (Savitzky & Golay 1964) to the *total* counts collected by these detectors when XMM-Newton was pointed at PSR J0740+6620. We used filters with window lengths of 130 channels for the pn data and 50 channels for the MOS1/MOS2 data. We then performed a Poisson draw from this fit and added the resulting counts, channel by channel, to the synthetic data.

Comparing the resulting number of counts in each energy channel of the three XMM-Newton detectors with

Table 1. Primary Parameters in the Pulse Waveform Model

Parameter	Definition	Assumed Prior	Best-Fit Value
$c^2 R_e/(GM)$	Inverse stellar compactness	$3.2 - 8.0$	4.52
M	Gravitational mass	$\exp[-(M - 2.08 M_\odot)^2/2(0.09 M_\odot)^2]$	2.067
$\theta_{c,1}$	Spot 1 inclination of center spot	$0.0 - \pi$ rad	0.834
$\Delta\theta_1$	Spot 1 radius	$0.0 - 3.14$ rad	0.087
$kT_{\mathrm{eff},1}$	Spot 1 effective temperature	$0.011 - 0.5$ keV	0.098
$\Delta\phi_2$	Spot 1 and 2 longitudinal offset	$0.0 - 1.0$ cycles	0.577
$\theta_{c,2}$	Spot 2 inclination of center spot	$0.0 - \pi$ rad	1.223
$\Delta\theta_2$	Spot 2 radius	$0.001 - 3.14$ rad	0.066
$kT_{\mathrm{eff},2}$	Spot 2 effective temperature	$0.011 - 0.5$ keV	0.103
θ_{obs}	Observer inclination	$1.44 - 1.62$ rad	1.561
N_H	Column density	$0.0 - 2.0 \times 10^{20}$ cm^{-2}	0.006
D	Distance	$\exp[-(D - 1.136\mathrm{kpc})^2/2(0.20\mathrm{kpc})^2], \mathrm{D} \geq 1.136\mathrm{kpc}$	1.151
		$\exp[-(D - 1.136\mathrm{kpc})^2/2(0.18\mathrm{kpc})^2], \mathrm{D} \leq 1.136\mathrm{kpc}$	

NOTE—This table details the pulse waveform model that Miller et al. (2021) fit to the NICER and XMM-Newton data on PSR J0740+6620, the priors on the parameters (with the exception that Miller et al. 2021 used flat priors of 0.0–3 rad for $\Delta\theta_1$ and $\Delta\theta_2$), and the values of the parameters in this model that they found gave the best joint fit to the NICER and XMM-Newton data. The 12 parameters listed here are the primary parameters in the pulse waveform model used in Miller et al. (2021) and in the present work; R_e is the equatorial circumferential radius. We generated synthetic pulse waveform data by sampling this best-fit waveform, assuming the same 346.532 Hz rotational frequency assumed by Miller et al. (2021). The priors listed here are the same as those we used when fitting this model to the two synthetic data sets we analyzed. See the text for further details.

the number of counts in each channel shown in Figure 8 of Miller et al. (2021) for their model of XMM-Newton's "blank-sky" counts, the average number of XMM-Newton background counts is a factor of ~ 5 greater. As a result, the total number of counts in this synthetic XMM-Newton background data is, on average, twice the total number of counts XMM-Newton detected when it was pointed in the direction of PSR J0740+6620. We chose this synthetic XMM-Newton background data to explore the effect on the estimate of the radius of PSR J0740+6620 if the XMM-Newton background is substantially greater than is normally estimated.

3. MODELING THE SYNTHETIC DATA

Following Miller et al. (2021) and Dittmann et al. (2024), we modeled the pulse waveform in our synthetic data using two uniform-temperature circular spots and assuming a non-magnetic hydrogen atmosphere. For the analyses reported here we used the atmospheric tables for emission by partially ionized hydrogen from Ho & Heinke (2009) (see Badnell et al. 2005 for the relevant opacity tables). This pulse waveform model has 12 primary parameters, which specify the mass and compactness of the neutron star, the radii, temperatures, and locations of the two spots, the phase offset between them, the observer's inclination, and the column density and

distance to the pulsar. Table 1 lists these 12 parameters, the ranges of the priors Miller et al. (2021) used when they fit this pulse waveform model to the actual NICER and XMM-Newton data, and their best-fit values of these parameters. As Table 1 notes, when we fit this model to the synthetic data considered in the present study we assumed slightly wider priors for $\Delta\theta_1$ and $\Delta\theta_2$.

Although the XMM-Newton data are unmodulated, these data are important in constraining the phase-independent background in each energy channel. This constraint helps disentangle astrophysical backgrounds from non-modulated surface emission, yielding more precise constraints on the hot spot properties and stellar compactness. In particular, knowing the background level from XMM-Newton allows the model to attribute the correct fraction of the observed counts in each channel to the modulated emission from the hot spots, tightening the posteriors on their temperatures and sizes and thereby tightening the constraints on mass-radius combinations.

In fitting this model to the actual NICER and XMM-Newton data on PSR J0740+6620, Miller et al. (2021) labeled the two spots '1' and '2' and allowed them to overlap without restriction. To avoid ambiguity when a given pixel is covered by both spots, they assumed that

such a pixel emits with the effective temperature of spot 1. They also assumed a stellar rotational frequency of 346.532 Hz, a NICER exposure time of 1, 602, 683.7 s, an XMM-pn exposure time of 6, 808.74 s, an XMM-MOS1 exposure time of 17, 959.5 s, and an XMM-MOS2 exposure time of 18, 680.7 s (Miller et al. 2021; Riley et al. 2021). We followed this same procedure when fitting the two waveform models considered in the present work to the two synthetic data sets we generated.

The two full pulse waveform models we fit to synthetic Data Set 1 and Data Set 2 each contain the same 12 primary waveform parameters (see Table 1), but differ in the assumptions we made about the phase-independent background in the XMM-Newton data. In one model, we assumed that the only non-spot contribution is the XMM-Newton blank-sky background, with the background in each energy channel modeled as a Poisson-distributed variable whose prior is given by the long-term blank-sky measurements. In the other model, we allowed for additional unmodulated background in each energy channel, in addition to the blank-sky counts. In both models the NICER background is treated identically: we marginalized over a phase-independent term in each energy channel using a Gaussian fit to the likelihood profile (see Section 3.4 of Miller et al. 2019 for additional details). Figure 1 shows the smoothed synthetic spectra we used to define the priors on the additional XMM-Newton background for the two synthetic data sets.

For a given model and data set, we adjusted the 12 primary waveform parameters and the background parameters to maximize the total log likelihood. The background parameters directly affect the predicted counts in each energy channel and thus the log-likelihood contributions from each instrument. For example, increasing an assumed background reduces the contribution from the hot spot, altering the modulation amplitude and shifting the maximum of the likelihood to a larger radius.

3.1. *Calculation of the XMM-Newton log likelihood when the only background is assumed to be from the blank sky*

For our first analysis we assumed that all the counts that do not come from the two hot spots are background counts and that these counts are fully described by the counts collected by XMM-Newton during its blank-sky observations. This was the assumption made by Miller et al. (2021) and Dittmann et al. (2024) when they analyzed the NICER and XMM-Newton data on PSR J0740+6620.

(a)

(b)
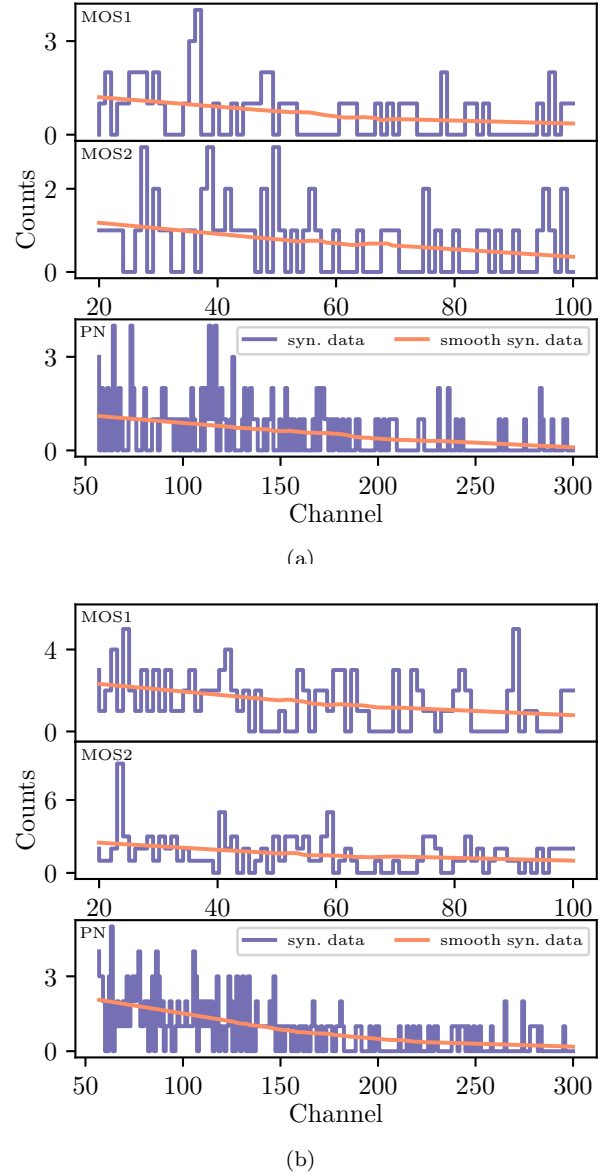
Figure 1. The data we used to define the priors on the backgrounds when fitting synthetic Data Set 1 and synthetic Data Set 2. In panel (a) the purple histograms show the counts that were generated by Poisson sampling the sum of the modeled spot emission and the XMM-Newton blank-sky background, while the orange curves show the same data smoothed using a Savitzky-Golay filter. When we analyzed synthetic Data Set 1, which includes only the XMM-Newton blank-sky background, using a model that allows for an additional unmodulated background, we used the smoothed data to define the prior for that additional background. In panel (b) the purple histograms show the counts generated by Poisson sampling the sum of the modeled spot emission, the XMM-Newton blank-sky background, and the additional unmodulated background we considered. Again, the orange curves show the smoothed data. In this case we used the smoothed data to define the priors on the additional backgrounds when we analyzed synthetic Data Set 2 using a model that includes the XMM-Newton blank-sky background and an additional unmodulated background.

XMM-Newton has made blank-sky observations for more than two decades, but the total number of counts it has collected in each energy channel of the XMM-pn, XMM-MOS1, and XMM-MOS2 detectors during these observations is small, so it is necessary to use Poisson statistics when treating them. The probability that the long-term average number of blank-sky counts over a time T_{bk} is between m_{bk} and $m_{\mathrm{bk}} + dm_{\mathrm{bk}}$ is

$$\mathrm{prob}(m_{\mathrm{bk}})dm_{\mathrm{bk}} = \frac{m_{\mathrm{bk}}^{N_{\mathrm{bk}}}}{N_{\mathrm{bk}}!}e^{-m_{\mathrm{bk}}}dm_{\mathrm{bk}} \ . \tag{1}$$

For our synthetic Data Set 1, we adopted this probability as the prior for the blank-sky counts in this energy channel and used it to estimate the number of blank-sky counts that, in addition to the counts from the hot spots, produced the total number of counts in that energy channel. Then, for an observation lasting a time T_{obs}, the expected number of blank-sky background counts is $b_{\mathrm{bk}} = m_{\mathrm{bk}}(T_{\mathrm{obs}}/T_{\mathrm{bk}})$.

If the number of counts from the hot spots in the specified detector channel is s, then the total number of counts predicted by the model (including both the counts contributed by the spots and the counts contributed by this candidate background) is $s + b_{\mathrm{bk}}$. For a detector channel with N_{obs} counts observed over time T_{obs}, the Poisson likelihood of the data, given these hot spot and background models, is

$$\mathcal{L}(N_{\mathrm{obs}}|s + b_{\mathrm{bk}}) = \frac{(s + b_{\mathrm{bk}})^{N_{\mathrm{obs}}}}{N_{\mathrm{obs}}!}e^{-(s+b_{\mathrm{bk}})} \ . \tag{2}$$

The posterior probability is then proportional to the product of the probability of m_{bk} and the likelihood:

$$P(s + b_{\mathrm{bk}}|N_{\mathrm{obs}}, N_{\mathrm{bk}}) \ \propto \ \frac{m_{\mathrm{bk}}^{N_{\mathrm{bk}}}}{N_{\mathrm{bk}}!}e^{-m_{\mathrm{bk}}}$$
$$\times \frac{(s+b_{\mathrm{bk}})^{N_{\mathrm{obs}}}}{N_{\mathrm{obs}}!}e^{-(s+b_{\mathrm{bk}})} \ . \tag{3}$$

The likelihood of the data given the model is then given by marginalizing $P(s + b_{\mathrm{bk}}|N_{\mathrm{obs}}, N_{\mathrm{bk}})$ over m_{bk}.

3.2. Calculation of the XMM-Newton log likelihood when we allow for the possibility of additional background

For our second analysis, we incorporated the possibility of additional unmodulated background in the XMM-Newton data on PSR J0740+6620. In addition to the XMM-Newton blank-sky background, we included an extra component that could account for sources such as backgrounds from nearby sources or intrabinary shocks in the pulsar system. The maximum allowed number of counts b_{max} from this additional background, in any given energy channel for each instrument, was

taken to be the smoothed count rate predicted from the pointed XMM-Newton observation in that channel. This smoothed spectrum was obtained by applying a Savitzky-Golay filter (Savitzky & Golay 1964) to the total counts in the synthetic data. This procedure produces an estimate of the mean counts per channel that serves as a physically motivated upper bound for the additional background, ensuring that the inferred background does not exceed the average counts measured in the observation.

We assumed that the prior on the additional background b_{other} in each channel was uniform between zero and the corresponding smoothed value b_{max}:

$$\mathrm{prob}(b_{\mathrm{other}})db_{\mathrm{other}} = \frac{1}{b_{\mathrm{max}}}db_{\mathrm{other}} \ . \tag{4}$$

This choice is independent of the specific hot-spot waveform model and is therefore acceptable for model comparison. When a model includes this additional background, the predicted counts per detector energy channel are given by the Poisson-sampled $s + b_{\mathrm{bk}} + b_{\mathrm{other}}$. The posterior probability for the model parameters in this case is then

$$P(s + b_{\mathrm{bk}} + b_{\mathrm{other}}|N_{\mathrm{obs}}, N_{\mathrm{bk}}) \propto$$
$$\frac{m_{\mathrm{bk}}^{N_{\mathrm{bk}}}}{N_{\mathrm{bk}}!}e^{-m_{\mathrm{bk}}}\frac{1}{b_{\mathrm{max}}}e^{-(s+b_{\mathrm{bk}}+b_{\mathrm{other}})}$$
$$\times \frac{(s + b_{\mathrm{bk}} + b_{\mathrm{other}})^{N_{\mathrm{obs}}}}{N_{\mathrm{obs}}!}. \tag{5}$$

We then obtained the likelihood of the data given the model by marginalizing $P(s+b_{\mathrm{bk}}+b_{\mathrm{other}}|N_{\mathrm{obs}}, N_{\mathrm{bk}})$ over both m_{bk} and b_{other}. This approach provides a way to test how the presence of such a background, particularly if it is not included in the model, can influence the inferred radius.

3.3. Fitting our two different pulse waveform models to Data Sets 1 and 2

We fit each of our two different full pulse waveform models to our two different full synthetic data sets (Data Set 1 and Data Set 2). Specifically, for each full pulse waveform model and synthetic data set, we calculated the log Poisson likelihood of the model by summing the log-likelihoods for each of the four data sets. The total log-likelihood is then

$$\log \mathcal{L}_{\mathrm{tot}} = \log \mathcal{L}_{\mathrm{NICER}} + \log \mathcal{L}_{\mathrm{XMM_{PN}}}$$
$$+ \log \mathcal{L}_{\mathrm{XMM_{MOS1}}} + \log \mathcal{L}_{\mathrm{XMM_{MOS2}}} \ . \tag{6}$$

Equation (6) assumes that the four different data sets are independent, which is reasonable given that NICER and XMM-Newton are separate instruments and that

the XMM-MOS1, XMM-MOS2, and XMM-pn detectors operate independently, using distinct hardware and readout systems. For each data set, we computed the log likelihood by summing the log Poisson likelihood across all the energy channels and (for NICER) all the rotational phase bins (see Section 3.4 in Miller et al. 2021).

We determined the pulse waveform models that best fit the joint NICER and XMM-Newton synthetic data by adjusting the 12 primary pulse waveform parameters listed in Table 1 to maximize the log posterior. In addition to the primary pulse waveform parameters, our fitting procedure introduced additional nuisance parameters.

For the part of the fit that was to the XMM-Newton synthetic data, these nuisance parameters describe the background, and are discussed in Section 3.1 (when the only non-spot contribution is the blank sky background) and Section 3.2 (when we allow for the possibility of additional background).

For the part of the fit that was to the NICER data, the rotational phase introduces an additional nuisance parameter and changes our approach to marginalizing over the background. This additional nuisance parameter is the overall rotational phase of the pulse waveform. To determine this phase, we computed a trial pulse waveform and then determined the log-likelihood of the data given the model for different assumed rotational phases. We then fitted a Gaussian to the resulting likelihood distribution and marginalized over the phase.

For the NICER background, we did not impose prior constraints on the unmodulated backgrounds present in the NICER data, although models for these backgrounds exist (e.g., the 3C50 model of Remillard et al. 2022). Instead, we included a phase-independent background independently for each energy channel, fitted a Gaussian to the likelihood as a function of the added backgrounds, and then marginalized.

Our models involve a substantial number of parameters, making posterior exploration in the resulting high-dimensional spaces a significant computational challenge. To address this challenge, we adopted a hybrid sampling approach.

We used the nested sampling algorithm MultiNest (Feroz & Hobson 2008; Feroz et al. 2009; Buchner 2016) but only as a starting point for our posterior sampling. Previous studies (e.g., Miller et al. 2019, 2021; Ih & Kempton 2021; Lemos et al. 2023; Dittmann 2024; Dittmann et al. 2024) have found that MultiNest often produces posterior credible regions that are too narrow, therefore underestimating the uncertainties in the parameters, particularly when the number of live points

$N_{\rm live}$ is insufficient and/or the sampling efficiency parameter η is too large (see, e.g., Dittmann et al. 2024).

As an example, Salmi et al. (2024) reported a headline radius of $12.49^{+1.28}_{-0.88}$ km using $N_{\rm live} = 4 \times 10^4$ and $\eta = 0.01$ when they required the radius to be less than 16 km, but when they used a smaller sampling efficiency ($\eta = 10^{-4}$), the radius uncertainty broadened to $12.55^{+1.37}_{-0.92}$ km, although they did not carry out analyses to the point of achieving converged and robust uncertainties; Dittmann et al. (2024), who used the same nested sampling + MCMC strategy as our current investigation, found $12.76^{+1.49}_{-1.02}$ km when they imposed the same ad hoc radius limits as Salmi+, though the full range was $12.92^{+2.09}_{-1.13}$ km.

For our initial analysis of the pulse waveforms studied here, we used $\eta = 0.01$ and $N_{\rm live} = 4096$, following Dittmann et al. (2024). While these choices for the values of these parameters made our initial exploration of the parameter space efficient, they did not necessarily lead to robust posteriors. As an example, for our initial analysis we carried out three independent MultiNest runs. The resulting best-fit values of only 18 of the 36 parameters (6 of 12 per realization) fell within the $\pm 1\sigma$ credible region predicted by the MultiNest-only analyses, versus the ≈ 24 expected given that the $\pm 1\sigma$ band contains 68% of the probability. This is consistent with the previously observed tendency for MultiNest to underestimate posterior uncertainties.

To refine our results, we initialized parallel tempered MCMC (PT-emcee; Foreman-Mackey et al. 2013) using walkers from the completed MultiNest run, thereby leveraging its output to accelerate convergence of the PT-emcee analysis. PT-emcee ensures detailed balance, allowing the MCMC chains to asymptotically sample the true posterior distribution. We assessed convergence by monitoring the 1st, 16th, 50th, 84th, and 99th percentiles of each posterior.

Our settings included 1024 walkers per temperature rung across four rungs, with temperatures equal to 1 (which thus produces an unmodified log likelihood), 2, 4, and 8 (totaling $1024 \times 4 = 4096$ walkers). This approach allowed us to combine the rapid exploration of the parameter space that MultiNest allows with the convergence properties of PT-emcee. We verified robustness by repeating the combined MN+PT-emcee analysis three times independently, each yielding consistent posteriors.

To assess our approach, we compared the width of the $\pm 1\sigma$ credible regions for the inferred radius, which we denote as $\Delta R_{\rm e}$, given by the MultiNest-only and the MultiNest+PT-emcee analyses for the four cases we investigated (see Table 3 for the definitions of these cases).

Table 2. Widths of the Radius Distributions from MultiNest-only and MultiNest+PT-emcee Analyses

Case	$\Delta R_{e,MN}$ (km)	$\Delta R_{e,MN+PT}$ (km)	% Broadened
1	3.23	4.59	42.11
2	2.85	3.18	11.58
3	4.41	5.89	33.56
4	4.71	5.98	26.96

NOTE—Comparison of the $\pm 1\sigma$ credible region width (ΔR_e), defined as the width of the $\pm 1\sigma$ credible region for the radius, between MultiNest-only ($\Delta R_{e,MN}$) and MultiNest+PT-emcee ($\Delta R_{e,MN+PT}$) analyses. See Table 3 for the definitions of the cases listed here.

Table 2 summarizes the results. In every case the initial, MultiNest-only estimate of ΔR_e was appreciably broadened by subsequent use of PT-emcee, reinforcing previous findings that MultiNest tends to give erroneously narrow credible intervals when N_{live} is too small and/or η is too large.

3.4. *Evaluating the models*

To evaluate our models, we used chi-square tests to assess the absolute goodness of the fit and the Bayesian evidence to compare models.

We computed the chi-squared statistic χ^2 using the model variance form originally proposed by Pearson (1900), namely

$$\chi^2 = \sum_i \frac{(m_i - d_i)^2}{m_i}. \tag{7}$$

Here m_i is the expected number of counts in bin i predicted by the model and d_i is the observed number of counts in bin i. This formulation differs from the data variance form, which replaces m_i in the denominator with d_i. Unlike the data variance form, the model variance form ensures that when the model is correct and d_i follows a Poisson distribution with mean m_i, the expected value of χ^2 equals the number of bins, regardless of m_i, although when m_i is small the χ^2 distribution is broader than when m_i is large.

Our assessments of the implications of the value of χ^2 apply only to the fits of the pulse waveform model to the NICER data—the XMM-Newton data have too few counts for an assessment of χ^2 to be meaningful. We compute χ^2 for the NICER part of the fit using both the full phase-channel data and the bolometric data; because χ^2 is a nonlinear measure, these provide different perspectives on the goodness of fit. For example, a fit which is consistently slightly high at a given phase will have much larger bolometric χ^2 at that phase than the sum of all of the channel-phase χ^2 values at that phase.

Interpretation of the χ^2 statistic requires evaluation of the probability—a lower probability implies a poorer fit. We took a probability less than 10^{-3} as indicating that the model may be inadequate. Conversely, a probability in the tens of percent range does not confirm correctness but suggests consistency between the model and the data. Intermediate probabilities require careful consideration in the context of the analysis.

The Bayesian evidence (Z) serves a different purpose by directly comparing models while accounting for model goodness of fit and complexity. Its logarithmic form, $\log Z$, is defined as

$$\log Z = \log \int \mathcal{L}(\theta)\pi(\theta)\,d\theta, \tag{8}$$

where $\mathcal{L}(\theta)$ is the likelihood function, $\pi(\theta)$ is the normalized prior probability distribution of the parameters θ, and the integral spans the parameter space. Since direct evaluation of this integral is often intractable, numerical techniques such as nested sampling and parallel tempering can help obtain useful results. Nested sampling estimates $\log Z$ by iteratively replacing low-likelihood samples with new ones at increasing likelihood thresholds, effectively transforming the integral into a weighted sum. Parallel tempering runs Markov chains at different temperatures, facilitating efficient exploration of the parameter space and improving the accuracy of the integration. A larger value of $\log Z$ indicates stronger support for a given model relative to alternatives (e.g., Sivia & Skilling 2006).

Because the Bayesian evidence inherently involves comparing models it complements χ^2, which measures how well a single model fits the data on its own, without directly evaluating alternative models. However, the value of the evidence can be sensitive to the prior volumes adopted for all the model parameters, since it is computed by integrating over the full prior space. Assuming broader priors for the parameters increases the volume of the prior space, which can penalize the evidence, even if the model fits the data well. In the present analysis, the parameters that affect the volume of the prior space include the background parameters discussed earlier. To ensure that model comparisons based on $\log Z$ are robust, it is important to check that any preference for one model over another persists across a range of reasonable choices for the priors of their parameters.

Table 3. Summary of Equatorial Radius Posteriors from MultiNest+PT-emcee Analyses

Case	Data	Model	Input R_e	$-2\sigma\ R_e$	$-1\sigma\ R_e$	Median R_e	$+1\sigma\ R_e$	$+2\sigma\ R_e$
1	extra[a]	extra[b]	13.823	11.29	12.7	14.61	17.29	20.76
2	extra[a]	standard[d]	13.823	9.67	10.38	11.5	13.56	17.25
3	standard[c]	extra[b]	13.823	12.44	14.69	17.3	20.58	23.51
4	standard[c]	standard[d]	13.823	10.84	12.32	14.7	18.3	22.3

[a]See Section 2.3 [b]See Section 3.2 [c]See Section 2.2 [d]See Section 3.1

NOTE—All equatorial radius values are given in km. We generated one synthetic data set for each background model (Data Sets 1 and 2) using the best-fit model of the PSR J0740+6620 pulse waveform described in the text and the specified background model. The pulse waveform model assumes two circular spots in a non-magnetic, partially ionized hydrogen atmosphere. In cases 1 and 4, the synthetic data were analyzed assuming the same background model that was used to generate the synthetic data, whereas in Cases 2 and 3 the synthetic data were analyzed assuming an incorrect background model (see text for details). Inferred equatorial circumferential radii (in km) are those given by the fits listed in columns 2 and 3. Definitions: Data Set 1 = XMM-Newton blank-sky background only; Data Set 2 = XMM-Newton blank-sky background plus an additional unmodulated background. Background Model 1 = blank-sky background only; Background Model 2 = blank-sky background plus an additional unmodulated background. Cases 1–4 correspond to the four possible combinations of Data Set and Background Model. The radius assumed in the model used to generate the synthetic pulse waveform data is within one standard deviation of the median inferred radius for Cases 1 and 4, and just outside one standard deviation for Cases 2 and 3.

4. RESULTS

We now present the results of our analyses of the four cases listed in Table 3. For each case we list the assumption about the background that was made when the synthetic data was generated and the assumption about the background that was made in the model that was used to analyze the synthetic data. Each case explores how the relationship between the background in the synthetic data and the background assumed in the fitted model affects the estimated radius of the star. These results improve our understanding of when and how a mismatch between the actual (in this case synthetic) background and the background assumed when analyzing the data affects the inferred stellar radius.

Table 3 reports the inferred median values and $\pm1\sigma$ and $\pm2\sigma$ credible regions for the equatorial circumferential radii given by these analyses. Table 4 shows the quality of the fits given by the phase-channel and bolometric χ^2 values and compares the Bayesian evidence for these fits computed using MultiNest. To verify the values of the Bayesian evidence computed using MultiNest we independently computed the Bayesian evidence using pocoMC (Karamanis et al. 2022). In all cases the logarithmic evidence computed using MultiNest and pocoMC differed by less than 0.7, confirming the consistency of the evidence values computed using these two codes. In the following subsections we describe our key results for each of the four cases listed in Table 3.

Table 4. Quality of Fits to the Synthetic Data

Case	$(\chi^2/\mathrm{dof})_{PE}$ (Prob.)	$(\chi^2/\mathrm{dof})_B$ (Prob.)	$\Delta\mathrm{logZ}$
1	2978.47/2901 (0.155)	29.85/25 (0.23)	-
2	2980.24/2901 (0.149)	29.01/25 (0.26)	-8.59 ± 0.11
3	2920.28/2901 (0.397)	39.24/25 (0.035)	-5.76 ± 0.10
4	2916.59/2901 (0.416)	37.41/25 (0.053)	-

NOTE—χ^2 values for the phase- and energy-resolved (PE) and bolometric (B) pulse waveforms and the probabilities of the best-fit models for each case. Fitting the models that were used to generate the synthetic data (cases 1 and 4) yielded higher log evidences than fitting the models with additional backgrounds and are therefore preferred over the models with additional backgrounds (cases 2 and 3). This suggests that a converged analysis can be a useful tool to identify the more promising models. All the fits to the bolometric and phase-energy channel pulse waveform appear acceptable. Note that these χ^2 assessments apply only to the fits of the pulse waveform model to the NICER data. None of the fits, even the ones that assumed incorrect backgrounds, provide a clear mismatch with the data.

4.1. The synthetic data include and the fitted model assumes an additional background

In Case 1, the synthetic XMM-Newton data were generated with a background larger than the standard XMM-Newton blank-sky background, and the fitted model explicitly allowed for this additional background. As expected for a model matched to the data generation
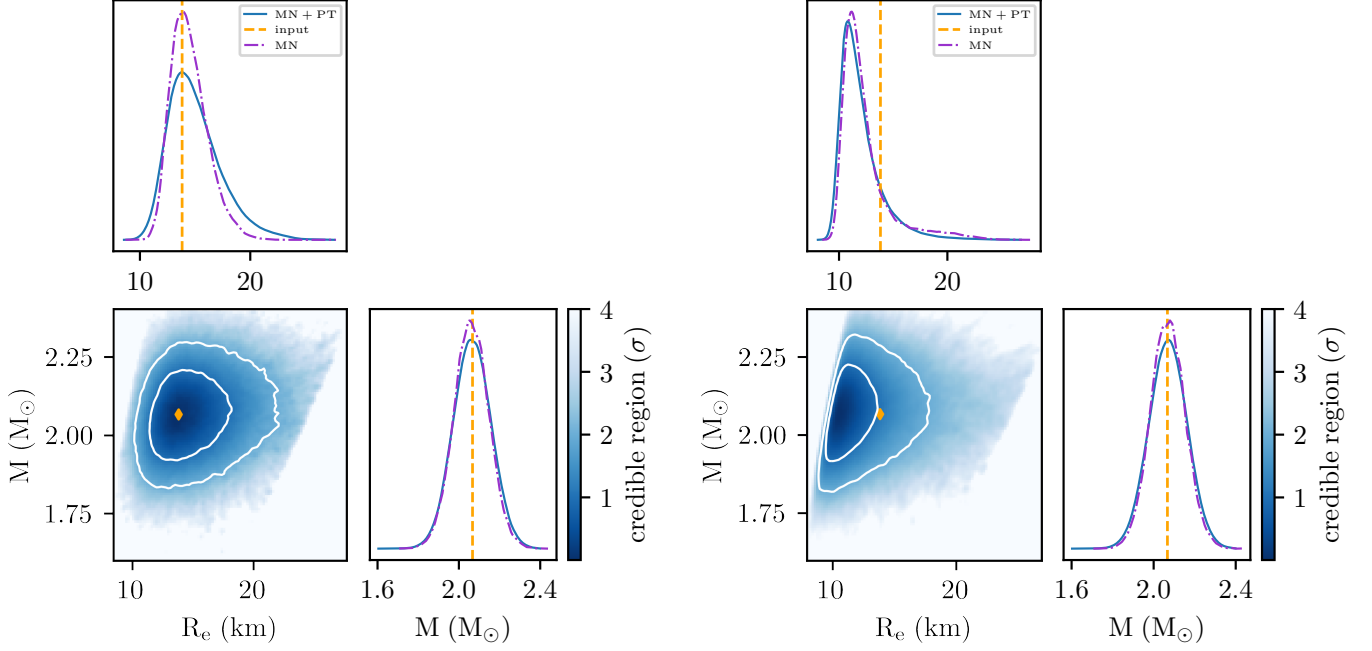
(a) Case 1: The synthetic data were generated with a background approximately five times the XMM-Newton blank-sky background, and the fitted model allowed for an additional background beyond the blank-sky component. The true radius lies at the 36th percentile of the PT-emcee radius posterior, near the peak of the distribution.

(b) Case 2: The synthetic data were generated with a background approximately five times the XMM-Newton blank-sky background, but the fitted model assumed only the blank-sky background. The radius posterior peaks at a value smaller than the true radius used to generate the synthetic data, which lies at the 86th percentile of the posterior distribution.

Figure 2. The mass and radius posterior probability distributions from our combined MultiNest and PT-emcee analysis for Case 1 (a) and Case 2 (b). In each case, the orange diamond in the 2D plot and the orange dashed lines in the 1D plots mark the mass and radius values assumed in generating the synthetic data. Solid blue and dot–dashed purple curves show the PT-emcee and MultiNest posteriors, respectively. As discussed in the text, the radius distribution broadens from the initial MultiNest result to the converged PT-emcee distribution, while the mass posterior remains dominated by the prior.

process, the inferred neutron star radius is accurate (see Table 4 and Figure 2a), and both the phase-channel and bolometric χ^2 values indicate satisfactory fits.

The phase-channel χ^2 quantifies the agreement between the model and the data across the full two-dimensional grid of rotational phase and energy, whereas the bolometric χ^2 provides a complementary measure by integrating over energy and comparing the phase-resolved pulse profiles. For the phase-energy resolved fit, the best-fitting two-spot model yielded $\chi^2 = 2978.47$ for 3008 phase–energy bins. Accounting for 12 primary parameters, 94 background-related nuisance parameters, and one overall phase parameter leaves $3008 - 94 - 12 - 1 = 2901$ degrees of freedom, giving $\chi^2/\mathrm{dof} = 2978.47/2901$. If the model is correct, the probability of obtaining a χ^2 at least this large is $p = 0.155$, which exceeds our threshold for considering a fit acceptable (Section 3.4).

The bolometric fit compares the 32-phase pulse profile from the best-fitting two-spot model to the corresponding synthetic bolometric NICER waveform. This yields

$\chi^2 = 29.85$. Following Section 4.3 of Dittmann et al. (2024) we adopt an effective number of ~ 5 primary parameters influencing the bolometric waveform. Including the overall rotational phase and a single unmodulated background term gives approximately $32 - 5 - 2 = 25$ degrees of freedom, resulting in $\chi^2/\mathrm{dof} = 29.85/25$. The corresponding probability of obtaining such a value or larger is $p = 0.23$, which is again acceptable.

As shown in Figure 2a, the inferred mass and radius credible regions encompass the input values used to generate the synthetic data, with the radius posteriors (Table 3) showing no apparent bias. The full posterior distribution for this case is displayed in Figure 4 in the Appendix.

4.2. The synthetic data include an additional background but the fitted model does not

In Case 2, the synthetic XMM-Newton data contained additional background beyond the blank-sky background, but the fitted model assumed only the blank-sky background. The meanings of the line and
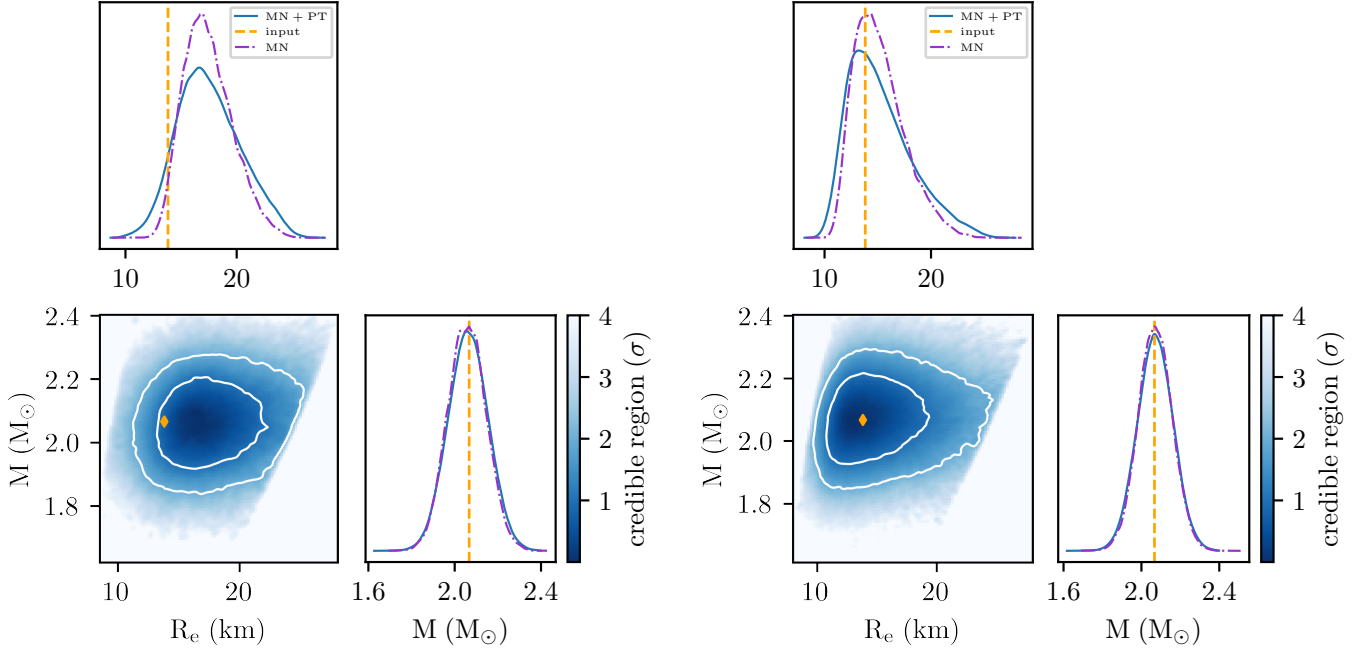
(a) Case 3: In this case the synthetic data contained only the XMM-Newton blank-sky background, whereas the fitted model allowed for an additional unmodulated background. This assumption leads to a posterior distribution that peaks at a radius larger than the true input value, with the true radius lying at the 8th percentile of the posterior distribution.

(b) Case 4: We generated the synthetic XMM-Newton data assuming a background comparable to the blank-sky background, and also assumed this when we fit the synthetic data. The posterior peaks around the radius value assumed during synthetic data generation, and the true radius lying at the 38th percentile of the posterior distribution.

Figure 3. The mass and radius posterior probability distributions from our combined MultiNest and PT-emcee analysis for Case 3 (a) and Case 4 (b). In each case, the orange diamond in the 2D plot and the orange dashed lines in the 1D plots mark the mass and radius values assumed in generating the synthetic data.

symbol types are the same as in Figure 2a. As shown in Figure 2b, the resulting radius posterior extends to smaller radii than the true value, producing a bias of roughly 1σ relative to the radius assumed in generating the synthetic data (see Table 3). This is measurable, but still within the 90% credible region.

This bias arises because neglecting the additional background in the model causes the inferred spot contribution to be overestimated, which reduces the inferred modulation amplitude. In turn, this yields an overestimate of the stellar compactness and, given the precise external mass constraint, an underestimate of the radius. The full posterior distribution is shown in Figure 5 in the Appendix.

The fits remain statistically acceptable. The χ^2/dof values for the phase–energy and bolometric fits are 2980.24/2901 and 29.01/25, with corresponding probabilities of 0.149 and 0.263, respectively. However, the Bayesian evidence for this model is lower than in Case 1 by $\Delta \log Z = -8.59 \pm 0.11$ (see Table 4), indicating that the model without an additional background is less favored by a factor of $e^{8.59} \sim 5,400$. This comparison

supports the conclusion that a waveform model including extra background provides a more accurate description of the data.

4.3. *The synthetic data do not include an additional background but the fitted model assumes an additional background*

In Case 3, we analyzed synthetic data that included only the XMM-Newton blank-sky background, but fit it with a model that assumed an additional background. When the model assumes a background that is higher than the background used to generate the synthetic data, the unmodulated XMM background is therefore too large. Given the fixed synthetic data, the best-fit spot waveform will therefore underestimate the total number of counts that come from the spots. In general, some of the counts from the spots will be modulated, and some can be unmodulated. The NICER data unambiguously tells us the number of modulated counts, which must come from the hot spots, so the ratio of modulated counts to the total apparent spot counts becomes artificially high. Thus, the modulation fraction (i.e., the number of modulated counts divided by the

inferred total counts coming from the spots) will appear to be larger than it should be. A fractional modulation that is too high leads to a larger inferred value of the inverse compactness (R/M). Given the mass constraint, the radius posterior extends to values larger than the radius that was assumed when the synthetic data were generated, producing a bias of about 1σ (see Figure 3a; the meanings of the line and symbol types are the same as in Figure 2a). The full posterior distribution is shown in Figure 6 in the Appendix.

The fit quality for this model is statistically acceptable: the phase-channel χ^2/dof is $2920.28/2901$ ($p = 0.397$), and the bolometric χ^2/dof is $39.24/25$ ($p = 0.035$) (see Table 4). The log evidence for this fit is $\Delta\log Z = -5.76 \pm 0.10$ relative to the Case 4 model (which assumes only the blank-sky background), indicating that the model with an additional background is disfavored by a factor of $e^{5.76} \sim 300$. This suggests that including an additional background in the model when the data do not contain one leads to a poorer fit.

4.4. *The synthetic data does not include and the fitted model does not assume an additional background*

In Case 4 we fit a model that assumes the XMM-Newton blank-sky background is the only non-spot background to a synthetic pulse waveform that was generated using a model that made this same assumption. In this case, the radius estimate agrees with the radius assumed in the model (see Figure 3b and Table 3). The phase-channel χ^2/dof is $2916.59/2901$, which has a probability of 0.416, and the bolometric χ^2/dof is $37.41/25$, which has a probability of 0.053 (see Table 4). Both probabilities are acceptable. The full posterior distribution is shown in Figure 7 in the Appendix.

5. SUMMARY AND CONCLUSIONS

In this work we have investigated how the assumption made about the X-ray background of non-accreting X-ray pulsars when fitting pulse waveform models jointly to NICER and XMM-Newton can affect estimates of the radii of these neutron stars, focusing on the potential for systematic error. We considered the four combinations that occur if one assumes that the background in the data is or is not solely the XMM-Newton blank-sky background and that the background assumed in the model is or is not solely the XMM-Newton background.

If the data contain a background (produced, e.g., by cosmic rays, AGN, or unresolved X-ray sources) that is missing from the model, the inferred stellar radius decreases. This is because a more compact star produces greater spacetime curvature and thus weaker modulation, mimicking the effect of a larger background. Conversely, if the model includes a background that is not present in the data, the inferred radius increases to preserve the fractional modulation. The radius estimate is unbiased compared to the actual stellar radius when the background in the model matches the background in the data.

We considered a large additional background in one of our synthetic data sets, allowing a background equal to the full smoothed synthetic count rate given by the pointed XMM-Newton observation. Even with this very large unmodeled background, the bias in the estimated stellar radius was small, corresponding to only about a $\sim 1\sigma$ offset from the true value. Also, in every case we considered the Bayesian evidence favored the correct model, which suggests its usefulness in choosing the best model for analyzing this kind of data. This was so even though our investigation considered only synthetic pulse waveform data similar to the NICER data from PSR J0740+6620, which has a relatively low signal-to-noise ratio.

These results add to the evidence that NICER-like analyses provide accurate measurements of neutron star radii as long as the statistical sampling is thorough and the data are fit well by the model.

REFERENCES

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, PhRvL, 116, 061102, doi: 10.1103/PhysRevLett.116.061102

Badnell, N. R., Bautista, M. A., Butler, K., et al. 2005, MNRAS, 360, 458, doi: 10.1111/j.1365-2966.2005.08991.x

Bogdanov, S., Heinke, C. O., Özel, F., & Güver, T. 2016, ApJ, 831, 184, doi: 10.3847/0004-637X/831/2/184

Bogdanov, S., Dittmann, A. J., Ho, W. C. G., et al. 2021, ApJL, 914, L15, doi: 10.3847/2041-8213/abfb79

Buchner, J. 2016, PyMultiNest: Python interface for MultiNest, Astrophysics Source Code Library, record ascl:1606.005

Carter, J. A., & Read, A. M. 2007, A&A, 464, 1155, doi: 10.1051/0004-6361:20065882

De, S., Finstad, D., Lattimer, J. M., et al. 2018, PhRvL, 121, 091102, doi: 10.1103/PhysRevLett.121.091102

Demorest, P. B., Pennucci, T., Ransom, S. M., Roberts, M. S. E., & Hessels, J. W. T. 2010, Nature, 467, 1081, doi: 10.1038/nature09466

Dittmann, A. 2024, The Open Journal of Astrophysics, 7, 79, doi: 10.33232/001c.123872

Dittmann, A. J., Miller, M. C., Lamb, F. K., et al. 2024, ApJ, 974, 295, doi: 10.3847/1538-4357/ad5f1e

Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449, doi: 10.1111/j.1365-2966.2007.12353.x

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601, doi: 10.1111/j.1365-2966.2009.14548.x

Fonseca, E., Cromartie, H. T., Pennucci, T. T., et al. 2021, ApJL, 915, L12, doi: 10.3847/2041-8213/ac03b8

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, doi: 10.1086/670067

Gendreau, K. C., Arzoumanian, Z., & Okajima, T. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8443, Space Telescopes and Instrumentation 2012: Ultraviolet to Gamma Ray, ed. T. Takahashi, S. S. Murray, & J.-W. A. den Herder, 844313, doi: 10.1117/12.926396

Guillot, S., Rutledge, R. E., & Brown, E. F. 2011, ApJ, 732, 88, doi: 10.1088/0004-637X/732/2/88

Guillot, S., Servillat, M., Webb, N. A., & Rutledge, R. E. 2013, ApJ, 772, 7, doi: 10.1088/0004-637X/772/1/7

Guillot, S., Kaspi, V. M., Archibald, R. F., et al. 2016, MNRAS, 463, 2612, doi: 10.1093/mnras/stw2194

Güver, T., Wroblewski, P., Camarota, L., & Özel, F. 2010, ApJ, 719, 1807, doi: 10.1088/0004-637X/719/2/1807

Ho, W. C. G., & Heinke, C. O. 2009, Nature, 462, 71, doi: 10.1038/nature08525

Hsu, S. D. H., & Reeb, D. 2010, International Journal of Modern Physics A, 25, 53, doi: 10.1142/S0217751X10047968

Ih, J., & Kempton, E. M. R. 2021, AJ, 162, 237, doi: 10.3847/1538-3881/ac173b

Kajava, J. J. E., Nättilä, J., Latvala, O.-M., et al. 2014, MNRAS, 445, 4218, doi: 10.1093/mnras/stu2073

Karamanis, M., Nabergoj, D., Beutler, F., Peacock, J., & Seljak, U. 2022, The Journal of Open Source Software, 7, 4634, doi: 10.21105/joss.04634

Kim, M., Kim, Y.-M., Sung, K. H., Lee, C.-H., & Kwak, K. 2021, A&A, 650, A139, doi: 10.1051/0004-6361/202038126

Lattimer, J. M., & Steiner, A. W. 2014, ApJ, 784, 123, doi: 10.1088/0004-637X/784/2/123

Lemos, P., Weaverdyck, N., Rollins, R. P., et al. 2023, MNRAS, 521, 1184, doi: 10.1093/mnras/stac2786

Lo, K. H., Miller, M. C., Bhattacharyya, S., & Lamb, F. K. 2013, ApJ, 776, 19, doi: 10.1088/0004-637X/776/1/19

Majczyna, A., & Madej, J. 2006, arXiv e-prints, astro, doi: 10.48550/arXiv.astro-ph/0601076

Marino, A., Degenaar, N., Di Salvo, T., et al. 2018, MNRAS, 479, 3634, doi: 10.1093/mnras/sty1585

Miller, M. C. 2013, arXiv e-prints, arXiv:1312.0029, doi: 10.48550/arXiv.1312.0029

—. 2016, ApJ, 822, 27, doi: 10.3847/0004-637X/822/1/27

Miller, M. C., & Lamb, F. K. 2015, ApJ, 808, 31, doi: 10.1088/0004-637X/808/1/31

—. 2016, European Physical Journal A, 52, 63, doi: 10.1140/epja/i2016-16063-8

Miller, M. C., Lamb, F. K., Dittmann, A. J., et al. 2019, ApJL, 887, L24, doi: 10.3847/2041-8213/ab50c5

—. 2021, ApJL, 918, L28, doi: 10.3847/2041-8213/ac089b

Nättilä, J., Miller, M. C., Steiner, A. W., et al. 2017, A&A, 608, A31, doi: 10.1051/0004-6361/201731082

Özel, F., Psaltis, D., Güver, T., et al. 2016, ApJ, 820, 28, doi: 10.3847/0004-637X/820/1/28

Pearson, K. 1900, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50, 157, doi: 10.1080/14786440009463897

Remillard, R. A., Loewenstein, M., Steiner, J. F., et al. 2022, AJ, 163, 130, doi: 10.3847/1538-3881/ac4ae6

Riley, T. E., Watts, A. L., Ray, P. S., et al. 2021, ApJL, 918, L27, doi: 10.3847/2041-8213/ac0a81

Romani, R. W., Kandel, D., Filippenko, A. V., Brink, T. G., & Zheng, W. 2022, ApJL, 934, L17, doi: 10.3847/2041-8213/ac8007

Saffer, A., Fonseca, E., Ransom, S., et al. 2024, arXiv e-prints, arXiv:2412.02850, doi: 10.48550/arXiv.2412.02850

Salmi, T., Vinciguerra, S., Choudhury, D., et al. 2022, ApJ, 941, 150, doi: 10.3847/1538-4357/ac983d

Salmi, T., Choudhury, D., Kini, Y., et al. 2024, ApJ, 974, 294, doi: 10.3847/1538-4357/ad5f1f

Savitzky, A., & Golay, M. J. E. 1964, Analytical Chemistry, 36, 1627, doi: 10.1021/ac60214a047

Sivia, D., & Skilling, J. 2006, Data analysis: a Bayesian tutorial (OUP Oxford)

Steiner, A. W., Lattimer, J. M., & Brown, E. F. 2010, ApJ, 722, 33, doi: 10.1088/0004-637X/722/1/33

Troyer, M., & Wiese, U.-J. 2005, PhRvL, 94, 170201, doi: 10.1103/PhysRevLett.94.170201

van Paradijs, J. 1979, ApJ, 234, 609, doi: 10.1086/157535

Wolff, M. T., Guillot, S., Bogdanov, S., et al. 2021, ApJL, 918, L26, doi: 10.3847/2041-8213/ac158e

7. APPENDIX: POSTERIOR DISTRIBUTIONS FROM ANALYSES OF SYNTHETIC J0740-LIKE NICER AND XMM-NEWTON DATA USING MULTINEST AND PT-EMCEE

Figures 4–7 show the full set of posterior distributions for the neutron star radii together with other model parameters, that are not included in Figures 2a–3b of the main text.
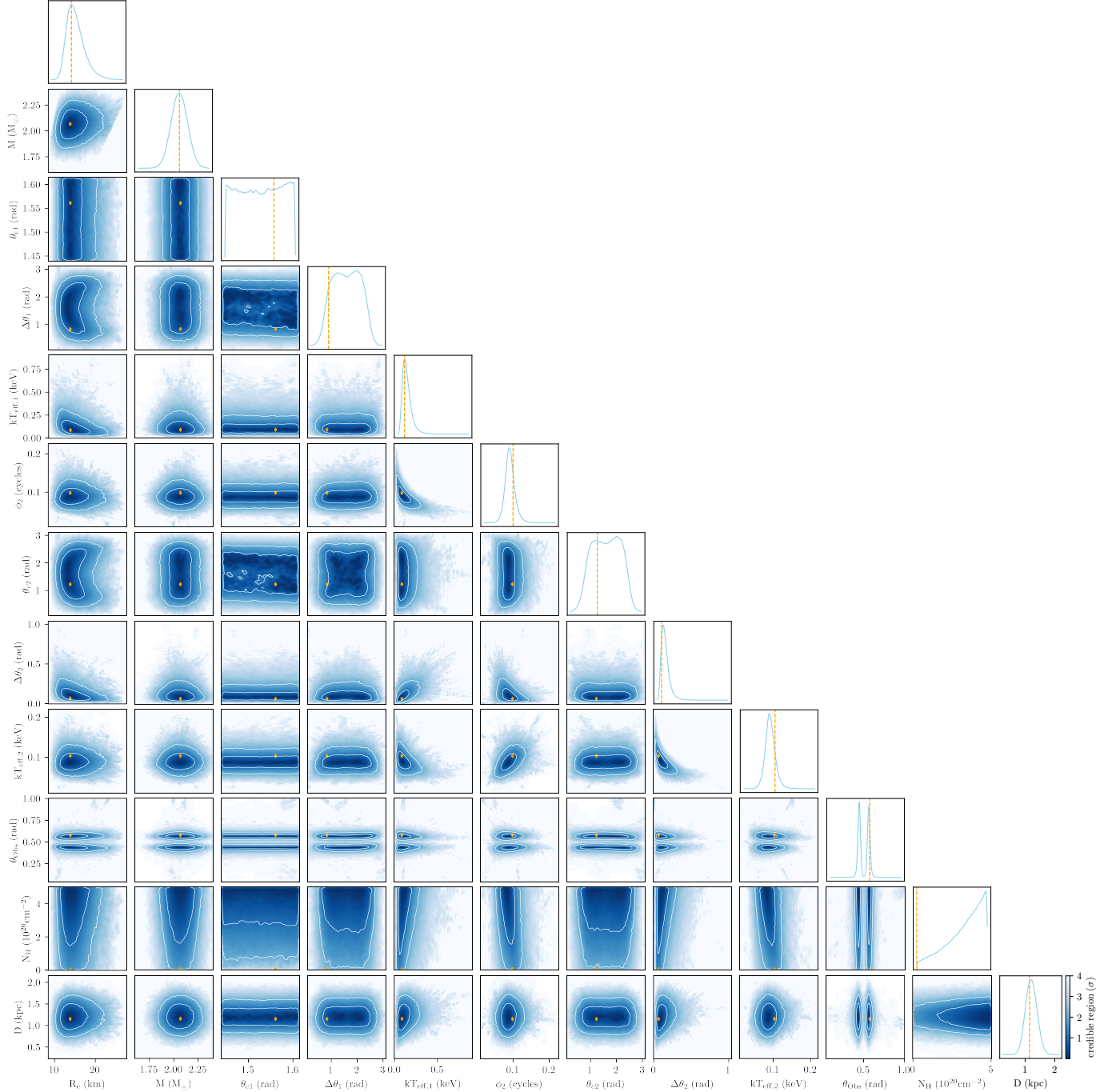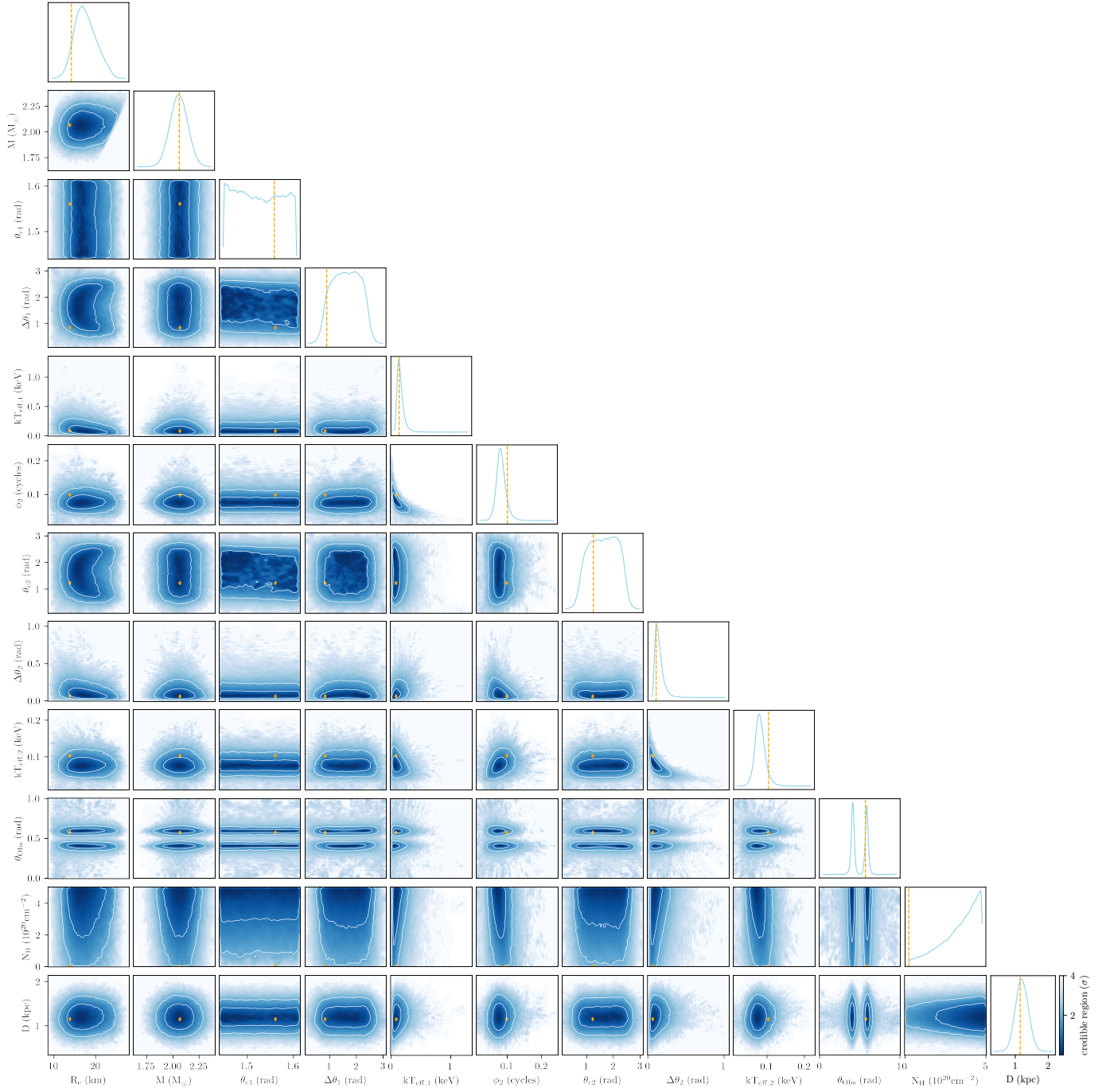

Figure 4. The full set of 1D and 2D posterior distributions from our MultiNest and PT-emcee analysis of Case 1. The synthetic data were constructed assuming that the background is the XMM-Newton blank-sky background plus an additional background (see text). The analysis model also assumes that the background is the XMM-Newton blank-sky background plus an additional background.
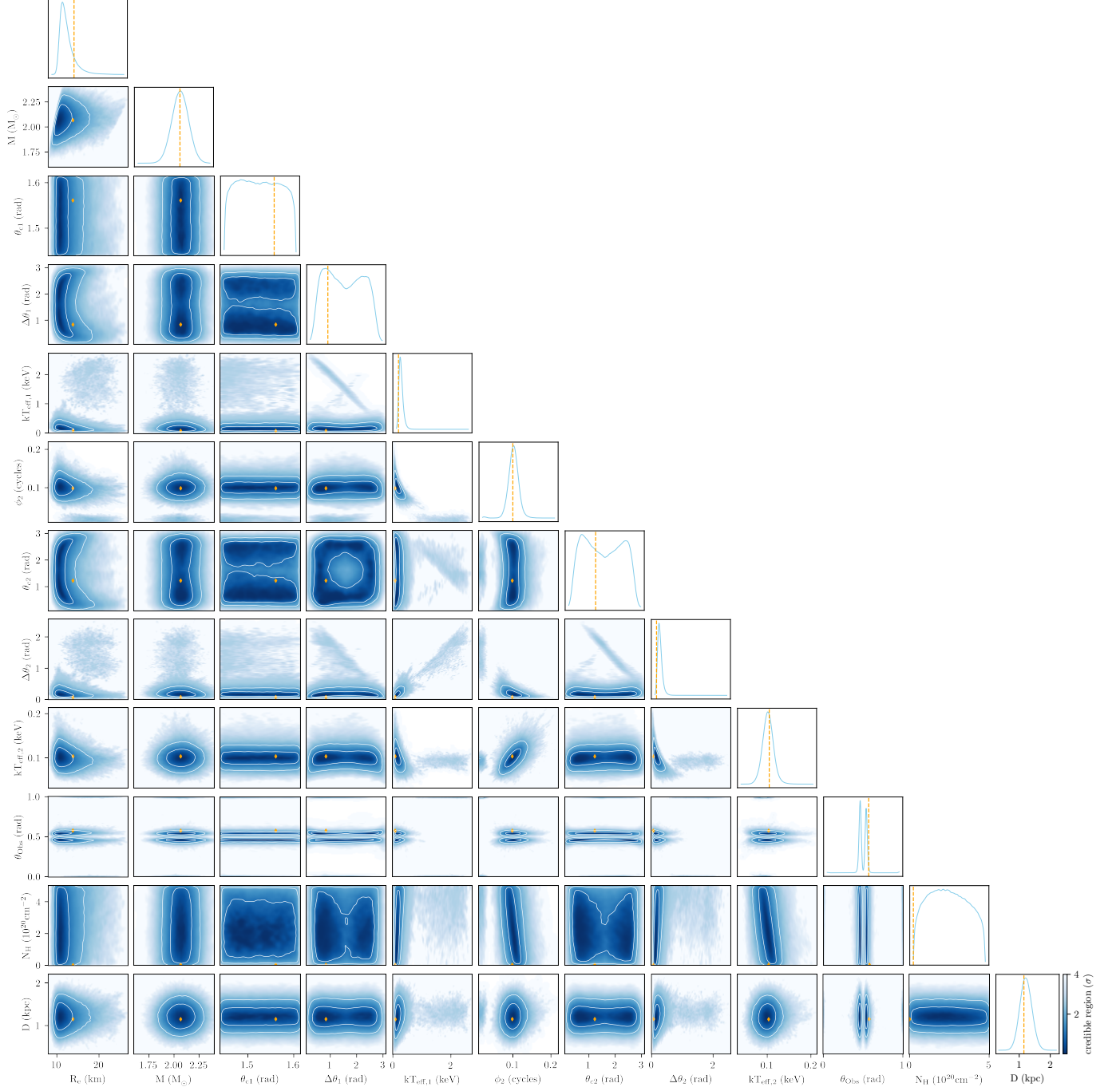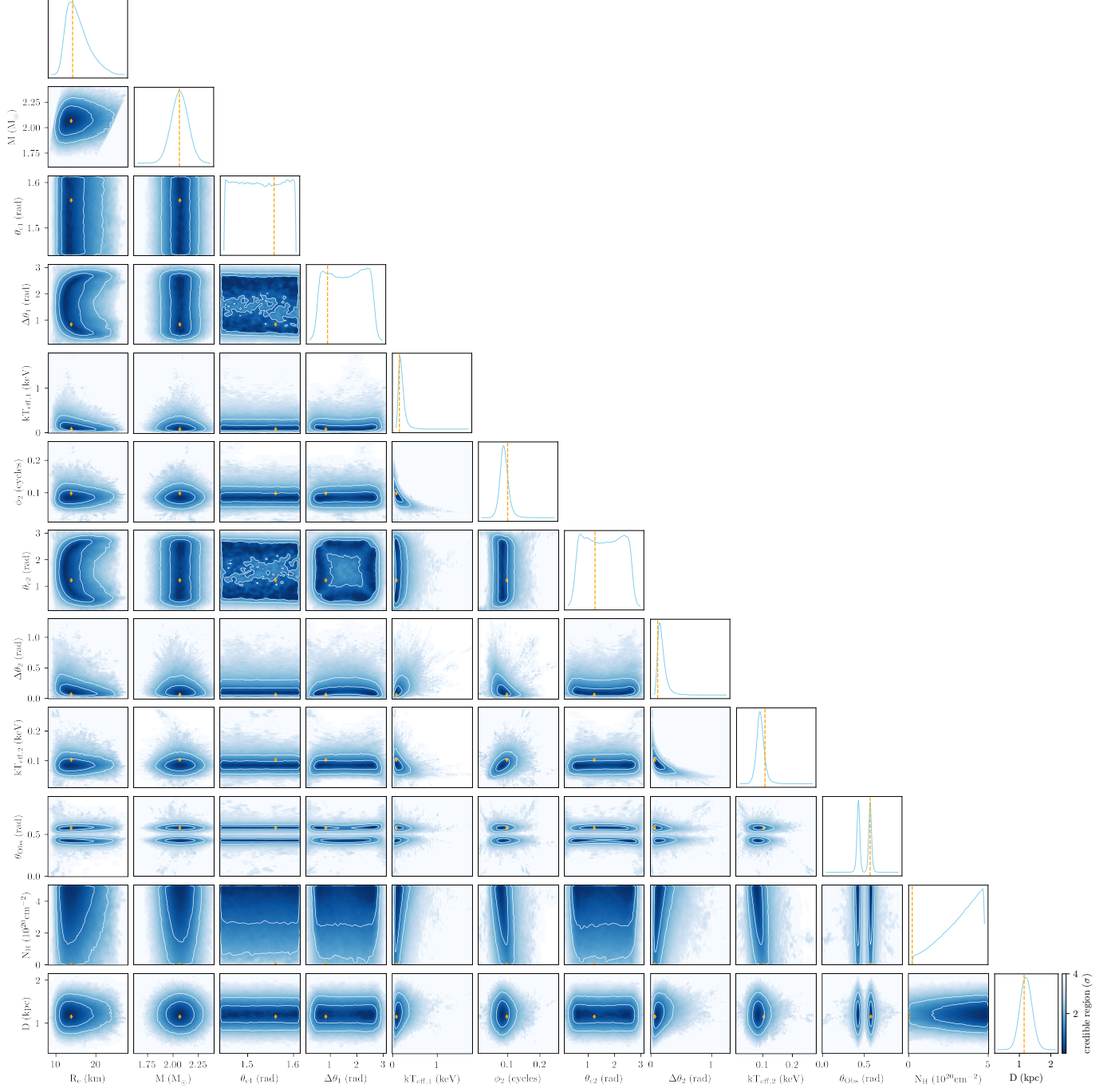
Figure 5. The full set of 1D and 2D posterior distributions from our MultiNest and PT-emcee analysis of Case 2. The synthetic data were constructed assuming that the background is the XMM-Newton blank-sky background plus an additional background (see text). The analysis model assumes that the XMM-Newton blank-sky background is the only background.

Figure 6. The full set of 1D and 2D posterior distributions from our MultiNest and PT-emcee analysis of Case 3. The synthetic data were constructed assuming that the background is the XMM-Newton blank-sky background. The analysis model assumes that the background is the XMM-Newton blank-sky background plus an additional background.

Figure 7. The full set of 1D and 2D posterior distributions from our MultiNest and PT-emcee analysis of Case 4. The synthetic data and the model both assume that the XMM-Newton blank-sky background is the only background.