The Astrophysical Journal Letters, 887:L21 (60pp), 2019 December 10 https://doi.org/10.3847/2041-8213/ab481c
© 2019. The American Astronomical Society. All rights reserved.

A *NICER* View of PSR J0030+0451: Millisecond Pulsar Parameter Estimation

T. E. Riley [1], A. L. Watts [1], S. Bogdanov [2], P. S. Ray [3], R. M. Ludlam [4,16], S. Guillot [5,6], Z. Arzoumanian [7], C. L. Baker [8], A. V. Bilous [1], D. Chakrabarty [9], K. C. Gendreau [7], A. K. Harding [10], W. C. G. Ho [11,12], J. M. Lattimer [13], S. M. Morsink [14], and T. E. Strohmayer [15]

[1] Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1090GE Amsterdam, The Netherlands; T.E.Riley@uva.nl, A.L.Watts@uva.nl
[2] Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA
[3] Space Science Division, U.S. Naval Research Laboratory, Washington, DC 20375, USA
[4] Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA
[5] IRAP, CNRS, 9 avenue du Colonel Roche, BP 44346, F-31028 Toulouse Cedex 4, France
[6] Université de Toulouse, CNES, UPS-OMP, F-31028 Toulouse, France
[7] X-Ray Astrophysics Laboratory, NASA Goddard Space Flight Center, Code 662, Greenbelt, MD 20771, USA
[8] Applied Engineering and Technology Directorate, NASA Goddard Space Flight Center, Code 592, Greenbelt, MD 20771, USA
[9] MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[10] Astrophysics Science Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[11] Department of Physics and Astronomy, Haverford College, 370 Lancaster Avenue, Haverford, PA 19041, USA
[12] Mathematical Sciences, Physics and Astronomy, and STAG Research Centre, University of Southampton, Southampton SO17 1BJ, UK
[13] Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794-3800, USA
[14] Department of Physics, University of Alberta, 4-183 CCIS, Edmonton, AB T6G 2E1, Canada
[15] Astrophysics Science Division and Joint Space-Science Institute, NASA's Goddard Space Flight Center, Greenbelt, MD 20771, USA
Received 2019 July 12; revised 2019 September 24; accepted 2019 September 25; published 2019 December 12

Abstract

We report on Bayesian parameter estimation of the mass and equatorial radius of the millisecond pulsar PSR J0030+0451, conditional on pulse-profile modeling of *Neutron Star Interior Composition Explorer* X-ray spectral-timing event data. We perform relativistic ray-tracing of thermal emission from hot regions of the pulsar's surface. We assume two distinct hot regions based on two clear pulsed components in the phase-folded pulse-profile data; we explore a number of forms (morphologies and topologies) for each hot region, inferring their parameters in addition to the stellar mass and radius. For the family of models considered, the evidence (prior predictive probability of the data) strongly favors a model that permits both hot regions to be located in the same rotational hemisphere. Models wherein both hot regions are assumed to be simply connected circular single-temperature spots, in particular those where the spots are assumed to be reflection-symmetric with respect to the stellar origin, are strongly disfavored. For the inferred configuration, one hot region subtends an angular extent of only a few degrees (in spherical coordinates with origin at the stellar center) and we are insensitive to other structural details; the second hot region is far more azimuthally extended in the form of a narrow arc, thus requiring a larger number of parameters to describe. The inferred mass M and equatorial radius $R_{\rm eq}$ are, respectively, $1.34^{+0.15}_{-0.16}\,M_\odot$ and $12.71^{+1.14}_{-1.19}$ km, while the compactness $GM/R_{\rm eq}c^2 = 0.156^{+0.008}_{-0.010}$ is more tightly constrained; the credible interval bounds reported here are approximately the 16% and 84% quantiles in marginal posterior mass.

Unified Astronomy Thesaurus concepts: Matter density (1014); Rotation powered pulsars (1408); Millisecond pulsars (1062); Pulsars (1306); X-ray stars (1823); Neutron stars (1108); Neutron star cores (1107); Nuclear astrophysics (1129)

Supporting material: figure sets

1. Introduction

Neutron star (NS) cores are thought to harbor nucleonic matter under extreme conditions: high in density, neutron rich, and potentially strange. Stable states of strange matter may either be bound in the form of hyperons, or deconfined as a mixture of up, down, and strange quarks. The density in the stellar core may reach up to several times the nuclear saturation density $\rho_{\rm sat} = 2.8 \times 10^{14}$ g cm^{-3}, a range for which we cannot yet calculate the state of nuclear matter from first principles.[17] Instead, theorists develop phenomenological models of particle interactions and phase transitions, which must be tested by experiment and observation. Heavy ion collision experiments explore the high-temperature and lower-density parts of the nuclear matter phase diagram; but NSs are unique laboratories for the study of strong and weak force physics in cold, dense matter (for recent reviews see Lattimer & Prakash 2016; Oertel et al. 2017; Baym et al. 2018).

The particle interactions on a microphysical scale emerge macroscopically as an equation of state (EOS)—in the context of cold dense matter, a relationship between pressure and (energy) density. The EOS forms part of the relativistic stellar structure equations that enable us, given a central density and a spin rate, to compute model NSs (e.g., Hartle 1967; Hartle & Thorne 1968). An EOS function thus maps to a sequence of stable global spacetime solutions, each controlled in the exterior domain by parameters such as the mass and equatorial radius (at low orders in a small dimensionless spin parameter;

[16] Einstein Fellow.

[17] Although note that calculations at sub-saturation densities (see, e.g., Hebeler et al. 2010), and perturbative quantum chromodynamics (QCD) modeling (valid for densities that are several times higher than the maximum expected in NS cores; see Kurkela et al. 2014) can impose some constraints on the low- and high-density limits of NS core parameter space.

Hartle 1967). In this work we constrain the total mass (sometimes referred to as the gravitational mass) and the equatorial radius of the star, respectively defined as the mass and coordinate radius in the Schwarzschild metric. If we can statistically estimate the masses and radii of a set of stars whose central densities span some sufficiently broad range, we can in principle map out the EOS and hence make inferential statements about the microphysics (Özel & Psaltis 2009; Steiner et al. 2010, 2013; Nättilä et al. 2016; Özel et al. 2016; Raithel et al. 2017; Raaijmakers et al. 2018; Riley et al. 2018; Greif et al. 2019).

The strongest statistical constraints[18] on NS masses are derived by timing radio pulsars in (compact) binaries, and rely on our well-established understanding of relativistic orbital dynamics. Every EOS function (corresponding to a parameter vector for a parameterized model) permits stable spacetime solutions with a maximum gravitational mass—associated with a specific central density—beyond which no stable solutions can exist. High-mass NSs with tight constraints can therefore effectively exclude[19] a subset of EOS function space, barring strong contention with future analyses of independently acquired data. The most informative pulsars in this regard are PSR J0348+0432, with mass $2.01 \pm 0.04\ M_\odot$ (Antoniadis et al. 2013, where the mass is derived by combining pulsar timing and models of the white dwarf companion), and PSR J1614−2230 with mass $1.908 \pm 0.016\ M_\odot$ (Arzoumanian et al. 2018, where the mass comes from Shapiro delay estimation).[20] More recently, Cromartie et al. (2019) have reported a higher—but at present more uncertain—mass of $2.14^{+0.10}_{-0.09}\ M_\odot$ for PSR J0740+6620. We note, however, that the Cromartie et al. (2019) measurement is not subject to the systematic uncertainty that should be added to the formal uncertainty on the mass of PSR J0348+0432 due to the latter's dependence on theoretical models of white dwarf evolution.

The radio pulsar timing of compact binaries has yet to deliver a radius constraint, although this is feasible and indeed anticipated, via moment of inertia estimation (Kramer & Wex 2009). There are, however, constraints on radius via X-ray spectral modeling of transiently accreting and bursting NSs (see, e.g., Steiner et al. 2013; Özel et al. 2016; Nättilä et al. 2017; Shaw et al. 2018; Baillot d'Etivaux et al. 2019); we refer the reader to Miller (2013) and Özel & Freire (2016) for detailed reviews that include an explanation of these X-ray modeling techniques and associated uncertainties. NS mass and tidal deformability estimates are now also being reported based on the first binary NS merger gravitational wave event, GW170817 (Abbott et al. 2017, 2018; De et al. 2018). These can be translated, usually by means of universal relations or EOS model assumptions, into constraints on mass and radius (see, e.g., Abbott et al. 2018; Annala et al. 2018; De et al. 2018; Most et al. 2018; Tews et al. 2018b). Generally, assuming that both NSs have the same EOS, it is found that their radii are nearly equal and (for the 68% credible interval) have the common value 11.9 ± 1.1 km.

NASA's *Neutron Star Interior Composition Explorer* (*NICER*; Gendreau et al. 2016), a soft X-ray telescope installed on the *International Space Station* (*ISS*) in 2017, was developed in part to estimate masses and radii of NSs using pulse-profile modeling of nearby rotation-powered millisecond pulsars (MSPs). Pulse-profile modeling is a technique that probes (approximations to) general relativistic effects on thermal emission from hot regions on the stellar surface (Bogdanov et al. 2019b); these effects are, predominantly, local radiation beaming[21] *due to the bulk motion* of material on the rotationally deformed surface,[22] and subsequent ray propagation on the exterior spacetime. Ray propagation includes the canonical *bending* of light, gravitational redshift, and the increasingly small imprints of rotational metric deformation: frame-dragging, a finite mass quadrupole moment, and higher-order (mass and current) multipole moments. As the star spins, the flux and spectrum of X-ray emission registered by a distant observer is modulated in a periodic manner: we can determine the rotational phase evolution of pulsars precisely and build up a pulse profile (X-ray counts per rotational phase bin per detector channel) by phase-folding X-ray events according to an ephemeris.[23] The mapping of surface emission into the pulse profile detected by a distant observer, via relativistic ray-tracing through the spacetime of a rapidly rotating (and hence oblate) star, is well understood (Pechenick et al. 1983; Miller & Lamb 1998; Poutanen & Gierliński 2003; Poutanen & Beloborodov 2006; Cadeau et al. 2007; Morsink et al. 2007; Bauböck et al. 2013; AlGendy & Morsink 2014; Psaltis & Özel 2014; Nättilä & Pihajoki 2018; Vincent et al. 2018). Thus, given a model for the surface emission (e.g., a geometrically thin atmosphere of some chemical and ionic composition together with a local comoving effective temperature field as a function of surface coordinates), one calculates the expected pulse profile for a given exterior spacetime solution and a given instrument. By coupling such light-curve models to statistical sampling software via efficient software implementations, we can use Bayesian inference to derive posterior probability distributions for spacetime parameters such as mass and equatorial radius directly from pulse-profile data.

For the pulse-profile modeling technique to deliver tight constraints on mass and radius, rapid spin ($\gtrsim$100 Hz) is desirable (Psaltis et al. 2014; Miller & Lamb 2015; Stevens et al. 2016), and one needs high-quality phase- and energy-resolved pulse profiles with time resolution $\leqslant 10\ \mu$s and a large number of photons. The precise number of photons needed to deliver constraints on mass and radius at levels of a few percent—and by extension tight constraints on EOS models—depends on the geometry of a given source, but is $\sim 10^6$ pulsed photons (Lo et al. 2013; Psaltis et al. 2014; Miller & Lamb 2015). For the brightest of the rotation-powered MSPs targeted by *NICER*, it is feasible to collect sufficient data with observation times $\sim$1 Ms. The hot regions on rotation-powered MSPs in theory arise as

[18] Or *measurements*, or *estimates*; in any case, this means some probabilistic measure that is a function of, or otherwise pertains to, model parameters. See Section 2.3 for the probabilistic measures that we consider in this present work.

[19] In a Bayesian context, by truncating the mass likelihood function only far in the tails, leading to a finite but small marginal posterior density for EOSs that do have substantially smaller maximum supported masses.

[20] Note that PSR J1614−2230 was initially reported as having mass 1.97 ± 0.04 (Demorest et al. 2010); the inferences have since been updated via analysis of newly acquired data.

[21] Note that local effective gravity in local comoving frames (instantaneously inertial during rotation) also enters calculation of *atmospheric* beaming of radiation emergent from the local comoving photosphere.

[22] Where for statistical applications the surface is either self-consistently computed via matching to a numerical interior solution to the field equations, or is embedded via a quasi-universal relation in an ambient spacetime solution (for an overview see, e.g., Riley et al. 2018, and references therein).

[23] Note a key difference to the X-ray spectral modeling mentioned two paragraphs earlier: pulse-profile modeling involves phase-resolved spectroscopy; spectral modeling is phase-averaged, and does not fully leverage the temporal dimension of information provided by the star's rotation.

magnetospheric currents—including return currents—deposit energy in the surface layers of the star; the resulting surface radiation field is a priori highly uncertain (Harding & Muslimov 2001; Gralla et al. 2017; Bauböck et al. 2019). *NICER* pulse-profile modeling can therefore also help to constrain the characteristics of the hot regions.

In this Letter we undertake pulse-profile modeling of *NICER X-ray Timing Instrument* (*XTI*) observations of the rotation-powered MSP PSR J0030+0451. Discovered as a radio pulsar by Lommen et al. (2000) and then identified as an X-ray pulsar (Becker et al. 2000), PSR J0030+0451 has a spin frequency of 205 Hz and lies at a distance of 325 ± 9 pc (Arzoumanian et al. 2018). There are no independent prior constraints on either mass or radius. Our analysis uses the *X-ray Pulsation Simulation and Inference* package (*X-PSI*[24] `v0.1`; Riley & Watts 2019). *X-PSI* is a software package for Bayesian modeling of astrophysical X-ray pulsations generated by the rotating, radiating surfaces of relativistic compact stars. *X-PSI* couples X-ray pulsation likelihood functionality to open-source statistical sampling software for use on high-performance computing systems; we apply nested sampling (Skilling 2006) in our analysis (refer to Appendix A). The work presented here is based on usage of a 500,000 core-hour grant on the Dutch national supercomputer Cartesius.[25]

Section 2 outlines the modeling choices and introduces details specific to this analysis, including issues associated with the PSR J0030+0451 surface radiation field parameterization, instrument response, prior definition and implementation, and the consequences for computational efficiency of posterior sampling. In particular we restrict this analysis to models with two distinct hot regions with various structures. While our choices are physically motivated, it is important to emphasize that our inferences are conditional upon these choices. Posterior[26] inferences for the models are presented in Section 3, including the inferred posterior probability distributions for the spacetime parameters (mass, radius, and their combination into compactness), surface radiation field parameters (e.g., heating distribution and resulting temperature field), and instrument parameters. The Bayesian evidence for each model is reported, and we also summarize the computational resources required for each parameter estimation run. We also compare our inference to predictions derived via earlier study of PSR J0030+0451 using *XMM-Newton* (Bogdanov & Grindlay 2009). Sections 2 and 3 are long and detailed; in Section 2.1 we provide a brief overview of some of the key aspects from those sections, to help orient the reader. In Section 4 we discuss the implications of our results for our understanding of dense matter, pulsar emission mechanisms, and stellar evolution. We conclude with a discussion of future work for PSR J0030+0451: variations in the model that should be considered, tests and cross-checks, and the potential for improving the constraints for this source via longer observations or more in-depth analysis.

2. Modeling Procedure

2.1. Executive Summary of Modeling Procedure and Inferences

This Letter (and its companions) are the first pulse-profile modeling analyses to emerge from the *NICER* mission. We have therefore provided (in Sections 2 and 3) a very detailed description of the methodology, the flow of the analysis, and the results. Since this is lengthy, we summarize the key aspects in this subsection.

The pulse-profile modeling technique requires us to define a model for the data-generating process, incorporating the physics that we initially assume to be most important. For a given choice of parameters, this model can be used to generate synthetic pulse-profile data sets. The model in part defines the likelihood function (the probability of the data as a function of parameters); the model also defines the prior probability distribution of parameters entering in the likelihood function. The posterior probability distribution of the model parameters (conditional on the observed data) is then sampled during the inference process. We must also pay attention to model complexity, in order to keep the computational load tractable.

In this Letter we assume that there are two separate hot regions on the stellar surface; this choice was motivated by the presence of two distinct pulses in the observed (phase-folded) pulse-profile. However, we considered a number of different possible configurations for the shapes and temperature functions of the hot regions: circular spots, annuli (rings, both centered and off-centered), and crescents; with one or two temperature components. These choices were motivated by contemporary theories for pulsar surface heating distributions as a result of magnetospheric return currents. We tested configurations where we insisted that the two hot regions were antipodal and identical; and where the hot regions were completely independent and potentially non-antipodal. We then assumed a geometrically thin fully ionized hydrogen atmosphere model (`NSX`, Ho & Heinke 2009), which characterizes the beaming and spectrum of the emergent thermal radiation (see, e.g., Zavlin et al. 1996, for details).

To propagate the emergent radiation toward the observer via relativistic ray-tracing, we use the Oblate Schwarzschild plus Doppler approximation of Morsink et al. (2007) for the NS spacetime. This is sufficiently accurate for our analysis, given the rotation rate of PSR J0030+0451. We define a joint prior distribution of mass and radius (the key parameters specifying the spacetime) that facilitates the subsequent inference of EOS model parameters (Riley et al. 2018; Raaijmakers et al. 2019). For the distance to the source, we use the (Gaussian-distributed) value inferred from radio observations (Arzoumanian et al. 2018) as a prior in our modeling. We then need to model the instrument response matrix (which includes both the effective area and the way in which incident photons of a given energy are assigned to specific detector energy channels). We develop a parameterized model that includes both energy-independent and energy-dependent components. The former attempts to capture absolute calibration uncertainty; for the latter we base our parameterization on residuals derived from *NICER* observations of the Crab Nebula and pulsar (Ludlam et al. 2018). We also assume a non-source background component, which we treat as a rotational phase-independent channel-by-channel contribution, rather than invoking a specific physics-driven spectral model. There are no prior constraints on either observer inclination or interstellar absorption for PSR J0030+0451, so we adopt a wide and diffuse prior for both parameters.

Note that Miller et al. (2019) have made an independent analysis of the same data set using different modeling choices and methodology. The choices we have made in this Letter

[24] https://github.com/ThomasEdwardRiley/xpsi

[25] https://userinfo.surfsara.nl/systems/cartesius

[26] Appendix A provides an overview of the methodology used for posterior computation, and of the format used to present the posterior information.

differ in several regards from those made by Miller et al. (2019); some of the most notable differences are in the models for the hot region configurations and the instrumental response, and the specification of the prior on distance.

During the inference analysis reported in Section 3 we considered a sequence of increasingly complex models for the shape and temperature function of the hot regions. All of the other aspects of the modeling described above are shared between models. Model assessment and comparison then enabled the identification of a superior configuration. We use a combination of performance measures: the evidence (the prior predictive probability of the data); graphical posterior predictive checking (to verify whether or not a model generates synthetic data without obvious residual systematic structure in comparison to the real data); visualization of the combined signals from the hot regions; Kullback–Leibler (KL) divergences (a measure of the parameter-by-parameter information gain of the posterior over the prior); background-marginalized likelihood functions (useful in combination with evidence to assess whether additional model complexity is helpful); model tractability (posterior computational accuracy being higher for less-complex models); and cross-checking of the inferred background against earlier analysis of PSR J0030+0451 with *XMM-Newton*.

Before beginning our analysis, we had mapped out an initial route through the model space of different heating configurations. This was modified as we progressed, informed by the results of each stage. We began with the simplest model, with single-temperature circular spots. Having the spots be antipodal and identical was quickly ruled out due to large residuals between model and data. Relaxing the requirement that the spots be identical and antipodal largely resolved this issue. We then moved to a more complex model where each hot region consisted of a circular spot—a *core*—and a surrounding annulus with an independently determined temperature. This model was superior to the simpler one, based on the evidence, but appeared to be overly complex: one hot region was dominated by a small hot circular spot, with negligible emission in the *NICER* waveband from the cooler annulus; for the other, emission was dominated by a hot annulus, with a much cooler core making almost no contribution. Simplifying the model such that one hot region was a single-temperature circular spot and the other a single-temperature annulus (with a centered, non-emitting core) produced congruent inferences at lower computational cost. At this point (after assessing the contribution from this component, and our remaining computational resources), we elected to restrict the model for one of the hot regions to be a single-temperature circular spot. The other was restricted to a single temperature, but we increased the complexity of the shape, testing two additional models: an annulus with an off-centered non-emitting core, and a crescent.

The superior configuration to emerge from this sequence of models, in terms of the performance measures listed above, was the final one: one hot region a small circular spot (sufficiently small that we would be insensitive to shape changes), and the other was an extended thin crescent. The results that we report in the abstract for the mass and radius of PSR J0030+0451 are those associated with this configuration.

2.2. Data Pre-processing

The *NICER XTI* data set $\boldsymbol{d}$ associated with the rotation-powered MSP PSR J0030+0451 is necessarily a product of pre-processing, and its curation is largely described in Bogdanov et al. (2019a). However, there are a number of details that are specific to each data set and so we record a summary of these details below. The raw *XTI* event data are publicly available via the *NICER* archive (accessible via HEASARC[27]). The processed data set $\boldsymbol{d}$ may be found in the persistent repository of Bogdanov et al. (2019c).

For this analysis we consider only the detector channel[28] subset [25, 300)—meaning channels 25 through 299 inclusive—nominally corresponding to 0.25–3 keV. Below channel 25, there is increased "optical loading" contamination (electronic noise due to ambient light), and there is greater uncertainty in the detector readout triggering efficiency for valid X-ray events. Above channel 300 the soft thermal emission from PSR J0030+0451 becomes negligible relative to the non-source background.

For PSR J0030+0451 we use 49 out of the 52 active detectors (excluding the three detectors that are frequently "hot," i.e., prone to excess electronic noise) and only *NICER* pointings subtending an angle $>80°$ to the direction of the Sun. In addition to the standard filtering criteria applied to *XTI* events, we excluded times where the planetary K-index $K_P \geqslant 5$, and excluded regions in the *NICER* orbit where the cutoff rigidity (the COR_SAX parameter) was less than 1.5. We further screened for occurrences of elevated background and eliminated all 16 s time intervals in which the 0.25–8.00 keV count rate exceeded 3 counts s^{-1}, yielding an integrated exposure time of $T_{\rm exp}$: $=\sum_\ell \Delta t_\ell = 1,936,864$ s, where each Δt_ℓ is the time interval for the ℓth exposure.

Harnessing the flux and timing stability of PSR J0030+0451, the entire event data set in each detector channel is phase-folded coherently according to the best available radio pulsar timing solution from Arzoumanian et al. (2018) and using two approaches to check for consistency: using the TEMPO2 (Hobbs et al. 2006) `photons` plug-in and the PINT[29] `photonphase` tool. Differences of $\lesssim 1\ \mu$s as a phase-offset are observed, but are deemed negligible for the present analysis as such an offset corresponds to $\lesssim 0.02\%$ of the total phase, or less than 0.7% of a bin width. The resulting folded event list, obtained with the PINT `photonphase` tool, is summarized as count data in a set of rotational phase intervals and detector channels. We display in Figure 1 the count data for channels in the interval [25, 300), nominally corresponding to the energy range 0.25–3.00 keV. We reserve the remaining details of data-space definition for Section 2.4.3, wherein we formalize the likelihood function applied in this specific work.

2.3. Generative Modeling

We begin the description of our modeling process by outlining useful mathematical objects that are assigned statistical meaning, and which can interface with open-source computational machinery. In Section 2.4, we build upon this conceptual groundwork and assign astrophysical meaning to such objects.

[27] https://heasarc.gsfc.nasa.gov/

[28] Conventionally termed a pulse-invariant (PI) channel, or alternatively a charge pulse amplitude (or height) channel. Note the distinction between *channel*, which is only indicative of photon energy in units of 10 eV, and incident photon *energy intervals* that map probabilistically to the channel set via the redistribution matrix, a standard calibration product provided by the instrument team.

[29] https://github.com/nanograv/PINT

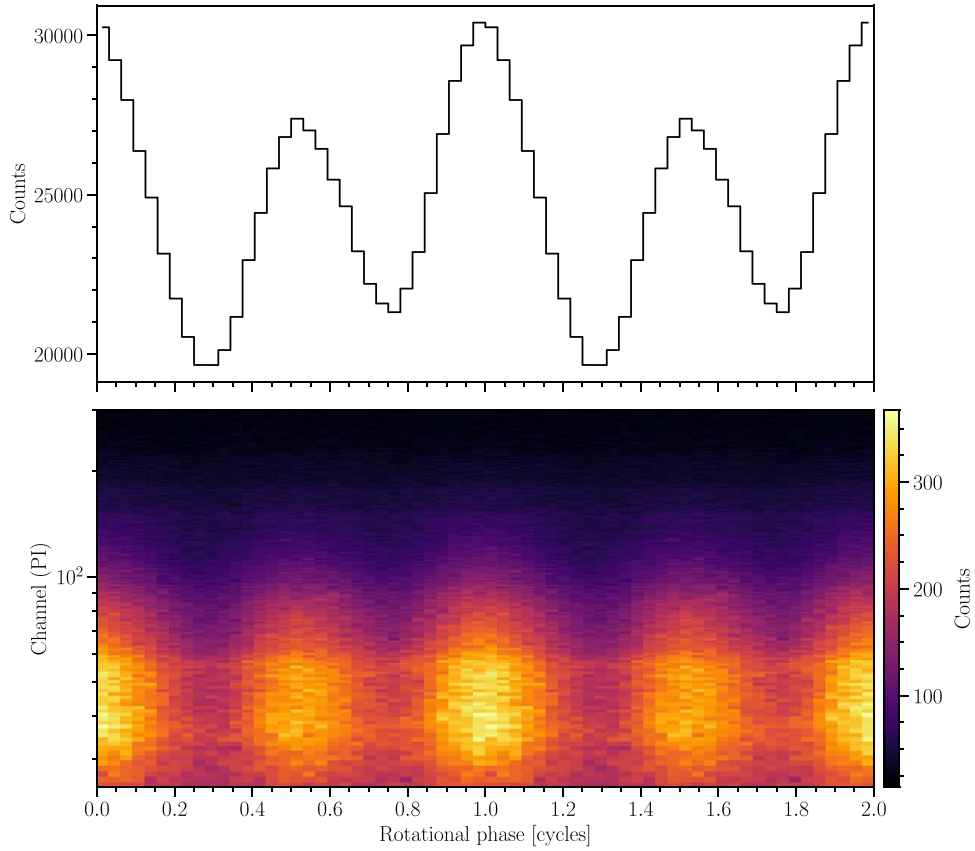

Figure 1. Phase-folded PSR J0030+0451 event data split over two rotational cycles for clarity: we use 32 phase intervals (bins) per cycle and the count numbers in bins separated by one cycle—in a given channel—are identical. The total number of counts is given by the sum over all phase-channel pairs. The top panel displays the pulse-profile summed over the contiguous subset of channels [25, 300). The bottom panel displays the phase-channel resolved count numbers for channel subset [25, 300). For likelihood function evaluation (see Section 2.4.3) we group all event data registered in a given channel into phase intervals spanning a single rotational cycle. Moreover, we do not indicate the count-number noise in the top panel to avoid confusion: the observed events are viewed as fixed random variates that do not have errors, and whose parameterized joint sampling distribution is to be modeled (see Section 2.3).

There exists an inherent degree of freedom as to the precise definition of a *model*. We consider a *generative* model for the X-ray data set $\boldsymbol{d}$ curated for PSR J0030+0451. Moreover, we opt to define a generative model as the union of the following components: (i) a data space $\mathscr{D}$ in which a data set $\boldsymbol{d} \in \mathscr{D}$ exists as a fixed vector of numbers; (ii) an abstract model space $\mathscr{M}$ of elements, such that each element completely specifies a joint sampling distribution on space $\mathscr{D}$ (i.e., the sampling distribution can be evaluated conditional on the element); (iii) a joint probability distribution defined on space $\mathscr{M}$ that is not conditional on data vector $\boldsymbol{d}$. Such a model is considered as generative because one can define a Bayesian joint distribution on the joint space of $\mathscr{D}$ and $\mathscr{M}$: the data vector and a model vector are both interpreted as (finite-dimensional) multivariate random variables, and in order to simulate (or generate) data sets, one can jointly draw random variates to populate the vector elements.

The model space $\mathscr{M}$ is in general a discrete-continuous mixed space, meaning that its elements form a discrete-continuous mixed set; it follows that one could in principle identify each element as a model. Ultimately, however, one can distinguish continuous subsets of $\mathscr{M}$ as models, which is perhaps more common. To formalize the framework that we are working in further, let us define a discrete-continuous mixed space $\mathscr{M}$ as a union $\bigcup_{m \in \mathscr{F}} \mathcal{M}_m$, where $\mathscr{F} \subset \mathbb{N}$ is the space of a discrete parameter m—a flag or label—and where each $\mathcal{M}_m$ may be considered as a model within $\mathscr{M}$. Let each model $\mathcal{M}_m$ define a continuous space $\Theta_m \subseteq \mathbb{R}^{n_m}$ such that $\mathscr{M}$ defines a space given by the union of Cartesian products

$$\Theta := \bigcup_{m \in \mathscr{F}} (\{m\} \times \Theta_m). \tag{1}$$

The number of dimensions of the space Θ_m is given by n_m; a parameter vector $\boldsymbol{\theta} \in \Theta_m \subseteq \mathbb{R}^{n_m}$ then has n_m elements. We thus consider each space in the union $\mathscr{M}$ as a model: unless explicitly stated otherwise, we hereafter consider a model as an element in a discrete set, where each element has an associated continuous parameter space Θ_m (among other constructs, e.g., probability measures such as a joint prior density distribution).

Working within the scope of a given model $\mathcal{M}_m$ with continuous parameter space Θ_m, the prior support $\mathcal{S}_m \subseteq \Theta_m$ is a (compact) subset of $\mathbb{R}^{n_m}$ on which the joint prior density is finite. The target distribution is the joint posterior density distribution, denoted by $\pi(\boldsymbol{\theta} \mid \boldsymbol{d}, \mathcal{M}_m)$, which is related to the joint prior density distribution, $p(\boldsymbol{\theta} \mid \mathcal{M}_m)$, via the probability identity (Bayes' theorem)

$$\begin{aligned} \pi(\boldsymbol{\theta} \mid \boldsymbol{d}, \mathcal{M}_m) p(\boldsymbol{d} \mid \mathcal{M}_m) &= p(\boldsymbol{\theta}, \boldsymbol{d} \mid \mathcal{M}_m) \\ &= p(\boldsymbol{d} \mid \boldsymbol{\theta}, \mathcal{M}_m) p(\boldsymbol{\theta} \mid \mathcal{M}_m), \end{aligned} \tag{2}$$

where $p(\boldsymbol{d} \mid \boldsymbol{\theta}, \mathcal{M}_m)$ is the likelihood function for model $\mathcal{M}_m$. The likelihood function is the sampling distribution on the space $\mathscr{D}$ evaluated at the fixed data vector $\boldsymbol{d}$, as a function of $\boldsymbol{\theta}$, conditional on model $\mathcal{M}_m$; the normalization of the posterior is

the prior predictive probability[30] $p(\boldsymbol{d} \mid \mathcal{M}_m)$ of the data conditional on model $\mathcal{M}_m$. We document the techniques implemented for posterior computation in Appendix A. It is necessarily the case that for construction of a generative model, a joint prior distribution injects a finite quantity of information about the elements of the model space. From a Bayesian perspective, models and parameters are random variables, whose joint prior distribution strictly encodes the information available before acquisition of—and computation given—data. Hereafter we do not need to typeset the symbol Θ (with or without a subscript) to represent a space, and thus we explicitly free it for a different use.

The model space that we consider is some mixture of $\mathscr{M}$-complete and $\mathscr{M}$-open (Vehtari & Ojanen 2012): we do not believe the true data-generating process exists in $\mathscr{M}$, nor do we believe that an element in $\mathscr{M}$ is the closest approximation (to the true data-generating process) that is achievable and tractable (in the future) given more resources (i.e., both cognition and computing time). However, we proceed as though we can imagine a true data-generating process to physically exist—i.e., is plausibly the product of simple physical laws and initial conditions—and that parameterized approximations can in principle approach the real process arbitrarily closely, at least in terms of predictive performance. Moreover, we view the model space for this specific work as effectively the best available to us given current resources, and consider it plausible that it is sufficiently rich to predict features in the procured X-ray event data. A subset of continuous parameters are shared by all models in the discrete set: these are considered to be of an (effectively) fundamental physical nature, such as the distance to a source system and parameters controlling the exterior spacetime solution of a compact star in general relativistic gravity. On the other hand, parameters are also defined that are of an overtly phenomenological nature and are intended to approximate reality only insofar as: (i) predicting—via interplay with the (almost) fundamental parameters—features observed in data, and (ii) supporting future development of physics models that are considered more realistic but crucially remain tractable for statistical falsification.

When we increment model complexity, it is intended to be in an intuitive and natural manner: we generally do so by breaking a form of symmetry, forming nested relationships between models (see Section 2.5 and then Section 2.5.7). Nevertheless, the form of the simplest model and the increments in complexity are ultimately of a subjective nature and exhibit a degree of arbitrariness—different practitioners would have defined their model spaces differently.

The discrete set of models that we condition on was determined partly out of consideration for available resources. Given that posterior computation cost generally increases with prior predictive complexity,[31] we consider it justified to organize a set of increasingly costly problems wherein each is an extension of the former. The salient advantage in this respect is facilitating robustness: by monitoring and analyzing sampling processes operating on gradually more difficult problems, tractability can be gauged for subsequent problems given available computing resources. Moreover, if a pair of models form an exact or approximate nested relationship then we expect the posterior parameter inferences to be consistent with that of the simpler model if the additional complexity is unhelpful.

Aside from resource management, and conditioned on the assumption that posterior computation is sufficiently accurate, one can pose the question of how much complexity is useful to capture structure in observational data. In a Bayesian framework one can in principle estimate prior predictive probabilities of data conditional on a model. It is generally argued that Occam's razor is inherent to prior predictive probability integrals: predictive complexity is penalized if predictions are not *expected* to be at least as commensurate with the data as a simpler (or nested) model. The interpretation of prior predictive probabilities is often fraught with problems, principally sensitivity to prior definition. In this work the joint prior distribution defined for the continuous parameters of each model is not rigorously chosen according to an information-theoretic criterion, nor to accurately quantify belief for overtly phenomenological parameters. However, the prior choices are viewed as being weakly informative[32] for most parameters of interest in the absence of existing constraints, and are viewed as being consistent between models.

As acknowledged above, a widely held view is that evidence estimation[33] does not solve the problem of model comparison. In order to evaluate model performance we thus employ both a form of graphical posterior-predictive checking, and prior predictive probabilities that hereafter we will refer to simply as *evidences*—a less accurate but canonical descriptor of $p(\boldsymbol{d} \mid \mathcal{M})$. If the evidence increases, it is generally accurate to conclude that additional complexity is warranted; if evidence does not increase, however, graphical posterior predictive checking on local modes is useful for determining whether or not facets of the higher-complexity model are a promising avenue to pursue in model development—i.e., if the likelihood function maxima are larger.

2.4. Overarching Definitions

In this section we describe model aspects that are generally shared between all models in the discrete set $\mathscr{M}$. These model facets are in some cases described in detail elsewhere in the *NICER* literature (Bogdanov et al. 2019a, 2019b, 2019c) and thus we are brief where possible. Due to the large number of symbols required to describe the models in this Letter, the symbols used to describe geometric variables shared with Bogdanov et al. (2019b) are different; Table 1 provides symbol translation from Bogdanov et al. (2019b), the theory in which underpins the present work.

2.4.1. Source

The pulsed sources are assumed to be thermally emitting, rotating hot surface regions of PSR J0030+0451.

Parameterization. The exterior spacetime solution is approximated as follows: we embed in each temporal hyperslice of an ambient Schwarzschild spacetime, a (quasi-universal) oblate two-surface, such that the geometric center coincides with the origin of the Schwarzschild coordinate chart (Morsink et al. 2007). The coordinate equatorial radius is

[30] The scalar expectation of the likelihood function with respect to the prior $p(\boldsymbol{\theta} \mid \mathcal{M}_m)$, commonly referred to as the evidence or fully marginal likelihood.

[31] Which is easily proven for (non-dynamic) nested sampling processes (see Appendix A for an outline of the sampling procedure).

[32] Also known as *vague* or *diffuse*.

[33] For calculation of Bayes' factors.

Table 1
Translation of Symbols for Angle Variables Typeset in *both* this Work and Bogdanov et al. (2019b)

Symbol	Description	Bogdanov et al. (2019b)
i	Earth inclination to pulsar rotation axis	ζ
Θ	colatitude of center of a circular hot spot[a]	θ_c
ζ, ψ	angular radius of a circular hot spot[b]	$\theta_{\rm spot}$
ϕ	pulsar rotational phase[c]	ϕ

Notes.
[a] We also use the symbol Θ, with subscripts, to denote the colatitudes of hot regions whose shapes are more complex than circular spots.
[b] We also use these symbols to parameterize hot regions with more complex shapes, such as rings with outer angular radius ζ and inner angular radius ψ.
[c] We also use the symbol ϕ, with subscripts, to denote the azimuthal coordinates of hot regions.

denoted by $R_{\rm eq}$, and the circumference of the equator is $2\pi R_{\rm eq}$; the total mass in the ambient spacetime is denoted by M. The polar axis of the Schwarzschild chart is defined as the pulsar rotational axis. Rotational deformation of the metric away from spherical symmetry is neglected; the current dipole and mass quadrupole moments of the exterior metric enter at first- and second-order in dimensionless angular velocity $\bar{\Omega}$ (Hartle 1967) but are sufficiently small—regardless of the EOS—that they can be neglected in pursuit of a tractable likelihood function (Morsink et al. 2007), especially during initial modeling and in a context where we do not expect the likelihood function to be sensitive to the finiteness of these moments. The perturbations, at constant baryon number, to both the total mass and circumferential radius of a nonrotating star are second-order in $\bar{\Omega}$ (Hartle 1967), and are thus small for a spin of 205 Hz (see also Section 4.2) but are implicitly accounted for. The perturbation to the polar coordinate radius is second-order in $\bar{\Omega}$, and the surface oblateness is controlled by an EOS-insensitive constraint equation (as is the effective gravity along the surface; Morsink et al. 2007). These small changes in the shape of the stellar cross-section induce tilt to the surface, which affects the rays that connect spacetime events at the surface to an observer; together with the change in the projected surface area of a tilted surface, and the change in effective gravity, the effect on light-curves manifests at first-order in $\bar{\Omega}$. Crucially, there is a performance floor for light-curve integration demarcated by a spherically symmetric exterior spacetime solution: embedding an oblate surface in such an ambient spacetime results in negligible increase in computation time per call to a light-curve integrator. We therefore are not concerned about quantifying the difference in our statistical inferences due to inclusion of oblateness over a spherical surface—more resources are required to quantify this rigorously than to simply account for oblateness.

A distant, static, and fictitious[34] instrument (see Section 2.4.2) is located at radial coordinate D, and subtends colatitude—hereafter termed inclination—denoted by i. Interstellar light–matter interaction is described by absorption within a column of material. The attenuation factor is parameterized solely by the column density $N_{\rm H}$ of neutral hydrogen and we assume relative abundances for the interstellar medium from Wilms et al. (2000); we implement the `tbnew`[35] model to precompute a set of lookup tables for attenuation as a function of photon energy.

In each model the spatial dependence of the surface radiation field is of a phenomenological nature: the aim is to introduce sufficient complexity so as to represent the basic notion of pulsar surface heating due to energy deposition by magnetospheric currents (in the vicinity of the magnetic poles). We are largely ignorant of spatial structure in the surface radiation field because the star is not spatially resolved; moreover, it is intractable for us to consider more self-consistent numerical models of the surface radiation field, in part due to the expense of statistical computation. In the simplest case the radiation field is constructed by filling two closed simply connected regions on the surface—which do not mutually overlap—with radiating material;[36] these regions may be interpreted to each result from magnetospheric polar cap heating. We only compute a radiative signal from these hot regions, and therefore in the context of each of our models, *hot region* can be viewed as synonymous with *radiating region*.

For all models a geometrically thin (and thus plane-parallel) fully ionized hydrogen `NSX` atmosphere is invoked for the radiating material (Ho & Lai 2001; Ho & Heinke 2009). The radiation field is precomputed and represented as a lookup table for cubic polynomial interpolation of specific intensity, $I_E/k_{\rm B}T_{\rm eff}^3$, with respect to four variables defined in a surface local comoving frame: effective temperature, $T_{\rm eff}$; effective gravity; photon energy, $E/k_{\rm B}T_{\rm eff}$; and the cosine of the ray zenith angle (to surface normal). A quasi-universal relation for surface effective gravity[37] is adopted from AlGendy & Morsink (2014) in order to evaluate local radiation field intensities. We do not explicitly compute emission from the stellar surface exterior to the closed regions (i.e., as a function of source parameters controlling the exterior radiation field),[38] but the phase-invariant background model that we invoke in all cases

[34] A notion borne from the nature of the event-data pre-processing (see, e.g., Riley & Watts 2019, for details pertaining to *X-PSI*).
[35] https://pulsar.sternwarte.uni-erlangen.de/wilms/research/tbabs/

[36] Note that although a cooler radiating hydrogen atmosphere should exist globally over the stellar surface (as observed for PSR J0437−4715, Durant et al. 2012; Guillot et al. 2016; Gonzalez-Caniulef et al. 2019, and PSR J2124−3358, Rangelov et al. 2017), we make no explicit reference to it when defining our likelihood function—i.e., we do not compute any radiative signal from the atmosphere exterior to the closed regions. The atmosphere cannot be globally uniform because local heating by magnetospheric currents will affect the local temperature and ionization degree; effective gravity also varies due to rotation. Reference to the global atmosphere is implicit due to the fluid properties required for containment of hot material in the closed regions.
[37] Equatorially reflection-symmetric.
[38] Such computation is supported by *X-PSI* with specialization to ensure (almost) exact areas as described above (see also Bogdanov et al. 2019b), but requires a choice of surface radiation field (e.g., atmosphere ionization degree and chemical composition). We therefore opt to capture the *non-pulsed* fraction of emission from the stellar surface exterior to the regions via our default background treatment. If evidence for unmodeled soft pulsed emission arises a posteriori one could then consider explicit computation of such emission.

can capture non-pulsed components of surface and off-surface emission (see Section 2.4.3).

Let us hereafter refer to the geometric configuration of the infinitesimal radiating surface elements simply as the *shape* of a hot region—including both exterior and interior boundaries. In more complex models each hot region is constructed using additional shape parameters and in some cases a second temperature component. We consider the shape and temperature of a hot region to result from the interaction of *two* closed regions—hereafter *members*. The members partially or wholly[39] overlap, and one member takes precedence when evaluating local radiation intensities along rays (null geodesics) that connect spacetime events on the rotating surface to a distant observer. We term one member as *ceding*, and the other as *superseding*. As discussed in the appendix of Bogdanov et al. (2019b), the *X-PSI* implementation (referred to as the AMS code in Bogdanov et al. (2019b)) of a radiating region is specialized for fast likelihood function evaluation when said function is a callback for sampling processes: while numerical approximations are necessary in general, it is relatively inexpensive to ensure that the proper area of each finite-element (discretized) radiating region is computed to a precision that (almost) exactly matches their mathematical definition.[40] When two members overlap to form a hot region (i) the area of a discretely represented superseding member is (almost) exact; and (ii) the area the discretely represented, *non-superseded subset* of a ceding member is also (almost) exact. The subset that is not superseded can itself be simply connected or non-simply connected depending on the model (where the set of configurations assigned finite prior density is model dependent). For the precise details of the hot regions, refer forward to Sections 2.5.1 through 2.5.7.

We opt for two disjoint hot regions: the two distinct pulses visible in the phase-folded event data (Figure 1) are suggestive of two such regions being widely separated. Initially, we impose parity in the complexity of each hot region—i.e., an equal number of shape parameters and temperature components. We then consider models in which the hot regions have unequal complexities because it becomes clear that increasing the complexity of a particular hot region is unwarranted. In all cases we define the support of the joint prior so as to exclude limiting configurations in which the hot regions overlap; the reason being that extension of scope to such configurations requires specification and implementation[41] of additional logical conditions for a complete order of precedence in local radiation intensity evaluation. For a subset of models we impose antipodal symmetry of the hot regions, in order to crudely represent a heating distribution that is consistent with symmetry in the physical mechanisms driving surface X-ray emission, such as a dominantly or perfectly centered-dipolar field configuration. A magnetic field with finite higher-order structure is viewed as a closer approximation of physical reality; we represent this case crudely by breaking antipodal symmetry and defining additional parameters for a secondary hot region that are not derived from parameters of a primary hot region. However, self-consistent coupling of the magnetosphere to the surface radiation field is beyond the scope of this work, partly due to the associated increase in complexity of efficient model implementation for posterior computation.

We consider models with three or more mutually disjoint and separated radiating regions to be a logical extension of the model space if it is deemed that incrementally increasing the complexity of only two such regions is yielding insufficient advancement in posterior predictive performance—i.e., is not satisfactorily capturing observed structure in data for the resource expenditure—when approaching or extrapolating to the limit of what is considered computationally tractable by a group executing posterior computation. A salient advantage of such an approach is that it is more exhaustive with ideas for two hot regions and incrementally breaks symmetries; an apparent disadvantage is that some small set of conceivable closely related models with equal (continuous parameter) dimensionality are not applied.[42] It is, however, necessary to be selective—inherent to which is subjectivity and arbitrariness.

The above choices for surface radiation field configuration are somewhat consistent with the notion of the source being a rotation-powered X-ray pulsar with two relatively small hot regions that are disjoint. Therefore, we consider the proposed model space to be logically structured and a reasonable representation of widely held conceptions of such stars that are yet to be falsified statistically.

Whether or not such a model space is tractable given algorithm properties and computing resources is highly sensitive to the choice of parameterization for posterior computation, especially in phenomenological contexts. If degeneracies plague the problem at hand, it may be considered an indicator that either (i) the model is simply ill-defined, leading to forms of invariance of the parameterized sampling distribution on the space of the data; (ii) the model is needlessly or at least *unhelpfully* complex for describing observations, because despite (physical) parameters having a finite effect on forward data-generation, one is ultimately insensitive to such model structure. Usually this equates—at least in part—to transforming away nonlinear likelihood function degeneracies where possible. A number of sophisticated open-source sampling software packages efficiently handle linear degeneracies, even in multi-modal contexts, but nonlinear degeneracy remains fiendish: certain sampling algorithms can perform accurately,[43] but coupled with an expensive (numerical) likelihood function and moderately high-dimensional sampling spaces, still require massive computing resources.

As we highlight in Section 2.5 (where we provide more precise definitions of surface radiation field structure), the choice of parameterization of a (largely phenomenological) hot region plays a crucial role in sampling-space definition. In Appendix A we summarize the techniques adopted for posterior computation: we opt to perform nested sampling (Skilling 2006). The natural space for nested sampling is usually that of a unit hypercube, which maps to an equal-dimensional physical parameter space according to an inverse transformation derived from a joint (prior) probability distribution on the physical space (e.g., Feroz et al. 2009). Given that nested sampling algorithms tend to operate in such a native space, a parameterization that approaches optimality involves

[39] Such that one member is a superset of the other.

[40] The overall numerical accuracy remains implementation dependent.

[41] Efficient finite-element representations.

[42] For example, configurations in which three single-temperature regions are disjoint.

[43] At reduced efficiency relative to simpler contexts—see the MultiNest sampler cited in Appendix A.

both the physical parameterization *and* the inverse transformation from the native to physical space, and is such that continuous likelihood function degeneracies, if existent, manifest effectively linearly in the native space.

Usually a joint prior distribution is chosen to be weakly informative—or "flat"—in the context of the likelihood function, and in some cases is defined as an absolutely flat density function with respect to a joint space. It follows that the mapping from the parameter space to the native sampling space will then approximately preserve the linear degeneracy of a posterior mode. An exception to this occurs if the boundary of the support of the joint prior satisfies some set of non-trivial constraint equations.[44]

Priors. We define the joint prior density distribution $p(M, R_{\rm eq})$ to be *jointly* flat with compact support: a prophylactic choice that eases future use of samples on the $(M, R_{\rm eq})$-subspace for computing an approximative marginal likelihood function, which in turn can be used for estimation of interior source-matter properties—principally EOS parameters (see Riley et al. 2018; Riley & Watts 2019). We also choose the boundary of prior support to be close to maximally inclusive in regards to theoretical EOS predictions: we impose hard bounds $M \in [1,\ 3]\,M_\odot$, and impose that $R_{\rm eq} \in [3r_g,\ 16]$ km, where $r_g = r_g(M) = GM/c^2$ is the gravitational radius.

Although we allow the boundary of the prior support to extend down to the photon sphere of the ambient spacetime solution, when computing pulse profiles we only integrate over the *primary* images (along rays with angular deflection $\leqslant\pi$) of radiating elements subtended on the sky of the instrument. For a spherical star of radius $R < 3.52r_g(M)$, multiple images of parts of the star will be visible (Pechenick et al. 1983; Bogdanov et al. 2019b) requiring that light from the primary, secondary, and higher-order images be included—at additional computational expense—in an exact calculation of the flux; this issue, and how it pertains to oblate stars, is discussed in more detail elsewhere (Bogdanov et al. 2019b). From a computational statistics perspective, one could view the inclusion of one or more higher-order images as a modeling refinement to be made if, a posteriori, a rotating star is favored to be sufficiently compact: e.g., a substantial fraction of posterior mass lies at $R_{\rm eq}/r_g(M) \lesssim 3.6$. The parameter inference reported in this work favors much less compact stars a posteriori (refer forward to Figure 19), so multiple imaging is not deemed important. However, when images are neglected, the issue of choosing the most appropriate support for a joint prior distribution of M and $R_{\rm eq}$ remains an open problem for statistical modeling.

A typical likelihood function for pulse-profile modeling will express many modes of dependence on the compactness $r_g(M)/R_{\rm eq}$, and will generally be more sensitive to this combination than to M (or $R_{\rm eq}$) individually. We ensure the mapping from the parameter space to the native space preserves such linear degeneracy between M and $R_{\rm eq}$. In Appendix B we provide implementation details for the joint density $p(M, R_{\rm eq})$.

Finally, there exists a constraint on the distance of the PSR J0030+0451 system (Arzoumanian et al. 2018) that we adopt—in approximation—as an informative prior.

[44] An example of trivial constraint equations here are those that generate a (hyper-)rectangular support boundary in parameter space.

2.4.2. Instrument

In defining a generative model, the data space is constructed by phase-folding X-ray events in each detector channel and grouping those events into a uniform set of phase intervals (bins) to curate a set *count* numbers, typically with cardinality $\mathcal{O}(10^3)$. The conditional joint sampling distribution of these count numbers is always constructed in terms of a phase-energy-resolved signal that is generated during a *single* rotation of the source. Together with an appropriate nuisance background model (see Section 2.4.3), it follows that the instrument in such an analysis represents the temporal-mean operation of all detectors collectively in response to the incident radiation field from the source during the observation time intervals. We reserve a more elaborate discussion on these modeling facets for Section 2.4.3 (and also refer the reader to Riley & Watts 2019).

For every model we invoke the instrument response model: the on-axis `v1.02` ancillary response function (ARF) and an updated version of the `v1.02` redistribution matrix file (RMF; private communication from James Steiner, see Hamaguchi et al. 2019, for details of the updates) to generate a reference (or nominal) response matrix $\mathcal{R}^\star$ derived from microphysical knowledge. Let detector channels increment with row number i, and energy intervals increment with column number j, such that an element of the reference response matrix is denoted by $\mathcal{R}^\star_{ij}$. We use this reference matrix as a basis for a parameterized family of response matrices, and aim to compute (for each model) a joint posterior density distribution of continuous source parameters and continuous instrument (response matrix) parameters. It is well founded to parameterize the instrument because despite its synthetic nature, we do not consider its microphysical operation to be sufficiently *known*;[45] nevertheless, we define far fewer parameters for the instrument than for the source. While the following model ensures that operational uncertainty is included, the continuum of response models and the associated prior density distribution does not attempt to rigorously represent uncertainty in microphysical knowledge; given a close approximation to the radiation field incident on the telescope, we would expect the model to be conservative in terms of *prior* predictive performance.

We parameterize the response matrix using a calibration product derived from observations of a calibration source. For this work we use instrumental residuals derived from *NICER* observations of the Crab. These residuals are derived using the observations and following the procedure outlined in Ludlam et al. (2018), modified to use the appropriate number of detectors, ARF, RMF, and Sun-angle cut consistent with the PSR J0030+0451 data set (refer to Section 2.2). We acknowledge that the Crab is a remarkably different source to PSR J0030+0451: the expected operation of *NICER* (and X-ray instruments in general) in response to incident radiation fields is a function of its properties. The Crab exhibits a very different spectrum to PSR J0030+0451, being harder, more absorbed, and subject to astrophysical features; the Crab is also an extended source, not a point source, and is much brighter than the rotation-powered MSPs targeted by *NICER*.

The calibration product is a channel-by-channel vector $\mathscr{R}$ of ratios of observed Crab count numbers to count numbers derived

[45] In-flight astrophysical calibration sources, for instance, are in practice far brighter than science targets, and operation is conditional on the radiation field incident on the detectors.

using a theoretical incident spectrum and the reference response matrix $\mathcal{R}^\star$. Let the vector elements $\mathscr{R}_i := \mathcal{C}_i/(\mathcal{R}^\star \cdot \mathcal{F})_i$: in words, the element-wise division of an *observed* count vector $\mathcal{C}$ from some calibration source, by a vector $\mathcal{R}^\star \cdot \mathcal{F}$ where $\mathcal{F}$ is a vector of photon *fluences* (in the set of energy intervals inherent to the definition of $\mathcal{R}^\star$) computed given some theoretical model of said calibration source during the calibration observations. Note that the calibration product is derived from observations of a single chromatic source and thus is not resolved over elements of the matrix, only over the set of channels; therefore we apply the ratio $\mathscr{R}_i$ for the ith channel to all elements $\mathcal{R}^\star_{ij}$. The theoretical model is uncertain in the lowest 10 channels that we consider: the Crab is highly absorbed so that there is less data at low energies, and the telescopes used to generate the reference spectra for the residuals also perform poorly in this regime. In this work we therefore assume that $\mathscr{R}_i$ for $i \in [25, 35)$ is equal to $\mathscr{R}_{35}$.

We choose to construct the response matrix as a continuous three-parameter family, where the parameters are denoted by *NICER* α, *NICER* β, and *NICER* γ where possible, but reduced to the aliases α, β, and γ, respectively, for clarity of mathematical expressions. First we give the definition, and then we offer an interpretation in words. The parameterized matrix is defined as

$$\begin{aligned}\mathcal{R}_{ij}(\alpha, \beta, \gamma; \mathscr{R}, \mathcal{R}^\star) &:= (1-\beta)\gamma\mathcal{R}^\star_{ij} + \beta\alpha\mathscr{R}_i\mathcal{R}^\star_{ij} \\ &\equiv \gamma\mathcal{R}^\star_{ij} + \beta\underbrace{[\alpha\mathscr{R}_i\mathcal{R}^\star_{ij} - \gamma\mathcal{R}^\star_{ij}]}_{\text{calibration shift}} \\ &\equiv \mathcal{R}^\star_{ij}\underbrace{[\gamma + \beta(\alpha\mathscr{R}_i - \gamma)]}_{\eta_i}. \end{aligned} \tag{3}$$

In general an additional condition must be invoked: $\mathcal{R}_{ij}(\alpha, \beta, \gamma) := 0$ if $[(1-\beta)\gamma\mathcal{R}^\star_{ij} + \beta\alpha\mathscr{R}_i\mathcal{R}^\star_{ij}] \leqslant 0$ for $(\alpha, \beta, \gamma) \in \mathcal{S}_{(\alpha,\beta,\gamma)}$, where $\mathcal{S}_{(\alpha,\beta,\gamma)}$ is the joint prior support on the (α, β, γ)-subspace.[46] We note, however, that this parameterization exhibits a finite degree of degeneracy.

The parameter α scales the calibration vector $\mathscr{R}$, and manifests to target the assumption that the vectors $\mathcal{C}$ and $\mathcal{F}$ are *known*: in regards to $\mathcal{C}$, the assumption that the temporal-mean operation of the instrument between calibration and science observations is invariant (including, e.g., pointing vector relative to source line of sight and the flux-dependent effects); and in regards to $\mathcal{F}$, the assumption that the expectation of the incident radiation field from the Crab during the calibration observations is known. Thus, the parameter α represents the product of dimensionless element-invariant scaling factors applied, respectively, to vectors $\mathcal{C}$ and $\mathcal{F}$. The support of α is such that $\mathcal{S}_\alpha = \{\alpha: \alpha \in \mathbb{R}^{>0} \wedge 1 - \epsilon_\alpha \leqslant \alpha \leqslant 1 + \epsilon_\alpha\}$ where $0 < \epsilon_\alpha < 1$; the prior density function allocates mass mostly to the near vicinity of unity.

The parameter γ is an element-invariant scaling factor applied to the reference matrix $\mathcal{R}^\star$ alone. The reason that we choose $\alpha \neq \gamma$ is that γ operates on $\mathcal{R}^\star$, and thus because $\mathscr{R}$ is a function of $\mathcal{R}^\star$ (via the element-wise vector division written above), γ cancels. The support of γ is such that $\mathcal{S}_\gamma = \{\gamma: \gamma \in \mathbb{R}^{>0} \wedge 1 - \epsilon_\gamma \leqslant \gamma \leqslant 1 + \epsilon_\gamma\}$ where $0 < \epsilon_\gamma < 1$; the prior density function allocates mass mostly to the near vicinity of unity.

The parameter β is a weighting factor between matrices with elements $\gamma\mathcal{R}^\star_{ij}$ and $\alpha\mathscr{R}_i\mathcal{R}^\star_{ij}$—i.e., the element-invariant coefficient of the element-dependent calibration shift away from the $\gamma\mathcal{R}^\star$ matrix. Thus β may be interpreted as the degree to which the calibration vector $\mathscr{R}$ is used to modify the reference matrix $\mathcal{R}^\star$. Note that the prior support for a weighting parameter such as β is the unit interval $\beta \in [0, 1]$. We consider the limit $\beta \to 0$ as a useful safeguard against erroneous calibration—e.g., artifacts may be introduced by invoking a calibration source to which the instrument responds appreciably differently than it does to PSR J0030+0451. As $\beta \to 0$, $\mathcal{R}_{ij}(\alpha, \beta, \gamma) \to \gamma\mathcal{R}^\star_{ij}$, capturing a simple element-invariant scaling of $\mathcal{R}^\star$. On the other hand, in the limit $\beta \to 1$, we have $\mathcal{R}_{ij}(\alpha, \beta, \gamma) \to \alpha\mathscr{R}_i\mathcal{R}^\star_{ij}$.

Priors. To demonstrate the properties of the family of response matrices defined above, we (i) define a joint probability density distribution $p(\alpha, \beta, \gamma)$ that could be plausibly viewed as a *weak* prior on its compact support given that the instrument is an artificial system that has been closely studied, and (ii) then visualize the distribution in terms of derived properties (such as total effective area as a function on energy interval). Let $\alpha \sim N(1, 0.1)$ truncated such that $\alpha \in [0.5, 1.5]$; $\beta \sim U(0, 1)$; $\gamma \sim N(1, 0.1)$, truncated such that $\gamma \in [0.5, 1.5]$. In Figure 2 we display the prior distribution of the total on-axis effective area as a function of energy.

2.4.3. Likelihood Functions and Background

In the *X-PSI* documentation (Riley & Watts 2019) we offer a more complete overview of the supported class of generative models than is appropriate for this work—instead we adapt the *X-PSI* documentation to provide a summary. A generative model for raw on-board event data is eschewed by subsuming a non-parameterized portion of the modeling within a data pre-processing phase. In this work, a pulsar radio timing solution is invoked to transform events into a simpler time domain: a fictitious instrument, which is static (or Eulerian) and distant in the (Schwarzschild) spacetime of the source, is implicitly constructed to register events against the elapsed natural number rotations of the star, which is a clock related to the Schwarzschild coordinate time simply by an affine transformation. In all models we condition on the phase-resolved specific flux signal (incident on the instrument) generated by precisely *one* rotation of the star: accurate computation of such a signal, even when invoking spacetime spherical symmetry, is approaching the limit of what we consider tractable at present in terms of likelihood function callback cost for sampling processes in $\mathcal{O}(10)$-dimensional spaces.

In the context of a model for the joint probability distribution of observed events (see Riley & Watts 2019),[47] phase-folding said events is equivalent to computing one (average) rotational pulse and replicating it over the many rotational cycles in order to evaluate the likelihood function. It follows that, in this limit, no information is lost by transforming events to the unit interval because the underlying information content in the model is not summarized for comparison to data. In each instrument channel we choose to group events into a set of uniform-width phase intervals (bins) that are subsets of the unit interval, and define the data space as $\mathscr{D} := \mathbb{N}^{I \times K}$, where $I \in \mathbb{N}$

[46] Alternatively, one might define the joint prior support such that if for $(\alpha, \beta) = (\alpha', \beta', \gamma')$ the inequality is true for any (i, j), then the joint density is locally zero at $(\alpha', \beta', \gamma')$.

[47] Leading to an unbinned likelihood function.

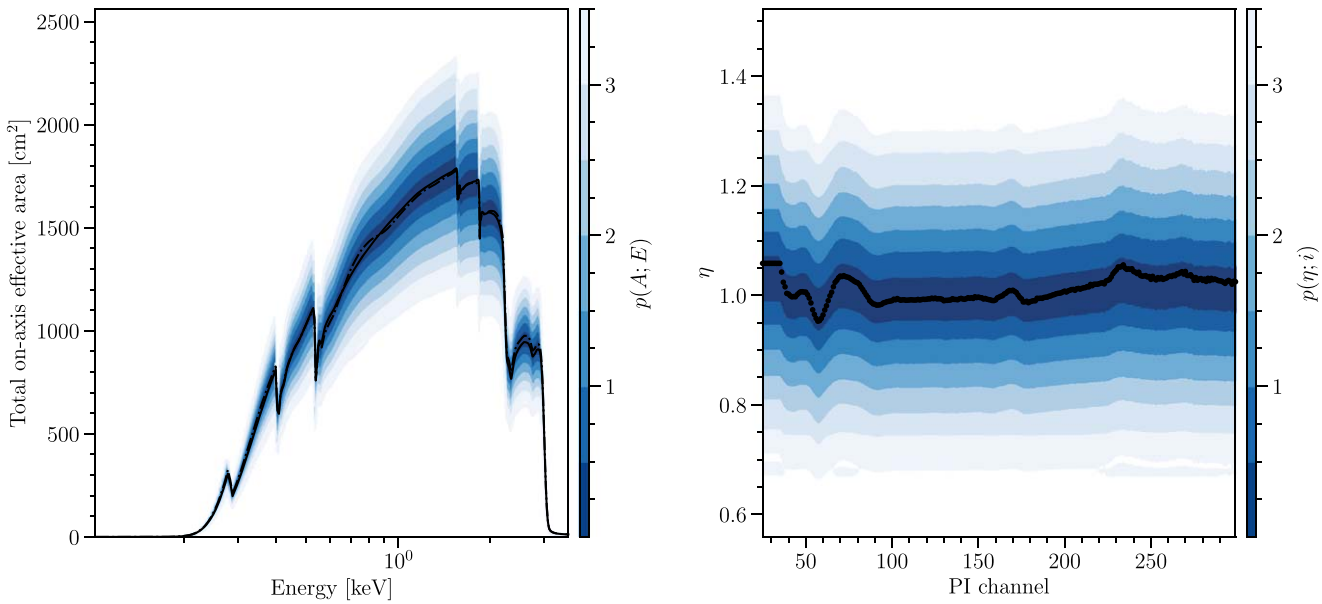

Figure 2. In the left panel we display marginal conditional prior probability density distributions of total on-axis effective area A as a function of energy (integrated over uniform intervals of width 5×10^{-3} keV), summed over the contiguous channel subset [25, 300). Note that the total *NICER* effective area at energies $\gtrsim$3 keV in particular is far greater than shown here. The prior $p(A; E)$ is represented by the blue contours. At energy E a set of (one-dimensional) highest-density credible intervals are estimated for $A = A(\alpha, \beta, \gamma)$; the credible intervals are connected as a function of energy, such that, e.g., the second-darkest band encodes the energy-dependence of the estimated 68.3% highest-density credible interval. We give the marginal prior distributions for the parameters α, β, and γ in Table 2. Note that the posterior information shown is not that of the joint distribution of effective areas over energy intervals: the effective areas are coupled by a functional form with three parameters. The range of the energy intervals is determined based on the RMF of the reference matrix $\mathcal{R}^\star$ and on the curated data set in the contiguous set of channels [25, 300), where these channels are summed over. The solid curve is that of the reference matrix $\mathcal{R}^\star$ (with $\gamma = 1$); the dashed–dotted curve is that of the calibrated matrix $\mathcal{R}_{ij} = \mathscr{R}_i \mathcal{R}^\star_{ij}$ (with $\alpha = 1$). In the right panel we display as black points the elements of the vector $\mathscr{R}$ of multipliers supplied for instrument parameterization based on the Crab as a calibration source. The prior $p(\eta; i)$, where i enumerates channels and η is defined in Equation (3), is represented by the blue contours; the probabilistic information is otherwise congruent in nature to that described for the left panel.

is the number of channels over which the folded events are distributed, and $K \in \mathbb{N}$ is the number of phase intervals; we choose $K = 32$ and $I = 275$ corresponding to channel subset [25, 300). Information loss is an inherent consequence of compression of events into a smaller set of summary quantities, but the phase resolution is sufficiently high here to mitigate our concern about the use of a binned likelihood function instead of an unbinned likelihood function. The conditional joint sampling distribution on the space $\mathscr{D}$ is assumed to be purely Poissonian and separable over channels: the *NICER* instrument exhibits sufficiently high-resolution event-timing capabilities for the Poissonian nature of the incident radiation field to be effectively conserved as an event arrival process in the on-board time domain (see Riley & Watts 2019, for a more explicit set of arguments pertaining to this matter).

Let the likelihood function for phase-folded and binned events be defined by (see also Miller & Lamb 2015)

$$L(\boldsymbol{\theta}, \boldsymbol{B}) \coloneqq p(\{\boldsymbol{d}_i\}_{i=1,\ldots,I} \mid \boldsymbol{\theta}, \{B_i\}) = \prod_{i,k} p(d_{ik} \mid \boldsymbol{\theta}, B_i), \quad (4)$$

where $i \in [1, I]$ enumerates channels of the instrument; each $\boldsymbol{d}_i$ is a data vector associated with the ith channel, constituted by count numbers $\{d_{ik}\}_{k=1,\ldots,K}$ where k enumerates phase intervals $\phi_k \subset [0, 1]$; B_i is the (nuisance) background count-rate parameter in the ith channel; and $\boldsymbol{\theta}$ are the continuous source parameters that constitute the sampling space, and on which source expected count numbers $s_{ik}(\boldsymbol{\theta})$ are dependent. Each B_i is defined as the expectation of a homogeneous—i.e., time invariant—Poisson arrival process: up to a known constant,

$$\ln L(\boldsymbol{\theta}, \boldsymbol{B}) = -2\sum_{i,k} d_{ik} \ln\left[s_{ik}(\boldsymbol{\theta}) + B_i\left(\sum_\ell \Delta t_\ell\right)\int_{\phi_k} d\phi\right] - s_{ik}(\boldsymbol{\theta}) - B_i\left(\sum_\ell \Delta t_\ell\right)\int_{\phi_k} d\phi. \quad (5)$$

Such background treatment is the default for *NICER* parameter estimation work and *X-PSI* (for implementation details see Appendix B of Riley & Watts 2019), and is based on Miller & Lamb (2015). In reality, the statistical properties of backgrounds do not exhibit time invariance, but exhibit long-term variation over the $\mathcal{O}(1)$ year observation epoch of MSPs target by the *NICER* mission, especially due to factors such as dynamical space weather (Bogdanov et al. 2019a). However, any background emission processes—and dynamical emission processes in the local vicinity of PSR J0030+0451—that are not harmonically coupled to the surface X-ray emission will *decohere* over the unit phase interval whose boundary is periodic. It follows that event phase-folding enables invocation of phase-invariant channel-by-channel (background) count-rate terms.

The background below $\sim$3 keV for *NICER* observations of PSR J0030+0451 observations consists of (Bogdanov et al. 2019a) (i) cosmic energetic particle events and diffuse X-ray emission over the $\sim$30 arcmin^2 field of view (Arzoumanian et al. 2014); (ii) many nearby X-ray point sources in the field

that make a small total contribution relative to the targeted MSP; and (iii) solar system contamination, including optical loading (pointing Sun-angle dependent), and high-energy non-cosmic particles and radiation. Considering a proper subset of detector channels and filtering background events during the pre-processing phase acts to reduce background contribution, but some subset of background events survive and must be modeled (Bogdanov et al. 2019a).

The source terms $s_{ik}(\boldsymbol{\theta})$ in Equation (5) are then derived as follows. Let $F(\phi, E; \boldsymbol{\theta})$ denote the incident specific photon flux from the source as a function of rotational phase ϕ. The function $F(\phi, E; \boldsymbol{\theta})$ is evaluated numerically at a regular discrete set of points in the joint space of energy and phase, as an approximating two-dimensional integral over the solid angle of the image of the source subtended on the sky of a point in the vicinity of the distant static instrument; given the discrete representation, a continuous representation is constructed via spline interpolation in *X-PSI*.

Let the symbol $\mathcal{R}_{\ell ij}(\hat{r})$ denote a temporal-mean point-source response matrix invoked for the ℓth observing interval, which is dependent on the radial coordinate unit vector $\hat{r}$ in the Schwarzschild chart at the location of the instrument—i.e., the pointing of the telescope relative to the source. The response matrix with elements

$$\mathcal{R}_{ij}(\hat{r}) = \sum_{\ell} \frac{\Delta t_{\ell}}{\left(\sum_{\ell} \Delta t_{\ell}\right)} \mathcal{R}_{\ell ij}(\hat{r}) := \mathcal{R}_{ij}(\alpha, \beta, \gamma) \tag{6}$$

is the exposure-time-weighted mean response matrix that is modeled as the matrix defined in Equation (3) and Section 2.4.2.

The source contribution to the expected number of counts is given by

$$\begin{aligned} s_{ik}(\boldsymbol{\theta}) &= \textstyle\sum_{\ell,j} \mathcal{R}_{\ell ij}(\hat{r}) \Delta t_{\ell} \int_{\phi_k \times \boldsymbol{E}_j} F(\phi, E; \boldsymbol{\theta}) dE d\phi \\ &= \textstyle\sum_j \left[\left(\sum_{\ell} \Delta t_{\ell}\right) \underbrace{\int_{\phi_k \times \boldsymbol{E}_j} F(\phi, E; \boldsymbol{\theta}) dE d\phi}_{\boldsymbol{F}_{jk}(\boldsymbol{\theta})} \right] \\ &\quad \times \textstyle\sum_{\ell} \frac{\Delta t_{\ell}}{\left(\sum_{\ell} \Delta t_{\ell}\right)} \mathcal{R}_{\ell ij}(\hat{r}). \end{aligned} \tag{7}$$

A matrix of source count numbers may thus be evaluated as the dot-product[48]

$$\boldsymbol{s}(\boldsymbol{\theta}) = \left(\sum_{\ell} \Delta t_{\ell}\right) \mathcal{R}(\alpha, \beta, \gamma) \cdot \boldsymbol{F}(\boldsymbol{\theta}), \tag{8}$$

where $\boldsymbol{F}(\boldsymbol{\theta})$ is a matrix of phase-integrated incident photon fluxes.[49] Note that the elements of the matrix $\boldsymbol{F}(\boldsymbol{\theta})$ may be approximated using instantaneous fluxes at points *within* the finite phase intervals instead of explicitly integrating over those intervals, provided that the intervals are determined to be sufficiently small.

We numerically marginalize the likelihood function given by Equation (5) over the subspace of nuisance background parameters $\boldsymbol{B}$ in order to improve tractability of the sampling process. The target distribution (the posterior) for sampling is written conditional on model $\mathcal{M} \subset \mathscr{M}$ as

$$\begin{aligned} \pi(\boldsymbol{\theta} \mid \boldsymbol{d}, \mathcal{M}) = \int \pi(\boldsymbol{\theta}, \boldsymbol{B} \mid \boldsymbol{d}, \mathcal{M}) d\boldsymbol{B} \propto p(\boldsymbol{\theta} \mid \mathcal{M}) \\ \times \underbrace{\int L(\boldsymbol{\theta}, \boldsymbol{B}) p(\boldsymbol{B} \mid \mathcal{M}) d\boldsymbol{B}}_{L(\boldsymbol{\theta})}, \end{aligned} \tag{9}$$

where $L(\boldsymbol{\theta}) = p(\boldsymbol{d} \mid \boldsymbol{\theta}, \mathcal{M})$ is the marginal likelihood function supplied as a callback for a sampling process, and the joint prior density distribution $p(\boldsymbol{\theta}, \boldsymbol{B} \mid \mathcal{M})$ is separable with respect to $\boldsymbol{\theta}$ and $\boldsymbol{B}$. The joint prior distribution $p(\boldsymbol{B} \mid \mathcal{M})$ is equivalent for all models: jointly flat and separable. Crucially, such a phenomenological background model exhibits a large prior complexity; for instance, the joint density at a background count rate vector $\boldsymbol{B}$, where the variation between channels is always small relative to the limiting instrument count rate, is equivalent to the joint density at a vector $\boldsymbol{B}$ whose elements exhibit vast channel-to-channel variations. It follows that a joint flat, separable prior is not considered representative of our prior belief. However, we consider—without proof—the prior to be weakly informative because (i) the conditional likelihood function (given a fixed source vector) exhibits a large curvature relative to the prior density function; (ii) the source photon flux signal always has few extrema in the joint space of energy and phase; and (iii) a posteriori the conditional likelihood function maxima do not wildly fluctuate as a function of channel because such structure does not exist in the data set. Nevertheless, if one were to compare, based on prior predictive performance, the models that we consider in this work with a model invoking a background component with far lower complexity, one should not be surprised if the former are strongly disfavored.

In general, the support of $p(\boldsymbol{B} \mid \mathcal{M})$ is compact and bounds can be specified on a channel-by-channel basis to truncate the marginalization integrals. Lower bounds may be derived, for example, from calibration observations of nearby fields that exclude the PSR J0030+0451 and are otherwise devoid of bright sources. Upper limits may, for example, be based on distinct *NICER* observations of the field containing PSR J0030+0451.

For this work, however, we define the lower bound as zero for each channel, and we eschew definition of an upper bound in each channel because the posterior is considered integrable: non-diverging on joint compact support, and the conditional likelihood function—$L(\boldsymbol{\theta}, \boldsymbol{B})$ for fixed $\boldsymbol{\theta}$—asymptotes to zero at large background count rates. If a set of sufficiently high upper bounds *were* specified (e.g., based on *NICER* count rate limits), the associated normalizing constant for the joint prior, equal to the reciprocal of the products of those bounds, would not modulate relative probability measures defined on $\mathscr{M}$.[50] We therefore do not view the improperness of the above prior as a misdemeanor, but it does mean that we should not describe our model as *generative* in the strictest sense.

[48] An approximation in many ways, one being the discrete representation of an instrument that responds in a continuous manner to input.

[49] In the *X-PSI* implementation phase integration is performed using splines after a dot-product operation on a matrix of instantaneous incident photon fluxes.

[50] The raw event count rate during the exposures used to curate the data set for this work is known to be far below limiting and thus the conditional likelihood function is always relatively small for near-limiting background count rates. A set of upper bounds defined in this limit therefore truncates the evidence integral in a regime where model-dependent sensitivity is negligible.

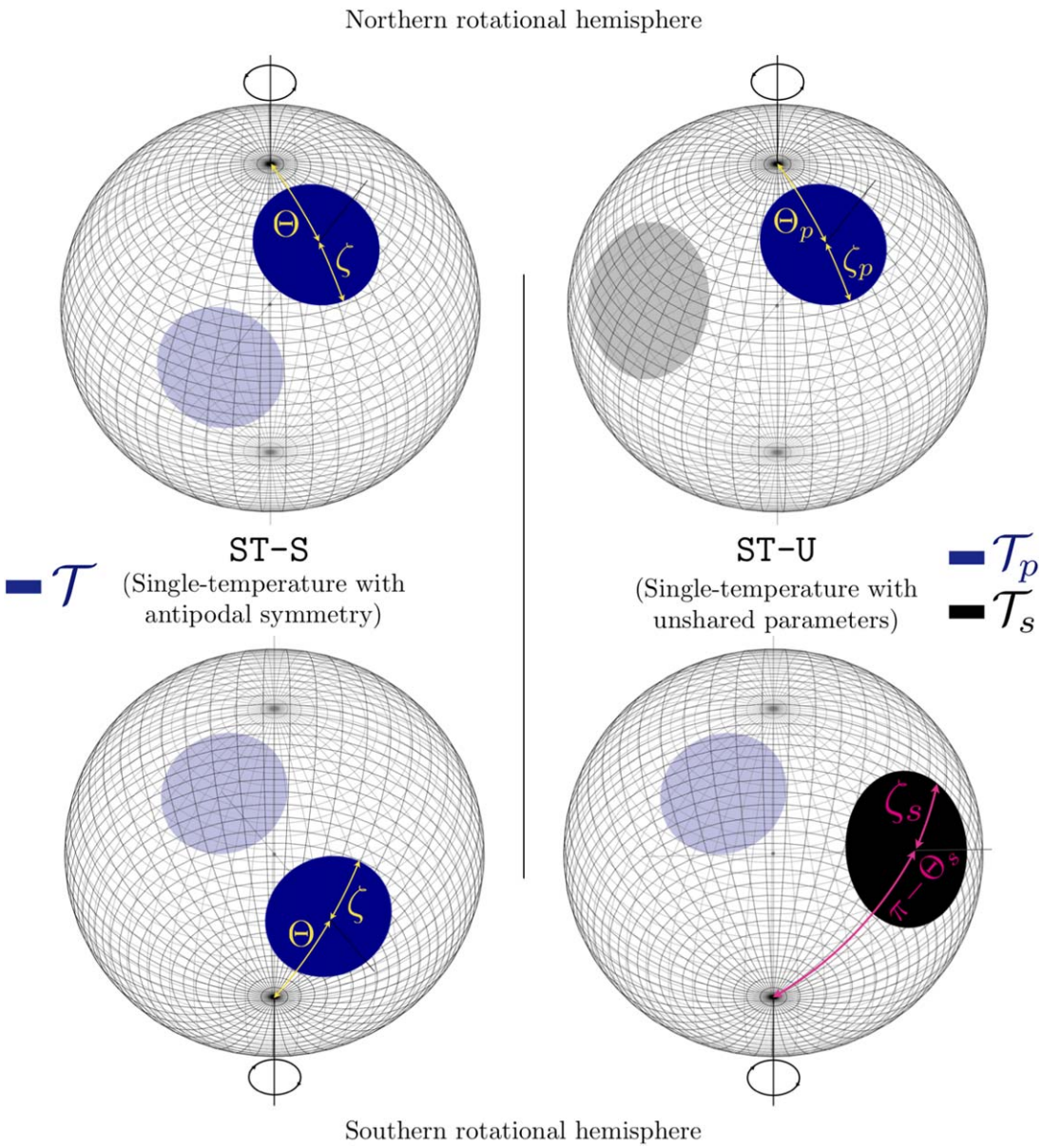

Figure 3. Schematic diagrams of models with single-temperature regions: ST-S defined by antipodal symmetry of the primary and secondary regions, and ST-U defined by the primary and secondary regions not sharing any parameters. Note that the two-surface shown is spherical: the hot regions are defined in spherical coordinates on a (unit) sphere and projected onto the rotationally distorted (oblate) spheroid representing the two-surface of the star, such that the proper areas of the regions are only implicitly defined. To construct the ST-U diagram from the ST-S diagram we displaced the secondary region (defined by having a center colatitude $\Theta_s \geqslant \Theta_p$) away from the antipode of the center of the primary region (with center colatitude Θ_p), decreased ζ_s such that $\zeta_s < \zeta_p$, and assigned the secondary region a distinct temperature parameter. Note that no numeric relationship is implied between the effective temperatures of material within each member—indeed, the prior is separable with respect to the temperatures ($\mathcal{T}_p$ and $\mathcal{T}_s$), and the support includes combinations $\mathcal{T}_s \leqslant \mathcal{T}_p$ and $\mathcal{T}_s > \mathcal{T}_p$. For clarity we display for ST-S a projection showing the southern rotational hemisphere; in subsequent diagrams we omit such a projection when antipodal symmetry applies.

The numerical marginalization operation implemented is described in Appendix B of the *X-PSI* documentation (Riley & Watts 2019).

2.5. Model-specific Definitions

The surface heating distribution by realistic magnetospheric (return) currents remains uncertain. The global magnetic field may be more complex than a simple dipole (at least in the near vicinity of the surface), while the mapping between currents and surface temperature field is not well determined by existing theoretical models (Harding & Muslimov 2001, 2011; Timokhin & Arons 2013; Philippov et al. 2015a; Gralla et al. 2017; Lockhart et al. 2019). We thus consider a set of simplified models that are representative of the various theoretical possibilities, albeit restricting our analysis to models with two distinct hot regions. We allow for the possibility of the hot regions being non-antipodal and non-identical (Pavlov & Zavlin 1997; Bogdanov et al. 2007, 2008; Bogdanov 2013); we also consider various hot-region shapes, including circles, rings, and crescents filled with material of uniform local comoving temperature.

While our choices are physically motivated, it is important to emphasize that our inferences are conditional upon these choices. However, posterior computation is computationally intensive and scales with model complexity. The work

presented here is based on usage of a 500,000 core-hour grant on the Dutch national supercomputer Cartesius; we thus find it pragmatic to disseminate information to the community at this point. Further exploration of model variants or execution of higher-resolution calculations requires additional resource allocation on high-performance systems, building upon the information offered here as guidance. In particular, it will be important to explore the sensitivity of marginal posterior estimates of fundamental physical parameters of interest—i.e., exterior and interior spacetime parameters—to expansion of the space of models that we have had the resources to consider.

In this section we detail the properties that distinguish the models in our model space. For each model we give the parameterization details and any remarkable prior details; notes on the *support* of the joint prior distribution are given where appropriate. We also discuss the existence of (continuous linear and nonlinear) degeneracy in posterior modes for a given parameterization, which we interpret as one indicator of unnecessary complexity for a surface radiation field with phenomenological spatial structure. For each model whose associated posterior distribution we compute, we provide a summary table containing a more precise definition of the joint prior distribution; these tables may be found in Section 3 and in Appendix C.

2.5.1. Single-temperature Regions with Antipodal Symmetry (ST-S) and with Unshared Parameters (ST-U)

Parameterization. The primary hot-region (refer to the leftmost panel of Figure 3) is simply connected and encloses radiating material—a fully ionized hydrogen NSX atmosphere with effective temperature $\mathcal{T}$. The boundary of the region is *circular*: i.e., given a *center* point on the surface with colatitude Θ_p and azimuth ϕ_p (in a spherical coordinate basis whose polar axis is defined as the stellar rotation axis), the boundary is the locus of points that are equidistant[51] in angular space from the center point.

Hereafter we use the alias ST-S, parsed as *Single-Temperature-Shared*. The surface radiation field associated with the secondary hot region is derived exactly by applying antipodal symmetry to the primary region: there are no free parameters associated with the secondary region.

Similarly, we use the alias ST-U, parsed as *Single-Temperature-Unshared*. The primary region (refer to the rightmost panel of Figure 3) definition is retained from ST-S as defined above. The secondary region, however, is now endowed with distinct parameters—i.e., the region is *not* derived from the primary region under antipodal symmetry. The parameters of the secondary region have an otherwise equivalent meaning—in terms of surface radiation field specification—to their primary-region counterparts.

Degeneracy. We note that a discrete degeneracy—multimodality—can in principle arise for a source such as PSR J0030+0451, but may only be weak when there is detectable asymmetry between the two component pulses over the course of one rotational cycle. There may exist two phase solutions, each corresponding to a distinct mapping between hot regions (distinguished by colatitude) and the pulse components in the event data. For instance, the primary (lower-colatitude) region could in principle generate either of the component pulses, while the secondary region generates the other. Fortunately, a number of open-source sampling software packages are designed to handle multi-modality efficiently (at least in the absence of nonlinear degeneracy). If the asymmetry between the component pulses is clear, the posterior mass in one mode may be entirely dominant.

Priors. For ST-S, we eliminate a region-exchange degeneracy by imposing a constraint $\Theta_p \leqslant \pi/2$ on the prior support. The primary region is uniquely defined as the region whose center subtends the smallest colatitude, Θ_p, to the rotational axis, if the region colatitudes are different.

For ST-U, we eliminate a region-exchange degeneracy by imposing a constraint $\Theta_p \leqslant \Theta_s$ on the support of the joint prior distribution. The primary region is uniquely defined as the region whose center subtends the smallest colatitude, Θ_p, to the rotational axis, if the region colatitudes are different; the regions are distinguishable when $\Theta_p = \Theta_s$ according to the subset of parameters that controls their physical manifestation. The joint prior support is such that the two regions cannot overlap but otherwise are not restricted to be antipodally symmetric.

2.5.2. Concentric Single-temperature Regions with Antipodal Symmetry (CST-S) and with Unshared Parameters (CST-U)

Parameterization. The primary hot region (see the leftmost panel of Figure 4) is a non-simply connected annulus (or ring) with outer angular radius ζ, which contains material with effective temperature $\mathcal{T}$. The non-radiating *hole* with angular radius ψ is concentric (in angular coordinates) with the radiating annulus. We thus recover the shape defined for the ST-S and ST-U variants in the limit $\psi \to 0$ (which is at the boundary of the prior support).

A generally useful way to distinguish the hole and the annulus—in particular for further increments in complexity—is as follows. Recall the term *member* from Section 2.4.1: consider two simply connected partially overlapping member regions, each wholly filled with radiating material, but impose the logical condition that when evaluating radiating intensities at a spacetime event on the stellar surface, one member—the hole—takes precedence if the event falls within its boundary. In this case (for the models here described), let the temperature of the material in the hole be (effectively) zero so that no signal need be computed for the hole. The statements in Section 2.4.1 pertaining to the proper areas in finite-element representations of radiating regions apply here: the annulus is a subset of a (simply connected circular) ceding member that is *not superseded* by the hole when evaluating local radiation intensities, and its proper area is computed (almost) exactly. These constructions are useful for further extension of the model.

Hereafter we use the alias CST-S, parsed as *Concentric-Single-Temperature-Shared*. For CST-S, the surface radiation field associated with the secondary region is derived exactly by applying antipodal symmetry to the primary region: there are no free parameters associated with the secondary region. The annuli share an outer angular radius ζ, and the holes share a fractional angular radius f such that the hole angular radii are $\psi := f\zeta$.

Similarly, we use the alias CST-U, parsed as *Concentric-Single-Temperature-Unshared*. For CST-U, the primary region (refer to the rightmost panel of Figure 4) definition is retained

[51] The ambient spacetime is static, and with respect to a Schwarzschild chart, the points are equidistant in angular coordinates. However, when projected from a spherical two-surface onto that of a rotationally deformed (oblate) spheroid, the spacelike separation—on a Schwarzschild temporal hyperslice—between the center point and boundary points is *not* invariant.

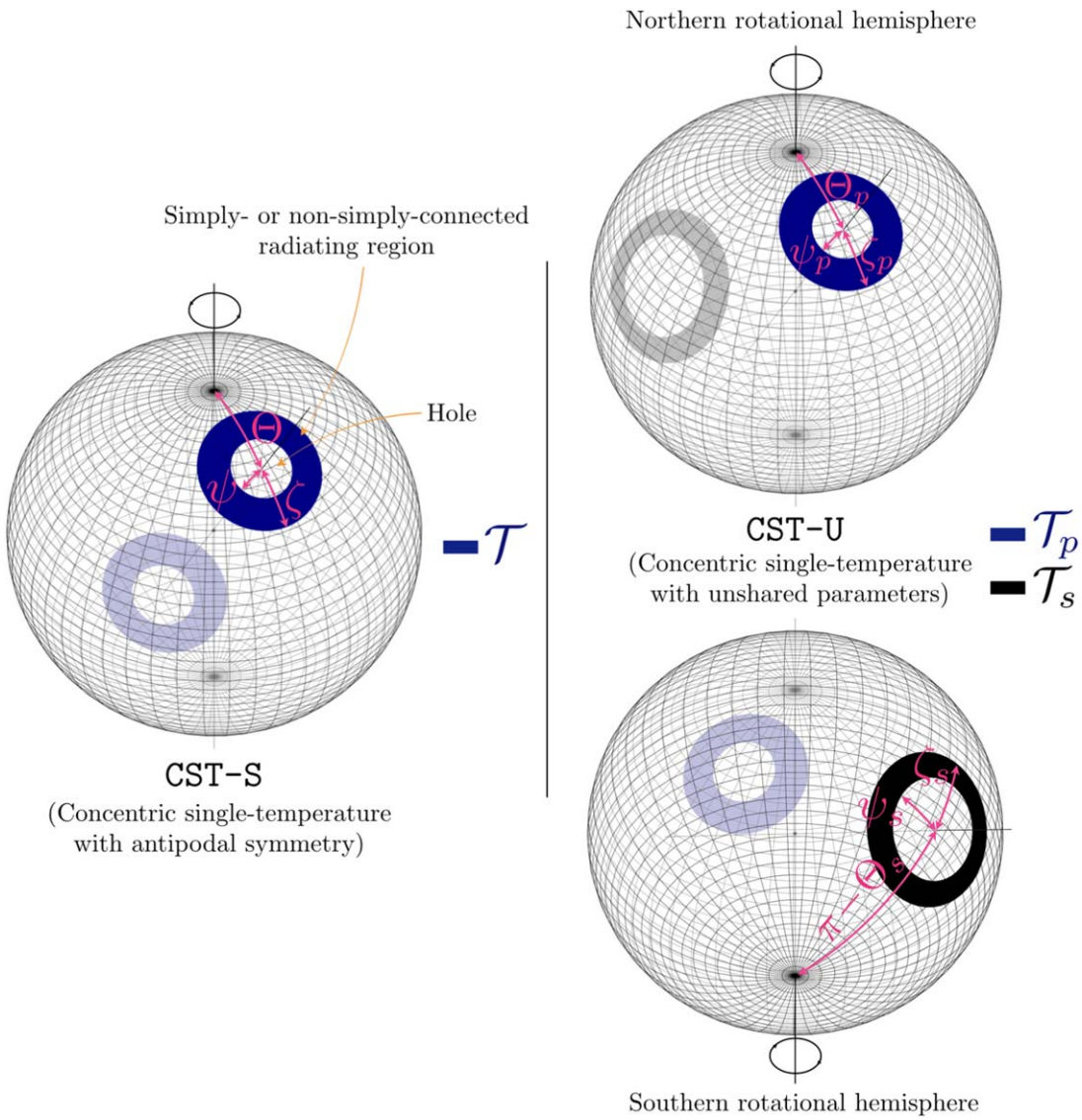

Figure 4. Schematic diagrams of models with concentric single-temperature regions: `CST-S` defined by antipodal symmetry of the primary and secondary regions, and `CST-U` defined by the primary and secondary regions not sharing any parameters.

from `CST-S` as defined above. The secondary region, however, is now endowed with distinct parameters—i.e., it is *not* derived from the primary region under antipodal symmetry. The parameters of the secondary region have an otherwise equivalent meaning—in terms of surface radiation field specification—to their primary-region counterparts.

Degeneracy. We now consider the continuous degeneracy labeled I in Figure 5. When the angular extent of a radiating region is small, the signal generated by that region—as registered by a distant detector that does not spatially resolve (image) the star—is insensitive to its shape (detailed spatial structure). Sensitivity is here a measure in terms of the likelihood: i.e., the *total* variation of the parameterized joint sampling distribution of a set of random variables,[52] in response to motion along a certain set of curves[53] in parameter space, is small as summarized by the scalar likelihood. It follows that if the shape is parameterized with more than a single degree of freedom, the likelihood function is degenerate with respect to shapes that satisfy a constraint on the solid angle subtended by the radiating region at the center of the star, and thus which satisfy a constraint on the proper area of the radiating region.

The degeneracy is not in this instance exact, but holds approximately. Consider a ceding member of angular extent $\cos\zeta \lesssim 1$ and hole with angular extent $\cos\psi > \cos\zeta$: the degeneracy is such that the solid angle, $\varpi(\zeta, \psi) = \cos\psi - \cos\zeta$, of the hot region (the annulus), is approximately $\varpi = 1 - \cos\tilde{\psi}$, where $\zeta = \tilde{\psi}$ is the angular radius in the limit $\psi \to 0$. In other words, the relative size—and indeed existence—of the hole is at most weakly constrained on small angular scales. This degeneracy in the (ζ, ψ)-subspace is nonlinear for $\zeta \to \tilde{\psi}$, but linearizes for increasing $\zeta > \tilde{\psi}$; however, with increasing ζ the signal generated by the region evolves away from that generated in the limit $\zeta \to \tilde{\psi}$ with $\psi \to 0$. Note that if the superseding member is not concentric with the ceding

[52] And whose number usually exceeds the number of parameters.

[53] Where those curves may more generally together generate m-dimensional surfaces in an n-dimensional space where usually $n > m$—if $m = n$ then no facet of the model is constrainable and the model is arguably not useful unless it can meaningfully tested in some other manner.

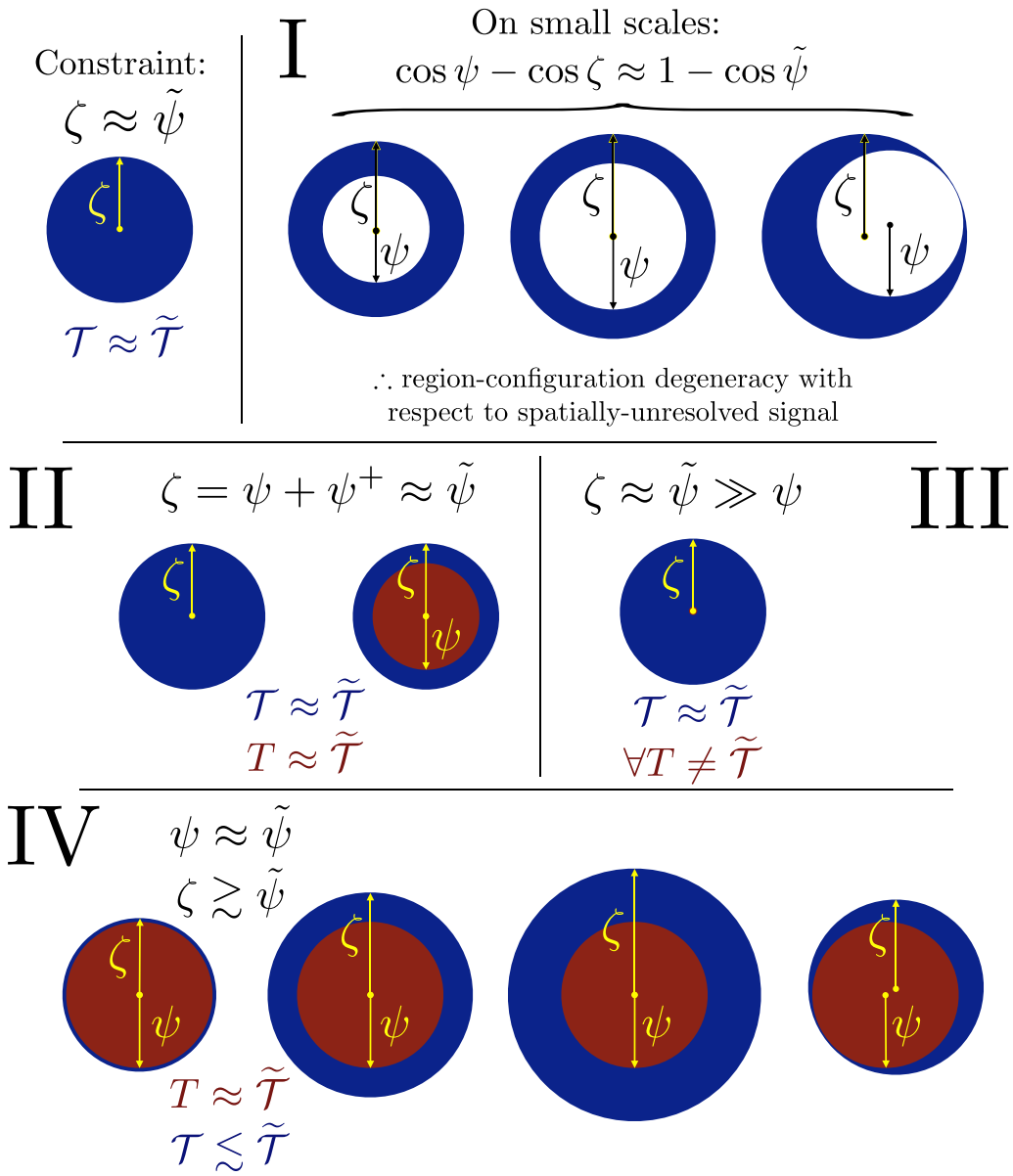

Figure 5. Schematic diagram of (approximate and exact) continuous degeneracies between hot-region structure parameters with respect to the energy-phase-resolved signal incident on a distant telescope. Degeneracy type I is discussed in detail in Section 2.5.2, while types II through IV are discussed in Section 2.5.3; here we focus on defining the components of the diagram. In all cases, let us suppose that there exists a strong posterior constraint on the (coordinate) solid angle ϖ subtended by a component of a hot region (at the stellar origin) with temperature $\tilde{\mathcal{T}}$. Type I—we consider a hot region with a single-temperature component, and represent such a region as simply connected with a strongly constrained radius $\zeta = \tilde{\psi}$. If the angular extent of the headed region is sufficiently small that we are insensitive (in the absence of spatial imaging power) to structural details other than solid angle ϖ, then we can introduce a hole of angular radius ψ such that the hot region has the topology of a ring and a likelihood function degeneracy $\varpi = \cos\psi - \cos\zeta =$ const. manifests (on small angular scales). Similar arguments can be made for other parameterized structural modifications to the hot region. Type II—let us consider a hot region constructed from an annulus and a hole (concentric or eccentric), each filled with material of local comoving temperature $\mathcal{T}$ and T, respectively. Let us assume a strong constraint on the outer angular radius $\zeta = \tilde{\psi}$ of the annulus, where $\tilde{\psi}$ is now not restricted to small angular scales; let us further assume a strong constraint on the structure of the hot region as simply connected circular with uniform temperature $\tilde{\mathcal{T}}$, where only one temperature component is useful. It follows that degeneracy manifests for $\mathcal{T} \approx T \approx \tilde{\mathcal{T}}$ and $\psi + \psi^{+} \approx \tilde{\psi}$, where $\psi^{+} := \zeta - \psi$. Type III—as type I, but in the limit that $\zeta \gg \psi$, such that the signal generated by the hole, whose temperature is degenerate over a range dependent on ψ, is effectively turned off. Type IV—the signal generated by the hole now satisfies the constraints assumed for type II—i.e., $T \approx \tilde{\mathcal{T}}$ and $\psi \approx \tilde{\psi}$—while the solid angle, temperature, and coordinates of the annulus exhibit degeneracy with respect to configurations that effectively turn off the signal generated by the annulus.

member, and the overlap is only partial, the form of the degeneracy—the constraint equation satisfied—has additional dependence on the coordinates of the center of the hole relative to the center of the ceding member.

It is important to be aware of such degeneracy for the purpose of efficient posterior computation—inherent to which is accuracy. In our case, if the type of signal that is superior for describing the event data is generated by radiation from a localized region on the star, the constraint on the solid angle of the region can be viewed as dragging a posterior mode through parameter space along the type I degeneracy direction. Thus, with this parameterization, nonlinearity will exist that will reduce efficiency to some degree.[54]

[54] Which can only be robustly learned during computation.

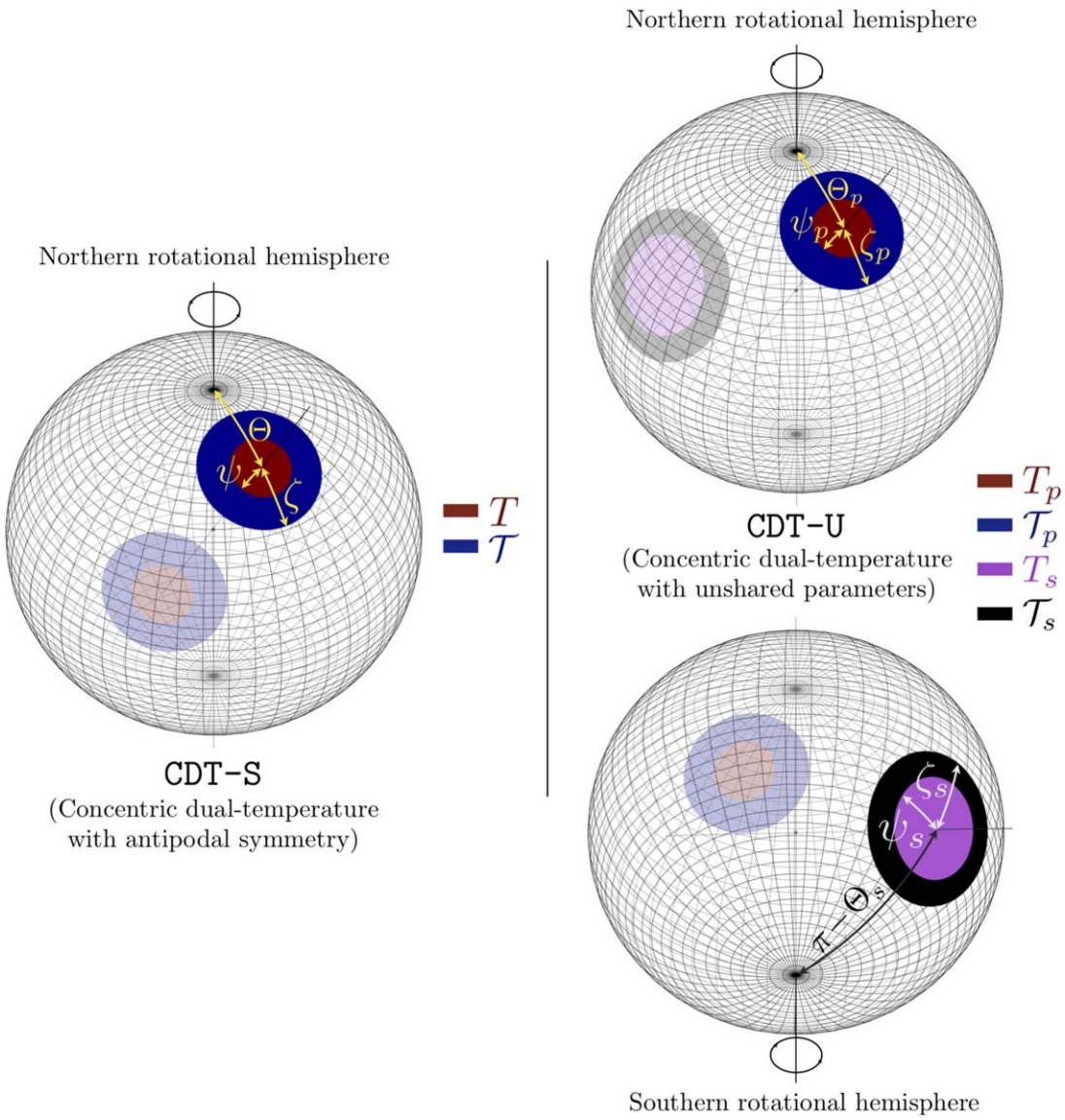

Figure 6. Schematic diagrams of models with concentric dual-temperature regions: `CDT-S` defined by antipodal symmetry of the primary and secondary regions, and `CDT-U` defined by the primary and secondary regions not sharing any parameters. To construct the `CDT-U` diagram from the `CDT-S` diagram we displaced the secondary region (defined by having a center colatitude $\Theta_s \geqslant \Theta_p$) away from the antipode of the center of the primary region (with center colatitude Θ_p), decreased ζ_s such that $\zeta_s < \zeta_p$ while leaving $\psi_s = \psi_p$, equivalently meaning $f_s > f_p$, and assigned distinct colors to the hole and annulus of the secondary region. We retained the caveat that no numeric meaning is implied regarding the effective temperature of the material in the members—indeed, the prior is separable with respect to all temperatures (T_p, $\mathcal{T}_p$, T_s, $\mathcal{T}_s$), and the support includes combinations $\mathcal{T}_s \leqslant T_s$ and $\mathcal{T}_s > T_s$.

If the posterior predictive performance is maximal for signals that are generated by localized emission, and degenerate posterior structure is observed, then clearly the most effective manner in which to achieve efficiency increase is to *reduce* the complexity of the structure of the radiating region. One thus inserts a simpler model into the model space, with the caveat that while estimation of ulterior model parameters should be insensitive to this reduction in complexity, the evidence may not be.

Alternatively, working with an integral summary variable such as the solid angle ϖ of the (uniform temperature) radiating region is useful from the perspective of eliminating degeneracy by parameterizing directly in terms of variables to which we are statistically sensitive. On the other hand, the mapping from ϖ to variables that directly control the shape of the radiating regions can behave undesirably, and can thus complicate the action of extending models.

One potential avenue for efficiency improvement by linearizing the degeneracy, is to sample in the space of $(\cos\zeta, \cos\psi)$, which will eliminate the emergent small-scale nonlinear degeneracy. The cost is complication of the joint prior definition and implementation: in this case, a singularity[55] exists in the mapping for $\psi \to 0$, which is the boundary at which the radiating region reduces to being simply connected as for both `ST-S` and `ST-U`. If we define finite joint prior density $p(\zeta, \psi)$ at points where $\psi = 0$, the joint density $p(\cos\zeta, \cos\psi)$ is divergent (although integrable, possibly in closed form).

Priors. An issue to be aware of is that the mapping from the native sampling space—that of the unit hypercube—would need to avoid introducing nonlinearity, otherwise the effort to

[55] Note that Taylor-expanding on small angular scales to work in the joint space of (ζ^2, ψ^2) by definition cannot bypass the singularity.

improve efficiency may be in vain (see the discussion in Section 2.4.1 regarding implementation of the joint prior distribution of M and $R_{\rm eq}$).

We define the prior as separable on the joint space of f and ζ. Specifically, we condition on $f \sim U(0,\ 1)$ and $\zeta \sim U(\epsilon_\zeta, \pi/2 - \epsilon_\zeta)$ where $\epsilon_\zeta \ll \pi/2$. For posterior computation, in order to ease prior implementation, we opt to transform the prior onto the joint space of ψ and ζ and accept the nonlinear degeneracy in the limit $\zeta \to \tilde{\psi}$ for small $\tilde{\psi}$. The chosen joint prior is not separable on this chosen space, but remains straightforwardly implementable.

For CDT-U, we eliminate a region-exchange degeneracy by imposing a constraint $\Theta_p \leqslant \Theta_s$ on the support of the joint prior distribution. The primary region is uniquely defined as the region whose center subtends the smallest colatitude, Θ_p, to the rotational axis; the regions are distinguishable when $\Theta_p = \Theta_s$ according to the subset of parameters that controls their physical manifestation. The joint prior support is such that the two regions cannot overlap but otherwise are not restricted by antipodal symmetry. Note that imposing that the regions are non-overlapping modifies the marginal prior density $p(\zeta)$—and thus $p(\psi)$—by redistributing prior mass to lower angular radii.

2.5.3. Concentric Dual-temperature Regions with Antipodal Symmetry (CDT-S) and with Unshared Parameters (CDT-U)

Parameterization. We extend the CST models by filling the *holes* (the superseding members) with radiating material at finite temperature and evaluating the signal generated by both the superseding and ceding members. The hot region is simply the union of the members together with an order of precedence, and can be considered to have two heated *subregions* with distinct temperatures, each of which generates a *component* of the signal.

Hereafter we use the alias CDT-S, parsed as *Concentric-Dual-Temperature-Shared*. For CDT-S, the surface radiation field associated with the secondary region is derived exactly by applying antipodal symmetry to the primary region: there are no free parameters associated with the secondary region. The annuli share an outer angular radius ζ, and the holes share a fractional angular radius f such that the hole angular radii are given by $\psi := f\zeta$.

Similarly, we use the alias CDT-U, parsed as *Concentric-Dual-Temperature-Unshared*. For CDT-U, the primary region (see the rightmost panel of Figure 6) definition is retained from CDT-S. The secondary region, however, is now endowed with distinct parameters—i.e., it is *not* derived from the primary region under antipodal symmetry. The parameters of the secondary region have an otherwise equivalent meaning—in terms of surface radiation field specification—to their primary-region counterparts. It follows that parameter vectors that correspond to, e.g., the primary region having a hotter annulus while the secondary region has a hotter hole, are assigned finite local joint prior density.

Degeneracy. Introducing a second temperature component also introduces degenerate structure in the likelihood function—a fundamental problem with such phenomenological descriptions of the surface radiation field. We now consider the continuous degeneracies labeled I, II, III, and IV in Figure 5.

Degeneracy of type I manifests on small angular scales when the temperature T of the superseding member is small relative to the temperature $\mathcal{T}$ of the ceding member. In this case, the signal from the superseding member is dominated and we effectively recover the structure from the CST models; the reader should imagine the hole in panel I of Figure 5 as colored and labeled $T \ll \mathcal{T}$.

Degeneracy of type II manifests because the temperatures of the ceding and superseding members can be approximately and exactly equal. Thus, given a strong constraint on the temperature, the solid angle ($\varpi \approx 1 - \cos\zeta$) constraint on the radiating region can be satisfied $\forall\, \psi \leqslant \zeta$. Note that unlike degeneracy of type I, degeneracy of type II can be exact (for $T = \mathcal{T}$) and exists invariantly of angular scale. The smaller the solid angle constraint, the weaker the constraint on topology of the hotter subregion (annulus or hole), and thus degeneracy of type II should be present. On larger angular scales, the signal offers a much stronger constraint on topology: if the hottest subregion is ring-like,[56] degeneracy of type II cannot be present a posteriori.[57]

Degeneracy of type III manifests because the smaller the superseding member is in relation to the ceding member, the greater the dominance of the signal from the ceding member, and thus the greater the range of values T can assume while leaving the signal approximately invariant. Thus, given a strong constraint on the temperature and solid angle ($\varpi \approx 1 - \cos\zeta$) of the dominant component, the constraint can be satisfied $\forall\, \psi \ll \zeta$. Again note that unlike degeneracy of type I, the existence of degeneracy of type II is insensitive to angular scale. However, if there is a strong constraint on the topology of the subregion that dominates the joint signal, and that subregion is ring-like, degeneracy of type II will not be present a posteriori.

We now consider degeneracy IV and its implications for sampling. Suppose that the signal can be dominated by that from a simply connected subregion: if the superseding member satisfies both a temperature constraint and a solid angle constraint ($\varpi \approx 1 - \cos\psi$), the ceding member then forms an annulus whose width ψ^+ and whose temperature $\mathcal{T}$ can together assume a wide range of values. In Figure 5 we illustrate the case of $\mathcal{T} \ll T$ for varying ζ (and, in the last cartoon, for a shift of the center of the ceding member in angular space).[58] The superseding member need only be described by four parameters, while the non-superseded subset of the ceding member is described by five parameters, three of which are shared with the superseding member. If the ceding member is relatively cool, then the signal it generates is dominated by that of a much hotter superseding member; the signal is then degenerate with respect to $\mathcal{T}$. Crucially, if one writes the properties of the superseding member in terms of that of the ceding member, nonlinear degeneracy can arise in what otherwise may appear to be a natural parameterization, as we show below.

For posterior computation (see Appendix A for methodology) we explicitly consider a parameter space in which the superseding member is constructed using the minimal number

[56] Formed by imposing a relatively large, cooler superseding member (effectively a hole).

[57] Here we assume a unimodal posterior distribution, but more generally, the degeneracy will not be present for a *local* mode in which the topology is strongly constrained as ring-like.

[58] Note that a less important degeneracy occurs, which we do not illustrate: the annulus is arbitrarily hotter ($\mathcal{T} > T$) but $\psi^+ \to 0$ so that the superseding member also dominates the signal. In the $T \approx \mathcal{T}$ transition zone—which coincides with the type II degeneracy—ζ approximately satisfies the solid angle constraint.

of parameters necessary, which eliminates needless *nonlinear* degeneracy. For example, if we work in the joint space of f and ζ, the local direction of approximate signal invariance—which emerges when the superseding member with angular radius ψ is dominant—is *not* everywhere a basis vector, but is given by the gradient of $f = \psi/\zeta$ for constant $\psi \in \mathbb{R}^+$: $\partial f/\partial\zeta = -\psi/\zeta^2$, so the direction is $(1, -\psi/\zeta^2)$ at point (ζ, f). This degeneracy is linearized by working in the joint space of ψ and ζ, or alternatively, in the joint space of ψ and the angular annular width $\psi^+ := \zeta - \psi$.

We note that unlike degeneracy of type I, the existence of degeneracy of type IV is insensitive to angular scale. However, if there exists a strong constraint on the topology of the heated subregion that dominates the signal, and that subregion is ring-like, degeneracy of type IV will not be present a posteriori—in the same vein that degeneracy of types II and III are absent. This does not mean that the posterior modes are devoid of degeneracy if the hot regions are both constrained to be ring-like: the temperature of a superseding member can assume a wide range of values (in logarithmic space) lower than that of the ring, because the additional complexity beyond that of a non-radiating hole (i.e., devoid of heated material) is unwarranted. However, the degeneracy is linear and thus handled straightforwardly.

In many cases, degeneracies are characterized by allowing one temperature component (a heated subregion) to dominate the signal, while one temperature—or the solid angle subtended at the stellar origin by a subregion—assumes relatively low values. In these cases the degeneracy is not expected to be particularly problematic because it will be effectively linear, parallel to basis vectors of the parameter space, and truncated by the boundary of the prior support. Thus, provided that the mapping from the native sampling space to the parameter space preserves this behavior, nested sampling efficiency may not be affected as adversely as when posterior modes exhibit strong nonlinear degeneracy.

For instance, consider degeneracy such as type IV: let the angular scale be sufficiently large for the dominant component to be constrained to be simply connected. The degenerate subsets of the (ψ, ζ)- and $(\psi, \mathcal{T})$-subspaces will be linear and orthogonal to the $\hat{\psi}$ basis vector. The notable source of efficiency reduction for type IV will be the boundary of the degenerate subset of the $(\zeta, \mathcal{T})$-subspace: while there will be a hard lower limit on $\mathcal{T}$, the boundary is otherwise dependent on combinations of parameters and will not conform trivially to common nested-sampling active-point bounding algorithms.

On the other hand, if degeneracy of type I arises and the hole is filled with material of temperature T, we incur a nonlinear degeneracy in the (ζ, ψ)-subspace, *and* non-trivial boundaries of the degenerate subsets of the (ζ, T) and (ψ, T)-subspaces. In this case the potential for efficiency reduction is greater.

It is interesting to note that degeneracy can also arise when one temperature component does *not* dominate the signal. Let us suppose that two components with distinct temperatures are favored a posteriori, where those components both contribute non-negligibly to the total signal. We implement a numerical geometrically thin atmosphere where the effective temperature and effective gravity control the local comoving specific intensity as a function of photon energy and direction (i.e., control the spectrum and beaming); it follows that if a hot region exhibits small angular extent, two components (subregions) can in principle generate signals that are commensurate in total count rate, provided that the solid angles subtended by the subregions at the stellar origin have appropriate relative sizes. For the CDT models, a *discrete* degeneracy—multi-modality—can then arise on small angular scales: pairs of configurations of the components generate approximately the same total signal. In one configuration the hole hosts a particular component, while in the alternate configuration that component is hosted by the annulus; the relative solid angles subtended by the hole and annulus depend on which component is hosted. However, for a more complex model where the superseding and ceding members are not defined as concentric (see Figure 5, and Section 2.5.5 and beyond), it is clear that (approximate) continuous degeneracies would also arise and thus complicate matters.

We should also be aware of the relative prior masses associated with subsets of parameter space over which the signal is approximately invariant, governed roughly by the dimensionality of the degeneracy. Degeneracy of types II and III generally occupy a subset of parameter space with smaller prior mass than that of type I and IV, the former two being effectively to one-dimensional,[59] and the latter two being effectively two-dimensional. Note that the type I degeneracy is effectively two-dimensional *because* we fill the hole from the CST models with material whose temperature is finite and a model parameter.

Priors. We retain almost all prior definitions from the CST models (Section 2.5.2). We again transform the prior onto the joint space of ψ and ζ in order ensure a more optimal parameterization for our sampling algorithm of choice, as highlighted above. Crucially, we need to transform from the native space to the parameter space in a manner that preserves linearity of degeneracy IV. We achieve this by inverse sampling the marginal prior density $p(\psi)$ and the condition prior density $p(\zeta \mid \psi)$; we give these transforms in Appendix B. In some cases we opt to transform to the joint space of ψ and ψ^+ space in post-processing for optimality of kernel density estimation when ψ^+ approaches a boundary—e.g., due to degeneracy of type IV.

2.5.4. Interlude: On Degeneracy and Complexity

General. The continuous degeneracies discussed in Sections 2.5.2 and 2.5.3, when in mutual existence, generally form a connected structure in parameter space. These degeneracies could be characterized in more detail, and also with reference to ulterior parameters (including those of the other hot region). We do not study the degeneracies further here for lack of a clear way to simultaneously linearize each of them via transformation.

We instead observe a fundamental aspect of the modeling process *on a source-by-source basis*: it is judicious to evaluate the utility of additional complexity *on a region-by-region basis*, if a posteriori one infers localized emission and/or that a single-temperature component per hot region suffices in describing a particular pulse component in the phase-folded event data. In practice, this could include the following: (i) design a model space with increments in complexity, moving toward more sophisticated region topologies and/or boundaries, and/or an additional temperature component as warranted by the data; (ii) insert models into one's model space

[59] An unillustrated degeneracy, characterized by a hotter ring in the limit $\psi^+ \to 0$, is also effectively one-dimensional.

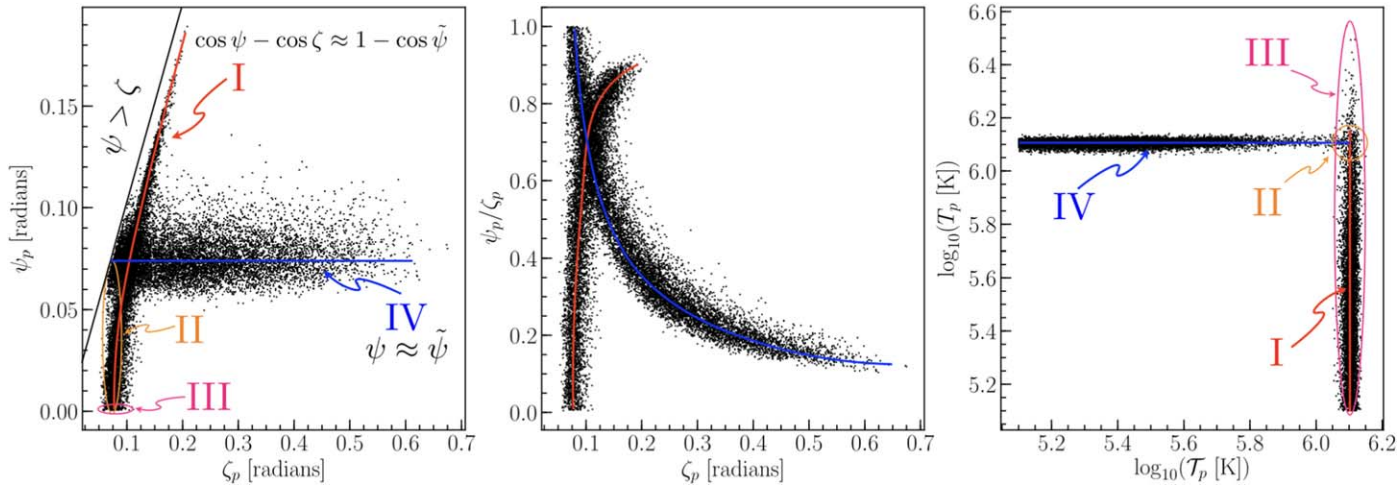

Figure 7. Conditional on CDT−U, we display points (nested samples) $\{\boldsymbol{\theta}_k\}$ that reported (background-marginalized) likelihood function values $L(\boldsymbol{\theta}_k) > L'$, where $\max_k \ln L(\boldsymbol{\theta}_k) - \ln L' \lesssim 13$; the choice of threshold L' is somewhat arbitrary beyond needing a large enough number of points to clearly resolve the form of the degeneracies. We display these points in a selection of two-dimensional spaces to focus on the degeneracies that emerge upon application of CDT−U. These degeneracies pertain to the structure of the primary (lower-colatitude) hot region, which generates a signal describing a particular pulse component visible in Figure 1; the hot region exhibits a small angular extent on the stellar surface, and we are statistically insensitive to the additional complexity offered by a CDT cap over a ST cap. In the left panel we show the points in the joint space (ζ_p, ψ_p)—i.e., in the joint space of the angular radii of the ceding and superseding members. We indicate which type of degeneracy illustrated in Figure 5 corresponds to which structure in parameter space; the curves (blue and red) each correspond to a constraint equation. Types II and III occupy far smaller prior masses than types I and IV, so we choose to indicate the structures associated with the former types with ellipses; in a finite-sample context these structures are allocated points more sparsely, in proportion to the associated prior mass in the full n-dimensional parameter space. Note that the region $\psi_p > \zeta_p$ is not a subset of the CDT−U prior support. In the center panel we display the points in the joint space of (ζ_p, f_p), where $f_p := \psi_p/\zeta_p$; our sampling processes (refer to Appendix A for details) were executed in a space in which likelihood function isosurfaces are structured nonlinearly as suggested in this center panel, leading to reduced posterior computation efficiency and higher nested-sampling error. In Sections 2.5.2 and 2.5.4 we discuss the interpretation and treatment of degeneracy: in hindsight, higher efficiency would have been achieved if we had worked in a joint space that linearized degeneracy of type IV. In the right panel we display the corresponding structures in a joint space trivially related to $(\mathcal{T}_p, T_p)$—i.e., in the joint space of the effective temperatures of the material enclosed by the ceding and superseding members. The points displayed here are a composite set from two sampling processes, where due to reduced efficiency and increased nested-sampling error, neither process individually resolved all of the degenerate structure present.

that reduce region complexity, if it is apparent a posteriori that a model with unhelpful complexity was first applied; or (iii) reduce region complexity as in point (ii), while increasing the complexity of another facet of the model, based for example on graphical posterior predictive checking.

To provide an example, suppose that a posteriori, given CDT−U, hot region $\mathcal{A}$ is degenerate with respect to multiple parameters controlling its surface radiation field, while hot-region $\mathcal{B}$ is only degenerate with respect to temperature of one relatively cool component. Specifically, suppose $\mathcal{A}$ is constrained to have a small angular extent and a single-temperature component, while offering satisfactory posterior predictive performance in comparison to a particular pulse component. Continuous degeneracies of type I and IV are then clear in the posterior distribution. Additionally, suppose that the secondary region has a strong constraint on its topology, with its dominant component being ring-like with moderate angular extent. On the basis of this inference, one might argue that it is justified to modify the model space: one proceeds to define a model wherein the complexities of $\mathcal{A}$ and $\mathcal{B}$ are unequal. In doing so one can reduce or redistribute complexity while improving computational efficiency. For the aforementioned example, one may define a model ST+CST, parsed as *Single-Temperature + Concentric-Single-Temperature*, which is intermediary in relation to ST−U and CST−U: $\mathcal{A}$ has one simply connected component, while $\mathcal{B}$ has one non-simply connected component (which can reduce to a simply connected component). The dimensionality of the parameterization is then $n = 9$, reduced from the CDT−U value of $n = 12$. One could also define additional models that increase the complexity of $\mathcal{B}$ (see Sections 2.5.5 and 2.5.6).

As a brief second example: if, given CDT−U, $\mathcal{B}$ was constrained to have *two* temperature components, one could instead define a model ST+CDT with dimensionality $n = 10$. One could then proceed to increment the complexity of $\mathcal{B}$.

An important remark here is that the performance of $\mathcal{A}$ cannot strictly be decoupled from the performance of $\mathcal{B}$ in relation to their respective target pulse components: the regions exist on the same rotating two-surface and share an ambient spacetime solution. It follows that the complexity of one region can in principle affect our conclusions about the level of complexity of the other. Such conclusions cannot be based solely on the existence of degenerate structure in the joint posterior distribution of the region parameters. If $\mathcal{A}$ exhibits degenerate structure while $\mathcal{B}$ does not, the structure may manifest—at least in part—*because* the complexity of $\mathcal{B}$ is insufficient.

A more robust basis for breaking parity between the hot-region complexities would be the existence of degenerate structure, *and* satisfactory posterior predictive performance of *both* regions in comparison to their respective pulse components. A practical basis for breaking parity, given the existence of efficiency-reducing degenerate structure a posteriori, may be resource availability and management.

Modeling PSR J0030+0451. For the source we focus on in this work, degeneracy emerged upon application of CDT−U. We display the corresponding structures, in parameter space, in Figures 7 and 8. In summary, the primary region was constrained to have a small angular extent and one dominant temperature component, and could generate data structurally consistent with a pulse component visible in the phase-folded event data (Figure 1); the primary region thus here corresponds

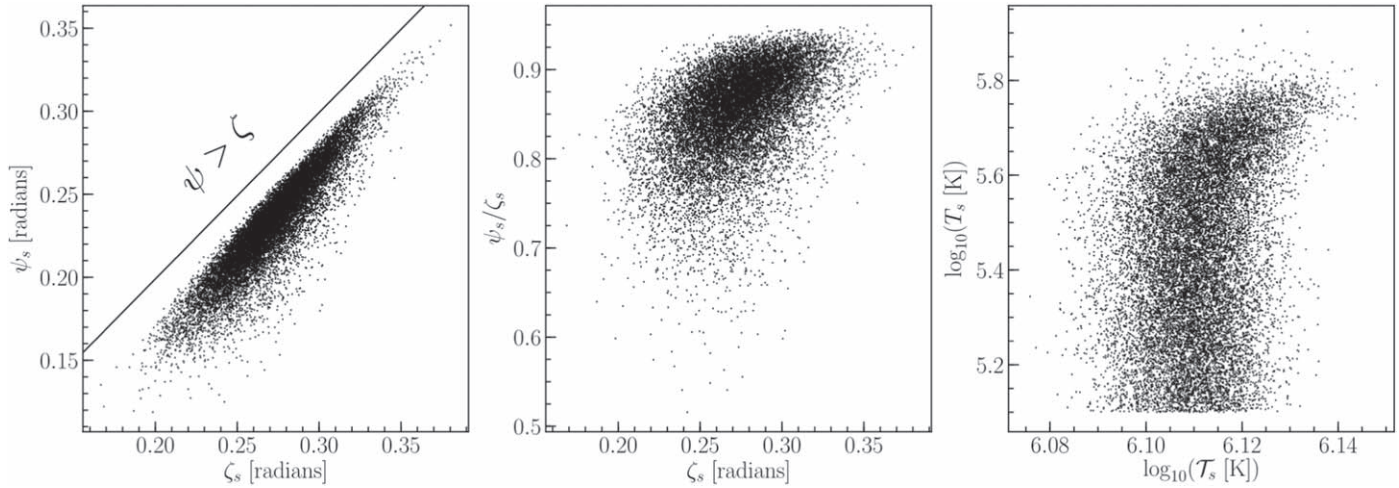

Figure 8. As for Figure 7, conditional on CDT−U, we display points (nested samples) $\{\boldsymbol{\theta}_k\}$ that reported (background-marginalized) likelihood function values $L(\boldsymbol{\theta}_k) > L'$, where $\max_k \ln L(\boldsymbol{\theta}_k) - \ln L' \lesssim 13$. We display these points in a selection of two-dimensional spaces to focus on the secondary (higher-colatitude) hot region. The secondary region exhibits a larger angular extent on the stellar surface than the primary region and, in contrast, we are statistically sensitive to the topology of the secondary region while being statistically insensitive to the additional complexity offered by a CDT cap over a CST cap. In other words, we do not strongly require the material in the hole to radiate when considering the *NICER* waveband.

to region $\mathcal{A}$ above. Degeneracy of types I and IV arose and dominated in prior mass (and by extension, in posterior mass). The nonlinear degeneracies in the native nested-sampling space also suppressed the efficiency of the sampling processes. We therefore now have grounds to design and apply models that break the parity in complexity between hot regions. We redistribute complexity from the primary region to the secondary region, such that the dimensionality of the parameter space does not exceed that of CDT−U. In Appendix D we define models that preserve parity between regions—a brief continuation of the above scheme for model extension.

In Sections 2.5.5 and 2.5.6 we explain how the complexity of a solitary hot region is extended. In Section 2.5.7 we briefly summarize and provide diagrams of the relevant models formed from two disjoint hot regions with unequal complexities. In Section 2.5.7 we then provide diagrams summarizing the relationships between all models considered in the scope of this work.

2.5.5. Eccentric Single- and Dual-temperature Regions (EST and EDT)

Parameterization. We extend the CST and CDT hot-region types by not requiring the ceding and superseding members to be concentric. For a CST region, the superseding member was a hole in a ceding member, devoid of radiating material; for a CDT region, the superseding member was similarly a hole, but filled with radiating material. For both CST and CDT regions, the *non-superseded* subset of the ceding member was a radiating annulus with concentric inner- and outer-boundaries; now the radiating annulus is *eccentric*. Further we require—via the joint prior support—that the ceding member is strictly a superset of the superseding member; in other words, the superseding member is a hole in the ceding member, again forming an annulus or ring.[60]

We illustrate these hot-region structures in the topmost panels of Figure 9; we also alluded to such configurations in Figure 5 and in Sections 2.5.2 and 2.5.3. Hereafter we use the aliases EST and EDT, respectively parsed as *Eccentric-Single-Temperature* and *Eccentric-Dual-Temperature*.

An additional two parameters must be defined to specify the eccentricity as the great-circle segment separating the centers of the members: the magnitude and direction of the offset. To do so, we consider a spherical coordinate basis such that: (i) the polar axis subtends an angle Θ to the stellar rotation axis; (ii) the southern rotational pole subtends angle $\pi - \Theta$ to the polar axis, and has azimuth *zero* (such that the northern rotational pole has azimuth π radians). The azimuth of the annulus center, φ, in the rotated basis uniquely selects the great circle and direction of the offset (depicted in Figure 9). We choose define the magnitude of the angular separation as a fraction ε of the *difference* $\psi^+ := \zeta - \psi \equiv \zeta(1 - f)$ between the angular radii of the hole and annulus (see Section 2.5.3). The magnitude of the angular offset between the centers—equal to the colatitude of the annulus center in the rotated spherical coordinate basis defined above—is thus $\vartheta := \varepsilon\psi^+$.[61]

Degeneracy. In a similar vein to the CDT parameterization, optimality of the parameterization for an EDT region should be considered. By this point in the modeling process, we have a handle on a region-by-region basis whether the additional complexity offered by an EST or EDT region is a justified modeling step. Nevertheless, one can construct the sampling space using the minimum number of parameters need to fully control the superseding member. The parameter vector for the regions is thus of the form $\boldsymbol{\nu} = (\psi, \Theta, \phi, \zeta, \varepsilon, \varphi)$, and the mapping from the native sampling space to the parameter space is formed in part using the marginal density $p(\psi)$ and the conditional density $p(\zeta \mid \psi)$.

[60] The hole reduces to a point at the boundary of the prior support for an EST region, while for a EDT region the hole is not permitted to reduce to a point. Also note that because the ceding and superseding members are circular, their boundaries may intersect at a maximum of one point.

[61] An alternative is to parameterize in terms of the fraction of the sum of their angular radii such that $\vartheta := \varepsilon(1 + f)\zeta$; it is this combination that is the limit of zero partial overlap when the superseding member lies partially exterior to the ceding member (see Section 2.5.6). In order to impose that the superseding member is a hole in the ceding member, we would then require the support constraint that $\varepsilon(1 + f)\zeta + f\zeta \leqslant \zeta$ and thus that $\varepsilon \leqslant (1 - f)/(1 + f)$.

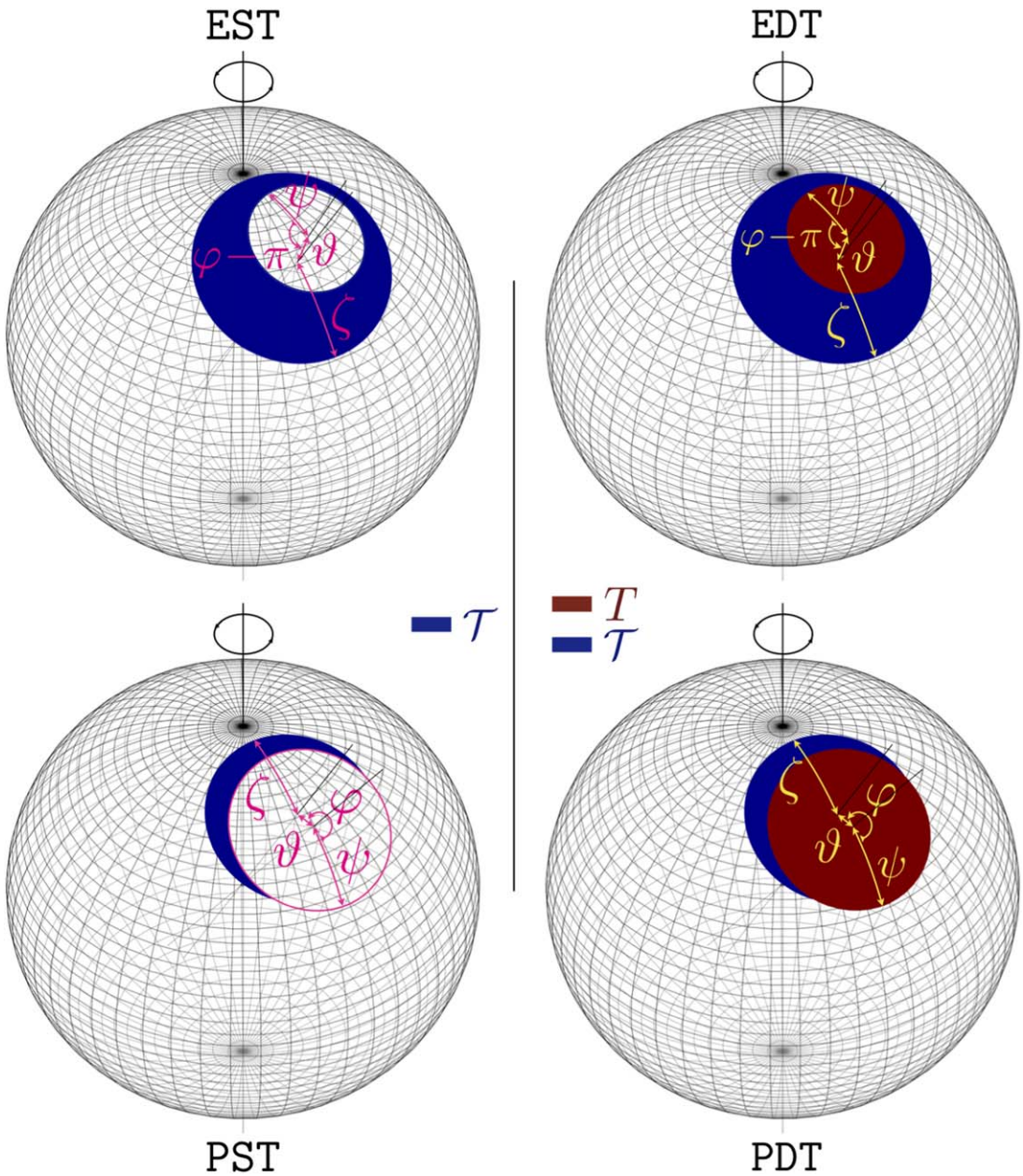

Figure 9. Schematic diagrams of solitary hot regions beyond the CST and CDT complexity levels, as described in Sections 2.5.5 and 2.5.6. We delineate the boundary of the superseding member for the PST region.

Priors. As for the CDT type of region (Sections 2.5.3), joint prior support is such that the material within the annulus may exhibit an effective temperature $\mathcal{T} < T$ or $\mathcal{T} \geqslant T$. Moreover, when an EST/EDT region shares the stellar surface with another hot region $\mathcal{C}$, the joint prior support does not include configurations wherein $\mathcal{C}$ overlaps with the ceding (annular) region of the EST/EDT region. Note that a continuous set of coordinate singularities exist in the mapping $(\varepsilon, \varphi) \mapsto (\Theta', \phi')$, where the coordinates (Θ', ϕ') are the colatitude and azimuth of the center of the annulus in a spherical coordinate system with polar axis defined as the stellar rotation axis. The mapping is singular for all points $(\varepsilon, \varphi) = (0, \varphi)$ at the boundary of the prior support, such that by defining a finite joint prior density at such singular points the joint density at $(\Theta', \phi') = (\Theta, \phi)$ diverges. We consider the spherical coordinates (ϑ, φ) to be more natural for increasing complexity and thus we display density functions with respect to (ε, φ) to avoid difficulty in representing density functions accurately.

2.5.6. Protruding Single- and Dual-temperature Regions (PST and PDT)

Parameterization. We extend the EST and EDT hot-region types by not requiring the ceding member be a strict superset of the superseding member. In other words, the superseding member need not be a hole in the ceding member, leading to radiating component with the topology of a ring. If the superseding member is *not* a hole nor a point, and the boundaries of the ceding and superseding members intersect at two points, then the *non-superseded* subset of the ceding member is simply connected *and* has a non-circular boundary: it cannot therefore always be considered an annulus, and can

assume a crescent- or arc-like morphology.[62] Note that we do not increase complexity here by incrementing the dimensionality of parameter spaces defined for the EST and EDT regions. We nevertheless choose to distinguish between EST/EDT and PST/PDT in view of a symmetry being broken: that of the circularity of the outer boundary of the region. The superseding member can also now subtend the larger (coordinate) solid angle at the stellar origin, and in the limit that the member boundaries touch at a point, the hot region is simply connected, with a boundary that is either circular or union of two circular subregion boundaries.

We illustrate these hot-region structures in the bottommost panels of Figure 9; we also alluded to such configurations in Sections 2.5.2 and 2.5.3. Hereafter we use the aliases PST and PDT, respectively parsed as *Protruding-Single-Temperature* and *Protruding-Dual-Temperature*: the superseding member can protrude from the ceding member, for parameter vectors within the joint prior support.

We modify the EST (and thus EDT) parameterization in order to permit the superseding member to: (i) protrude[63] from the ceding member; and (ii) subtend a larger (coordinate) solid angle at the stellar origin than the ceding member. Let us denote the angular radius of the *largest* member as ξ. Let us then consider the interval $f \in [\epsilon_f, 2-\epsilon_f]$, where ϵ_f is some small number: when $f \leqslant 1$ the angular radii of the superseding and ceding members are, respectively, $\psi = f\xi$ and $\zeta = \xi$, as in the EDT variants. However, when $f > 1$, the angular radii of the superseding and ceding members are, respectively, $\psi = \xi$ and $\zeta = (2-f)\xi$. The angular radii are thus piecewise in f, and are continuous at the transition point $f = 1$. It follows that the interval for the angular radius ξ of the largest member considered in the previous variants can be maintained, while the member that is largest switches. If one varies f through the interval $f \in [\epsilon_f, 2-\epsilon_f]$ while all other parameters are fixed, the solid angle subtended by the superseding member increases to match that of the ceding member; at the $f = 1$ transition the superseding member stops expanding, and for f increasing beyond unity, the solid angle subtended by the ceding member decreases.

We choose to define the magnitude of the angular separation as a fraction[64] $\varkappa$ of a combination of the angular radii of the superseding and ceding members. For $f \leqslant 1$, the combination is the *sum* of the angular radii of the superseding and ceding members, $\psi = f\xi$ and $\zeta = \xi$, respectively, for $f \leqslant 1$: $\vartheta(\xi, f \leqslant 1, \varkappa) := \varkappa(1+f)\xi = \varkappa(\psi+\zeta)$. It is this combination that is the limit of zero partial overlap between the members when the superseding member lies partially exterior to the ceding member. We define the prior support for $\varkappa$ simply as the unit interval: we permit only configurations wherein the members partially overlap (including at a point), with maximum separation $\vartheta = \psi + \zeta$.

We require the angular separation ϑ as piecewise in f: for $f > 1$, we let $\vartheta(\xi, f > 1, \varkappa) := \psi - \zeta + 2\varkappa\zeta = (f-1)\xi + 2\varkappa(2-f)\xi$. Note that the minimum separation here with respect to $\varkappa$, for fixed ξ and fixed f, is $\vartheta = \psi - \zeta$, equivalent to the transition from the ceding member being *partially superseded* to being *wholly superseded*; the minimum separation for the subinterval $f \leqslant 1$ is zero. The maximum separation (with respect to $\varkappa$ for fixed ξ and fixed f) is $\vartheta = \psi + \zeta$, which is equivalent to the expression for the maximum separation for $f \leqslant 1$. Note that at the transition $f = 1$, the piecewise components of $\vartheta(\xi, f, \varkappa)$ continuously match at a value of $2\varkappa\xi$.

The coordinates of the center of the ceding member remain written in terms of those of the center of the superseding member. We cannot define a pair of coordinates as being associated with the *larger* member (in the same vein that ξ is directly associated with the larger constituent) without generally introducing a discontinuous transition with respect to the configuration of the members at $f = 1$.

Note that the above prescription is equivalent in principle to defining a binary discrete parameter (with associated uniform prior probability mass function) that controls which member ξ in turn directly controls, and thus which member f directly controls, where the upper bound of the prior support of f remains as unity. In this equivalent alternative, the piecewise definitions of ϑ and the angular radii are required, but are recast with respect to the binary parameter. However, extension of the support of a continuous parameter to include these configurations eases posterior computation.

Degeneracy. The two clear exact degeneracies beyond those considered above[65] are: (i) equal temperatures where the superseding member is wholly enclosed by the ceding member, leading to three-dimensional degeneracy in the subspace of (f, $\varkappa$, φ); and (ii) the superseding member *wholly supersedes* the ceding member ($\varkappa = 0$ for $f \geqslant 1$), leading to a two-dimensional degeneracy in the (f, φ)-subspace. There are also discrete member-exchange degeneracies for a given region when the temperatures are equivalent.

Priors. The joint prior support is again such that the material within the ceding member may exhibit an effective temperature $\mathcal{T} < T$ or $\mathcal{T} \geqslant T$. Moreover, when a PST/PDT region shares the stellar surface with another hot region $\mathcal{C}$, the joint prior support could be defined to exclude configurations wherein $\mathcal{C}$ overlaps with the PST/PDT region. We note that due to choice of parameterization and prior support, the prior density function for the angular separation between the ceding and superseding members has changed from the corresponding EST/EDT density function.

2.5.7. Overview of the Model Space

In Figure 10 we provide a diagram of the relationships between the models—spanning the discrete model space—that impose equal complexities. A subset of these models are based on the extensions defined in Sections 2.5.5 and 2.5.6; diagrams of EDT-S, EDT-U, PDT-S, and PDT-U are given in Appendix D because we do not compute posterior distributions that are jointly conditional on these models and the event data.

We also define models wherein an ST region shares the stellar surface with a higher-complexity hot region. We provide diagrams of ST+EST and ST+PST in Figure 11. We provide diagrams of ST+EDT and ST+PDT in the online figure set associated with Figure 11. In Figure 12 we illustrate the relationships between these models. Note that these models

[62] It is now particularly clear why we supplant the *hole* and *annulus* descriptors with those in terms of simply connected *members* and evaluation precedence as noted in Section 2.4.1: one member is distinguished from the other simply by which takes precedence when evaluating radiation intensities along a ray (null geodesic) connecting the stellar surface to a distant observer.

[63] Or partially overlap with the ceding member.

[64] Equal to the colatitude of the center of the ceding member in the rotated spherical coordinate basis defined above.

[65] The degeneracies considered above should at this stage be of little to no concern if computation of the simpler (nested) models has *not* indicated that the complexity thus far introduced is unhelpful.

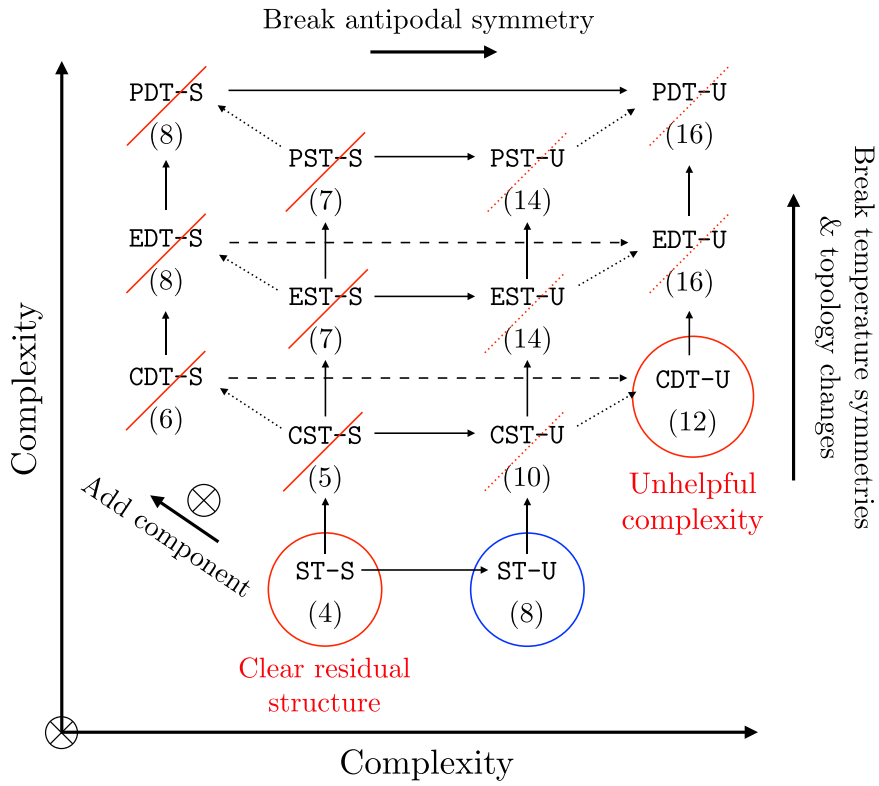

Figure 10. Diagram of relationship between a subset of models $\{\mathcal{M}\}$, where each $\mathcal{M} \subset \mathscr{M}$ and $\mathscr{M}$ is the model space as defined in Section 2.3. The models are ST-S and ST-U (Section 2.5.1), CST-S and CST-U (Section 2.5.2), CDT-S and CDT-U (Section 2.5.3), EDT-S and EDT-U (Appendix D), and PDT-S and PDT-U (Appendix D). The integers in parentheses are the number of continuous parameters controlling the hot regions; to obtain the total number of parameters constituting the sampling space for each model, add 8 to these numbers—corresponding to the vector (M, $R_{\rm eq}$, i, D, $N_{\rm H}$, α, β, γ) shared between all models. Prior predictive complexity increases with dimensionality and/or prior support expansion; complexity increments are achieved by breaking symmetries, allowing the morphology and topology of the radiating regions to change, and adding a second component with parameterized temperature. Solid arrows between model nodes delineate a nested relationship between models: the model at the tail is nested within the model at the head, the latter of which has greater complexity. Adding arrows head-to-tail at a node conserves such relationships—i.e., the model at the tail of the resultant arrow is nested within the more complex model at the head. The dashed arrows are in the background (visualize a third dimension of the graph as indicated by $\otimes$). The dotted arrows between single- and dual-temperature models indicate that the nested relationship is weaker: in the single-temperature models no material (or material with zero temperature) fills the superseding member and thus no signal is generated, whereas for the dual-temperature models we fill the superseding member with material of finite temperature $T \gtrsim \mathcal{O}(10^5)$ K, and a signal is thus physically generated and computed. The blue ring indicates the simplest model that can generate data that is visually comparable to the real data set (refer to Section 3 for discussion). The red annotations are to denote where we learn that a given model is not performing adequately or has unhelpful complexity. A strikethrough denotes that a model was not applied: (i) a solid strikethrough denotes that the model is considered as being incapable of generating synthetic event data that resembles the real *XTI* event data, based on performance of a simpler model with which some degree of symmetry is shared; (ii) a dotted strikethrough denotes that the model is considered to include unhelpful complexity for at least one hot region.

with unequal complexities exist within the scope of those models present in Figure 10.

When the hot regions are distinguishable by complexity level we can label the regions as *primary* and *secondary* according to that complexity: let the higher-complexity region be *secondary*. Our application of CDT-U suggested that evolving the lower-colatitude (primary) region from ST to CDT did not yield any improvement in describing the event data; the lower-colatitude region in this case describes the pulse component in the event data (Figure 1) that peaks at ∼0.5 rotational cycles. Moreover, for both ST+CST and ST+CDT, the secondary region is formed (not necessarily by trivial union) from two members—one or both of which radiate—that are concentric and thus share a single colatitude parameter. Therefore, for ST+CST and ST+CDT it is reasonable to retain the constraint on the prior support that $\Theta_p \leqslant \Theta_s$—i.e., that the colatitude of the *center* of the ST region is at most the colatitude of the *center* of the CST/CDT region.[66]

Notably, we cannot consider a model with an EST/EDT or PST/PDT region as nested within CDT-U, and there is thus not an obvious ordering of regions in colatitude that can be applied based on the CDT-U posterior information. We therefore do not impose such a constraint on the prior support, meaning that the order of the regions in model names ST+EST, ST+EDT, ST+PST, and ST+PDT does *not* indicate an order in colatitude.

Based on the CDT-U posterior information we are now interested in models wherein we couple the lower-complexity (primary) region to the pulse component in the event data (Figure 1) that peaks at ∼0.5 rotational cycles. We consider a practical reason to identify a particular hot region with a particular pulse component in the phase-folded event data: to ease posterior computation by focusing sampling resolution in the vicinity of such configurations. If we do not opt for this coupling, a local posterior mode (or modes) can absorb sampling resolution, despite contributing relatively low posterior mass and thus being ultimately uninteresting for both parameter estimation and evidence estimation; such a local mode is characterized by the ST region describing the pulse component at approximately zero rotational cycles (Figure 1).

[66] A model wherein the CST or CDT region is at *higher* colatitude can be continuously connected to to this model by simply by forgoing the constraint on the prior support that $\Theta_p \leqslant \Theta_s$.

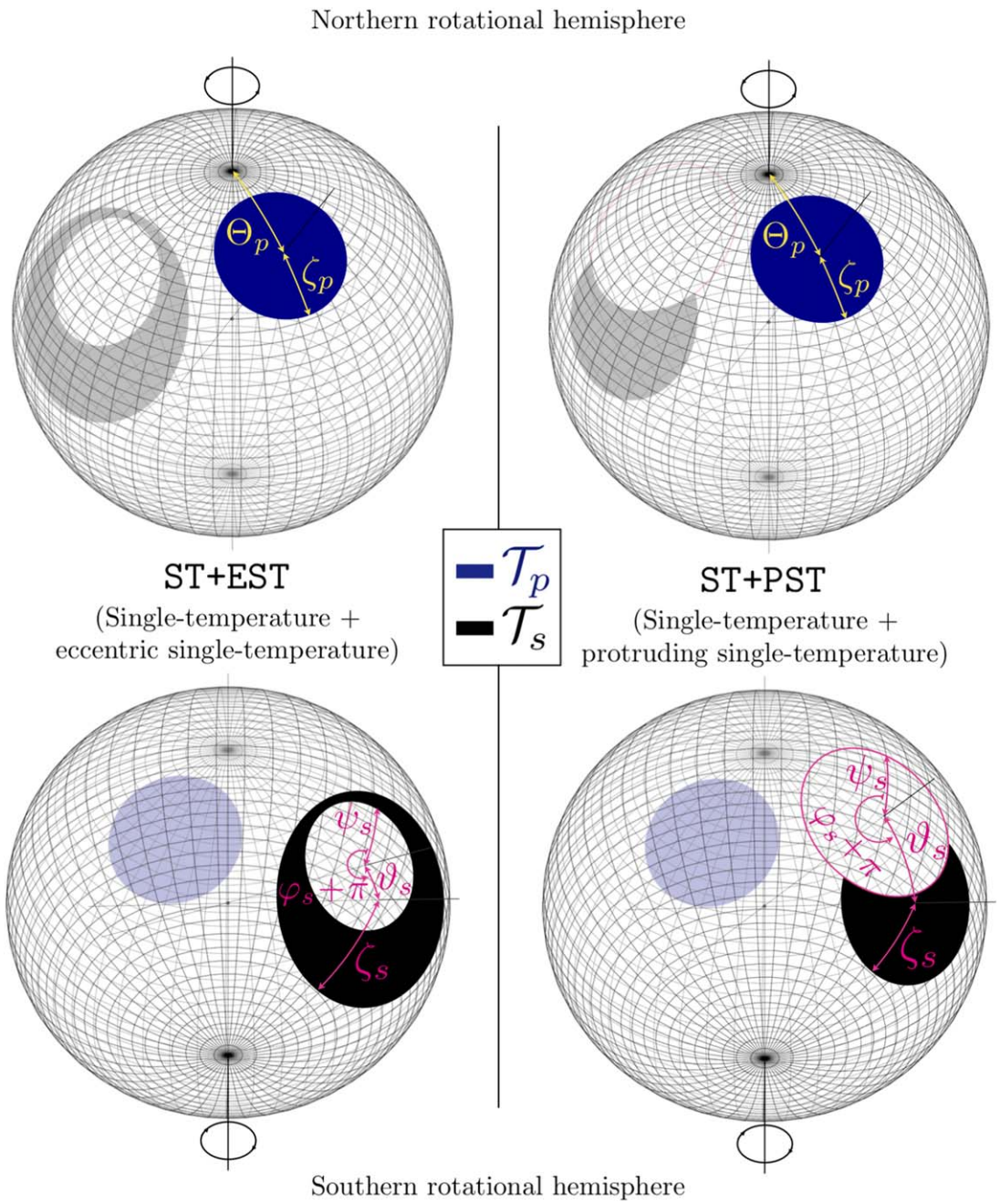

Figure 11. Schematic diagrams of models wherein an ST region shares the stellar surface with a higher-complexity EST or PST region.

Mitigating such an effect may then require additional active points and/or activation of the mode-separation MultiNest sampling variant (refer to Appendix A).

To intentionally couple the hot regions to pulse components, we restrict the prior support for the ST region to an interval $\phi_p \in [a, b]$ where $(b - a) < 1$ (such that the boundary of the support is not periodic with respect to ϕ_p). Remarkably, while such restriction may not affect parameter estimation,[67] the evidence does clearly depend on changes to the prior support, especially for weakly informative priors whose density does *not* fall to negligible values at the changing boundary of the support. It follows that in order to compare models based on evidences, we should obtain a lower bound on the evidence of, e.g., ST+EST, by accounting for the increased prior mass in the posterior mode due to contraction of the ST phase support.

We reserve the remaining prior implementation details for Appendix B.2.

3. Inferences

In this section we provide (posterior) summary information about each model applied and compare them. Tables giving numerical information for all but one model are available in Appendix C. Numerical files associated with the nested sample sets may be found in the persistent repository of Riley (2019). As regards model comparison, there is no clear maximally optimal measure for relative model performance (and certainly not for performance in an absolute sense). We deem ST+PST

[67] Provided that (i) the dominant posterior mode corresponds to a ST primary region, and (ii) that the marginal prior density $p(\phi_p)$ remains weakly informative relative to the likelihood function

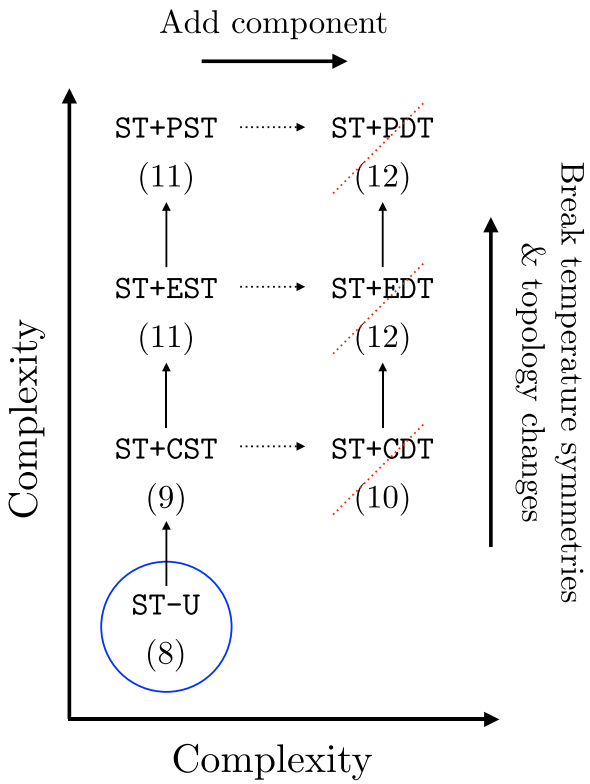

Figure 12. Diagram of relationship between dual-hot-region models with unequal complexities, constructed via extension of ST-U. These models exist within the scope of those displayed in Figure 10.

to be superior considering the following mixture of measures, both quantitative and qualitative.

Posterior predictive performance. Our crude graphical posterior predictive checking procedure suggests that even ST-U can generate synthetic event data that is structurally commensurate with the *XTI* event data—at least in the channel subset [25, 300). For reference, see Appendix A.2.1, together with Figure 13 and the associated figure set. Absent are obvious systematic differences in the (Poisson) standardized residuals over phase-channel intervals. Nevertheless, one could study the residual differences for modeling background event arrival processes, instrument operation, or noise properties of the event data—e.g., whether the event arrival processes could detectably deviate from being Poissonian (see also Figure 23). For the purpose of constructing a model that can simply generate event data that is similar to PSR J0030+0451 under visual inspection, ST-U is competitive with all higher-complexity models.

Visualization. In Figures 14 and 15 (and the associated online figure sets) we display, in posterior-expected form, various signals (derived quantities) generated by the hot regions; the reader can thus get a handle on the source contribution to the model displayed in Figure 13 (center panel). In Figure 16 we display the model-to-model evolution of the posterior-expected count-rate signals generated. To aid visualization, we provide in Figures 17 and 18 schematics of the hot regions on the surface that generate the aforementioned signals.

Evidence. We report evidence estimates in Tables 2 through 6. For the family of models that we have considered, the evidence strongly favors (adopting the guidelines of Kass & Raftery 1995) a model wherein antipodally reflection symmetry is *not* imposed; in other words, ST-S, and by extrapolation any such model with the -S extension (refer to Figure 10), is considered to be strongly disfavored. Moreover, for all computed models that do not impose antipodal reflection symmetry, posterior modes contain configurations characterized by both hot regions being located in the same rotational hemisphere—the opposite rotational hemisphere to the Earth direction. We also conclude that at least one region should be modeled with more complexity than offered by a single-temperature simply connected circular (ST) region; in other words, there is deemed sufficient evidence to disfavor ST-U. However, there is insufficient evidence to resolve between CDT-U, ST+CST, ST+EST, and ST+PST, especially when considering the evidence not as a scalar estimator, but as a random variable with a (simulated) distribution for estimating error intervals. The estimated error intervals (defined to contain 90% of evidence estimates based on nested sampling process realizations) typically overlap for these models, while the expected log-evidences are within $\sim$2 units.

Kullback–Leibler divergence. We report global and marginal divergence estimates in Tables 2 through 6, and in Figures 19–21. Information gain from prior to posterior is defined as a divergence integral over some subset of parameter dimensions: it is a non-negative real scalar measure of the difference between normalized density functions (see Appendix A.2.4 for a more detailed description). The larger the number, the larger the information gain, while a minimum divergence of zero indicates that the density functions are identical. The global information gain—the divergence integral over all dimensions—is comparable for all models. Based on comparison of parameter-by-parameter divergence estimates, and visual comparison of the marginal prior and posterior density functions hence summarized, we consider the joint prior distribution to be weakly informative in the context of the the likelihood function for most source parameters; the main exception is the distance, which is strongly prior-dominated. The other parameters whose marginal posterior distributions are entirely prior-dominated are the two of the instrument parameters, α and γ (refer to Figures 21 and 22). These informative prior distributions are however shared by all models, as are the improper flat prior density functions described in the background treatment (refer to Section 2.4.3). Note that more information is gained about the instrument parameter β—the weighting factor between two response matrices—and we can conclude that the instrument manifests a posteriori as a mixture weighted appreciably toward the nominal response for all posterior computations; we reserve discussion on why the instrument calibration may not be accurate for Section 4.1.3.

Likelihood function. We generally refrain from reporting parameter vector point measures such as the maximum a posteriori vector or the maximum likelihood vector for the purpose of formal quantitative model comparison (see Appendix A.2.1 for reasoning). We note, however, that the background-marginalized likelihood function exhibits the largest values in any posterior *typical set* (by $\sim$4 natural logarithmic units), across all models, for ST+PST. It follows that a subset of the additional complexity introduced by ST+PST is helpful—and suggestive of avenues for future modeling—while a subset of the complexity is unhelpful, leading to only commensurate evidence relative to ST+EST. In other words, the (background-marginalized) likelihood

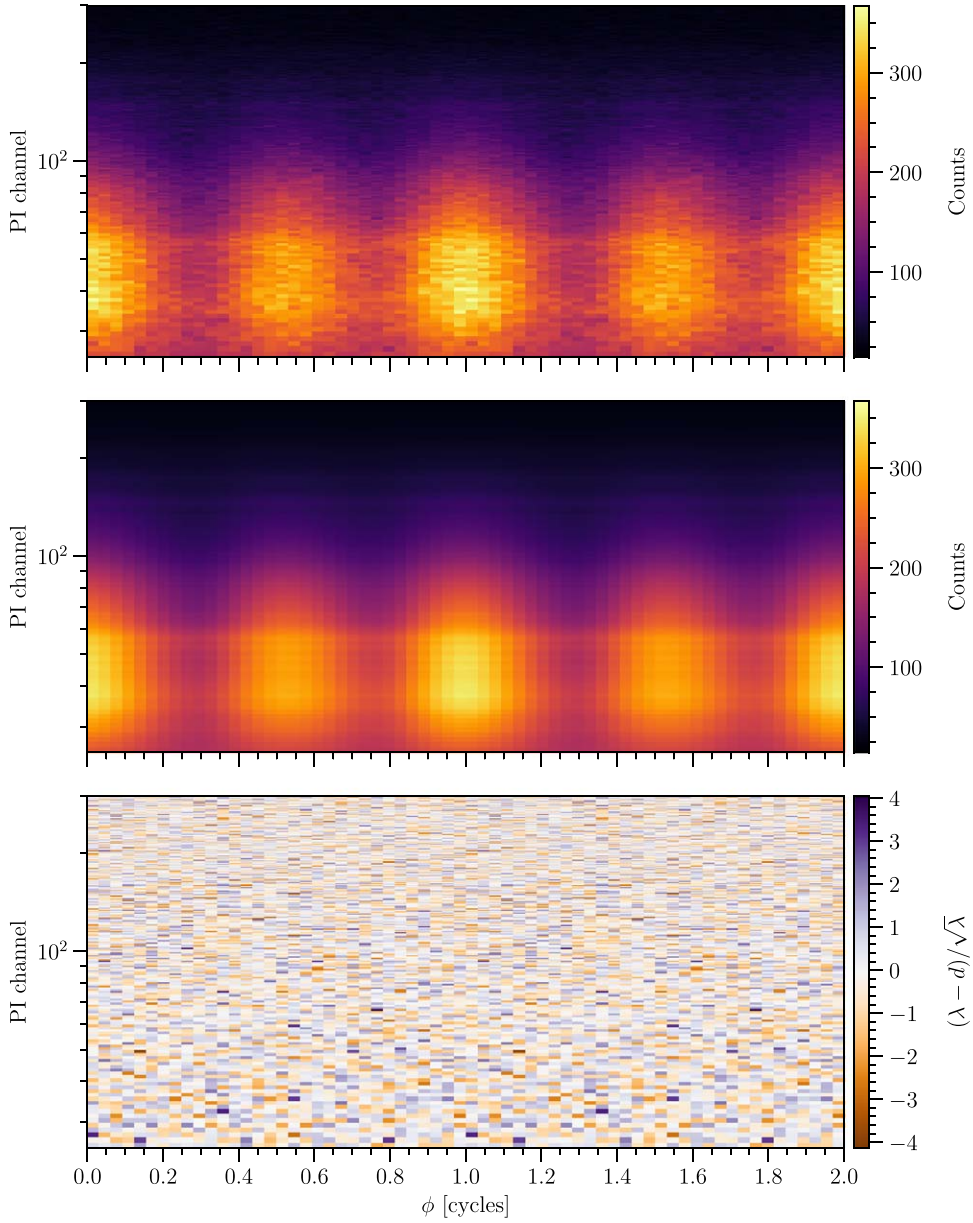

Figure 13. Count data $\{d_{ij}\}$, posterior-expected count numbers $\{\lambda_{ij}\}$, and (Poisson) residuals for ST+PST. Note that we split the count numbers in the upper two panels over two rotational cycles, such that the information on phase interval $\phi \in [0, 1]$ is identical to the information on $\phi \in (0, 2]$; our data sampling distribution, however, is defined as the (conditional) joint probability of all event data grouped into phase intervals on $\phi \in [0, 1]$. We display the standardized (Poisson) residuals in the bottom panel: the residuals for the rotational cycle $\phi \in [0, 1]$ were calculated in terms of all event data on that interval (as for likelihood definition), and simply cloned onto the interval $\phi \in (1, 2]$. In Appendix A.2.1 we elaborate on the information displayed here. The complete figure set (6 images) is available in the online journal, for the ST+PST, ST+EST, ST+CST, ST−U, ST−S, and CDT−U models.

(The complete figure set (6 images) is available.)

function, over a subset of the additional prior support, is at most commensurate with the prior expectation of the likelihood function for the other models; globally, the model does not therefore improve as measured solely by the evidence.

Tractability and complexity. Posterior computation accuracy and reproducibility are generally higher as the models become less complex. The ST+CST, ST+EST, and ST+PST models redistributed complexity based on the performance of CDT−U in pursuit of *helpful* complexity and thus higher-efficiency resource consumption. We demonstrate in Figure 16 that effectively all of the improvement of CDT−U over ST−U is captured by the intermediary ST+CST. As a consequence, ST+PST in particular exhibited substantial improvement in regard to the (background-marginalized) likelihood function in the posterior mode, while reducing the number of parameters by limiting the complexity of one hot region; it follows that it should be clearly preferred over CDT−U.

We remark that although the complexity of one hot region was ultimately limited at the ST-level, this does not mean that the signal generated by that region is fixed or performs maximally in some absolute sense. Indeed, the signal generated by one region cannot be decoupled (as stated in Section 2.5.4) from the signal generated by the other for the purpose of statistical inference: (i) the regions are restricted to exist on the

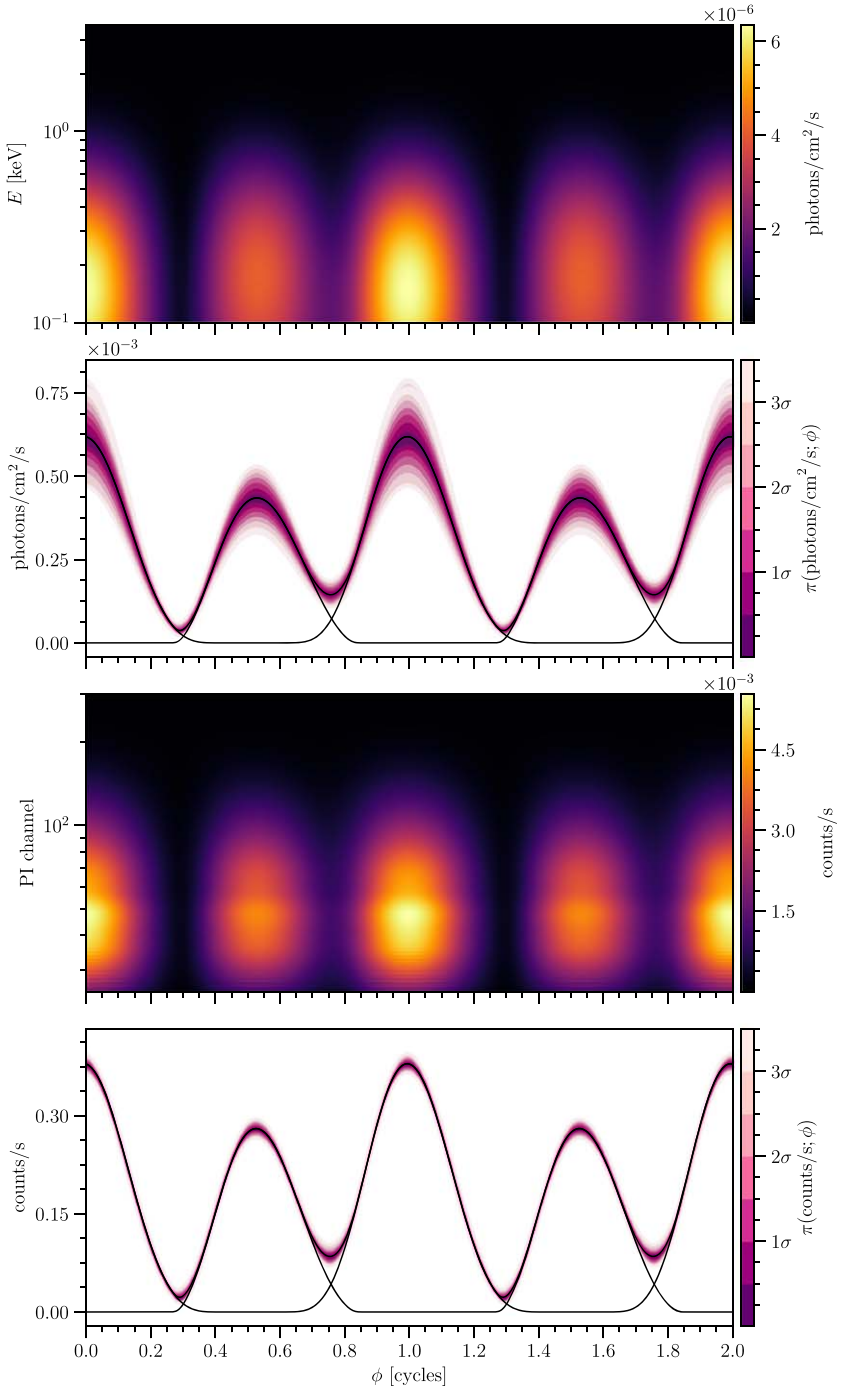

Figure 14. The posterior-expected signal for ST+PST, both incident on the instrument (top and top-middle) and as registered by the instrument (bottom-middle and bottom). The signal in the top panel has been integrated over the linearly spaced instrument energy intervals, and is effectively proportional to the specific photon flux. The black count-rate curves are the posterior-expected signals generated by each hot region separately, and in combination. We also represent the conditional posterior distribution of the incident photon flux (top-middle) and count-rate (bottom) at each phase as a set of one-dimensional highest-density credible intervals, and connect these intervals over phase via the contours; these distributions are denoted by π (photons cm^{-2} s^{-1}; ϕ) and π (counts s^{-1}; ϕ). Note that the fractional width of the credible interval at each phase is usually higher for π (photons cm^{-2} s^{-1}; ϕ) than for π (counts s^{-1}; ϕ) because of the variation permitted for the instrument model; in combination, the signal registered by the instrument is more tightly constrained. To generate the conditional posterior bands we apply the *X-PSI* package, which in turn wraps the fgivenx (Handley 2018) package. The complete figure set (6 images) is available in the online journal.

(The complete figure set (6 images) is available.)

same two-surface within an ambient spacetime; and (ii) for some subset of rotational phases, and for the configurations that report higher (background-marginalized) likelihood values, the images of both regions are simultaneously visible (a posteriori) and registered by the instrument in combination. As a result, upon examination of Figure 16, we observe that the posterior-expected count-rate signal generated by the ST region evolves along the ST−U, ST+CST, ST+EST, and ST+PST sequence,

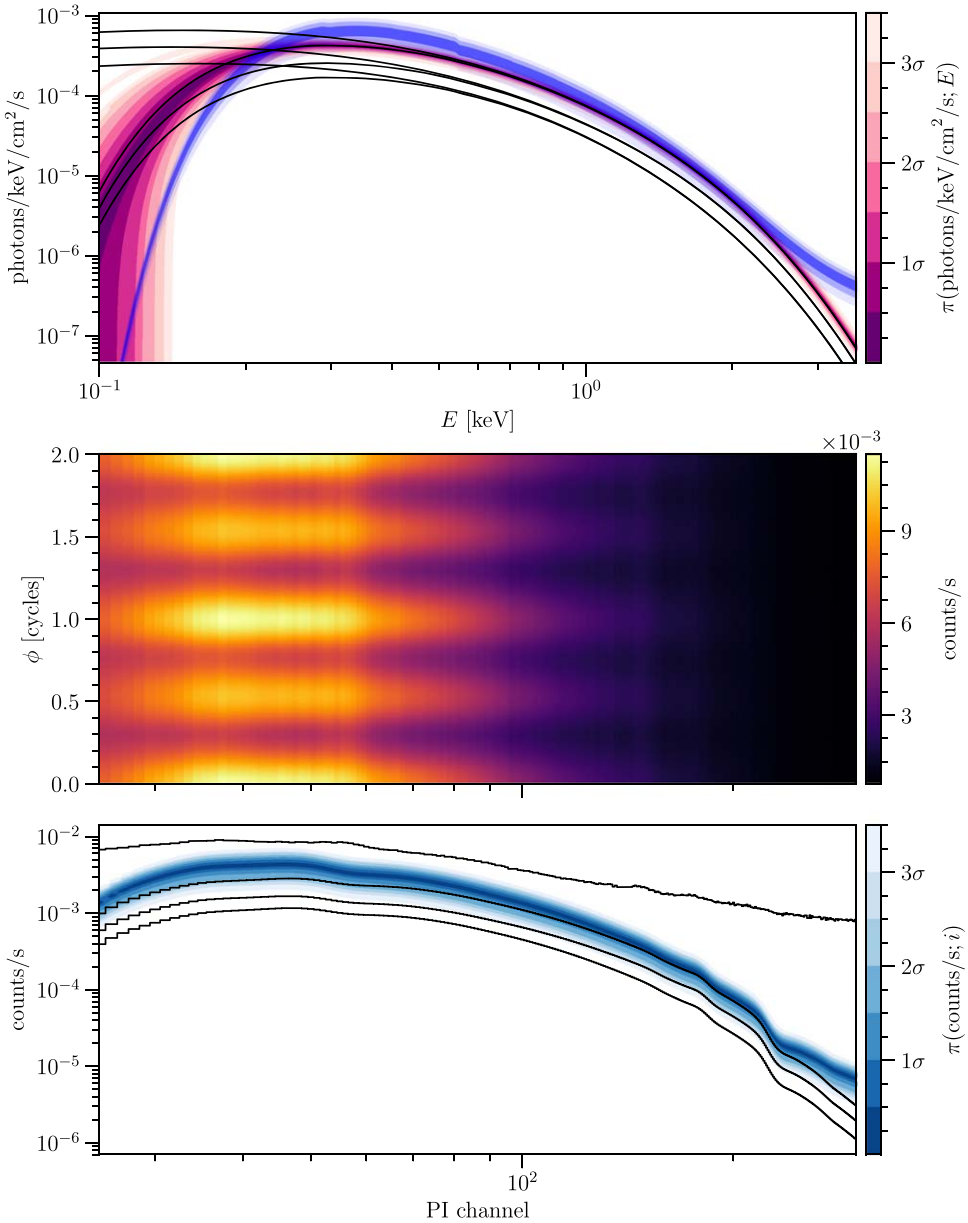

Figure 15. The posterior-expected spectrum for ST+PST, both incident on the instrument (top) and as registered by the instrument (middle and bottom). The black count-rate curves are the posterior-expected spectra generated by each hot region separately, and in combination. In the top panel we display as black curves the incident photon specific flux spectra both with and without interstellar absorption. We represent the conditional posterior distribution π (photons keV^{-1} cm^{-2} s^{-1}; E) of the absorbed incident photon specific flux at each energy as a set of one-dimensional highest-density credible intervals, and connect these intervals over phase via the contours (top); the energies displayed are those spanning the waveband of channel subset [25, 300). The credible intervals fan-out at the lowest energies because (i) conditional on event data for channel subset [25, 300), we are relatively insensitive to the details of the signal for $E \in [0.1, 0.2)$; and (ii) the interstellar attenuation factor is stronger the lower the photon energy, and thus the incident signal varies strongly as a function of the neutral hydrogen column density—a free parameter that operates as an exponent. In the top panel we overlay the incident absorbed spectrum inferred by Bogdanov & Grindlay (2009) based on a phase-averaged analysis of low-background *XMM* observations; the blue band denotes the estimated fitting uncertainty on this spectrum, at each energy, as a Gaussian with fractional standard deviation $\sigma/\mu = 0.15$, and the three opacity levels indicate intervals 1σ through 3σ. There is additional systematic *XMM* flux calibration uncertainty at the $\sim$10%-level that is not included here. In the middle panel we display the background-marginalized posterior-expectation of the source count-rate signal, plus the background count-rate terms that maximize the conditional likelihood function; the signal is equivalent to that displayed in the middle panel of Figure 13. In the bottom panel we display the posterior-expected count-rate spectra generated by the hot regions in combination and individually; we opt not to render the conditional posterior count-rate distribution for each channel because it is too narrow about the expected spectrum to be useful. Moreover, the topmost black step function is the phase-average of the center panel—it is effectively, but not exactly, the observed count-number spectrum divided by the total exposure time $T_{\rm exp}$. We combine the *XMM*-derived count-rate spectrum (and its associated uncertainty) with the marginal *NICER* instrument posterior on parameters α, β, and γ to simulate a conditional probability distribution π (counts s^{-1}; i) for the count-rate in the ith channel; these conditional distributions are connected via the contours in a manner congruent to the top panel described above. The complete figure set (6 images) is available in the online journal.

(The complete figure set (6 images) is available.)

both at phases where the non-ST region is invisible, and at phases where both regions are simultaneously visible. The reader should not therefore interpret our decision to limit the complexity of one region to ST as a statement that an ST region is forecasted to be sufficient for the purpose of future modeling efforts nor that exploration of different models will

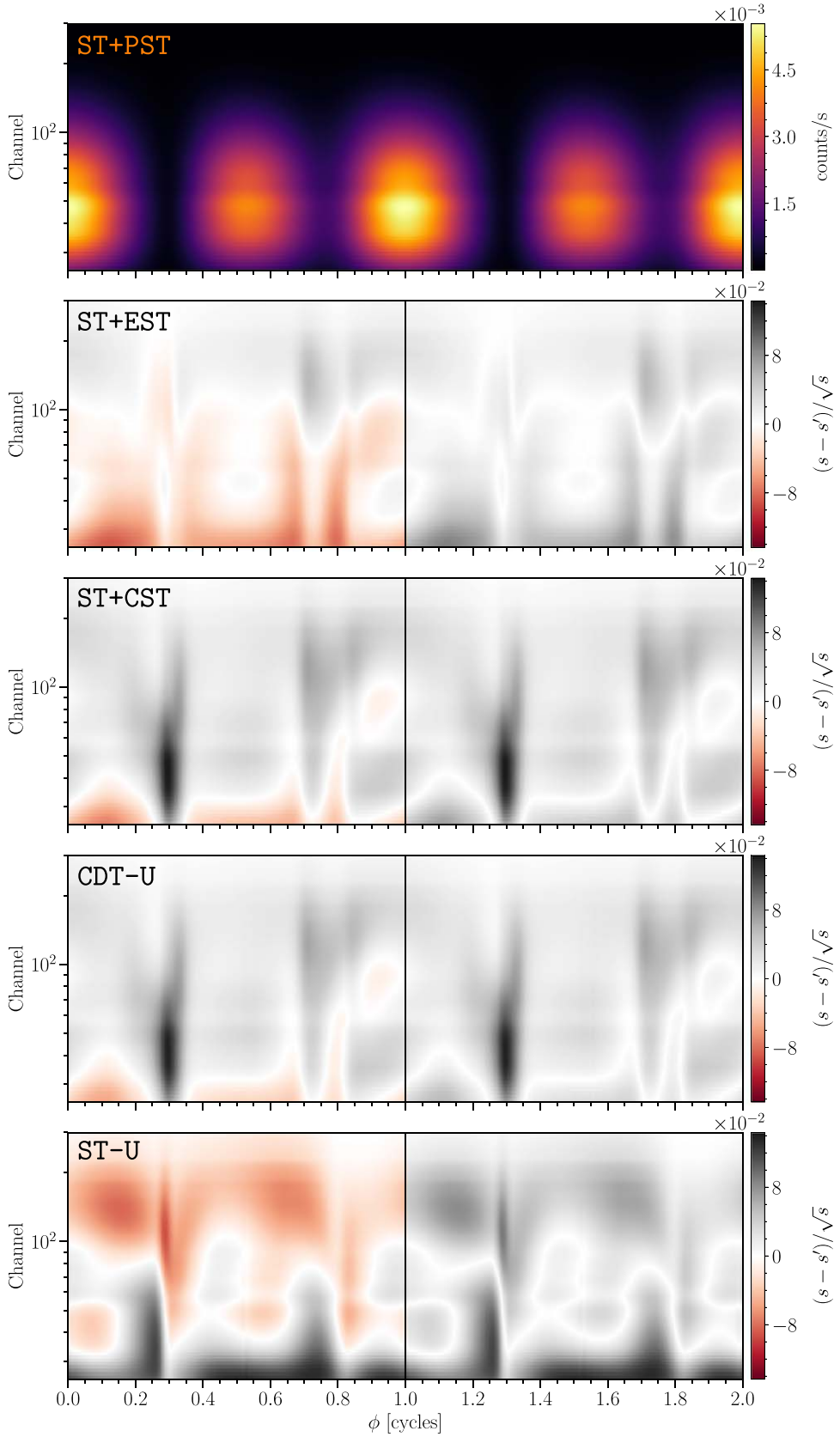

Figure 16. Top panel: the posterior-expected channel-by-channel source count rate signal generated as a function of rotational phase by ST+PST surface emission: $s \coloneqq \mathbb{E}_{\pi(\boldsymbol{\theta})}[g(\boldsymbol{\theta})]$, where $g(\boldsymbol{\theta})$ is a map from parameter space to the corresponding count rate signal. Others panels: the difference between signal s (for ST+PST) and the posterior-expected signal conditional on some other model: let $s' \coloneqq \mathbb{E}_{\pi(\boldsymbol{\theta})}[h(\boldsymbol{\theta})]$, where $h(\boldsymbol{\theta})$ is a map from parameter space to a count-rate signal conditional on some model other than ST+PST. The signals s and s' represent Poissonian arrival processes; we therefore opt to display the difference as $(s - s')/\sqrt{s}$, and scale the signals by $T_{\rm exp}/32$. A grayscale value is then locally representative of the absolute change (shown exclusively on the phase interval $\phi \in [1, 2]$), in units of the Poisson standard deviation, within a phase interval of width $1/32$ cycles as defined during event data pre-processing.

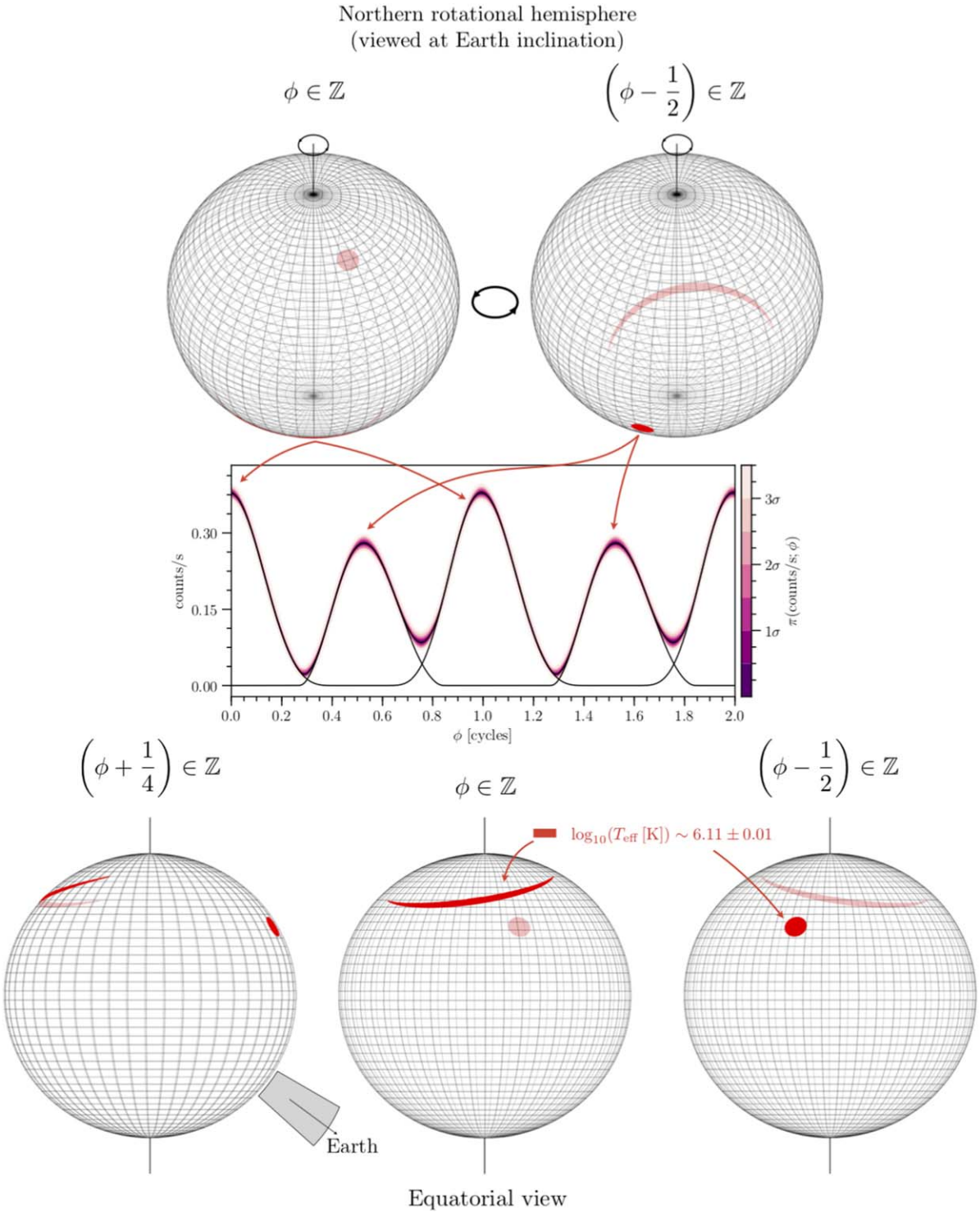

Figure 17. Top panel: schematic diagram of a surface heating configuration—and Earth inclination—representative of those corresponding to points in the ST+PST posterior mode. The configuration rendered here corresponds to the sample that reported the highest background-marginalized likelihood function value across all models (among the values reported by the set of all nested samples). We project the hot regions onto a (unit) sphere and view from the Earth inclination with no ambient gravitational field. The regions are constrained to exist in the same hemisphere, but with remarkably different morphologies. The hot regions are approximately equal in effective temperature and thus we define a new temperature symbol T_{eff} that is common to both. We also display the channel-summed count rate pulse generated by the source emission and indicate which region generates which component; we refer the reader to Figure 14, where this signal is also displayed, for more information. Bottom panel: note that we impose (via the prior support) that the Earth inclination lies within the northern rotational hemisphere, but an identical configuration (in terms of the physics that we consider and thus signal generation) is given via an equatorial reflection of both the Earth direction and radiating regions; we render this alternative configuration as viewed from the equatorial plane, and display the Earth inclination (but not azimuth) as the shaded angular interval bounded by the 16% and 84% quantiles in marginal posterior mass (see Figure 21 and Table 2 for the numerical interval).

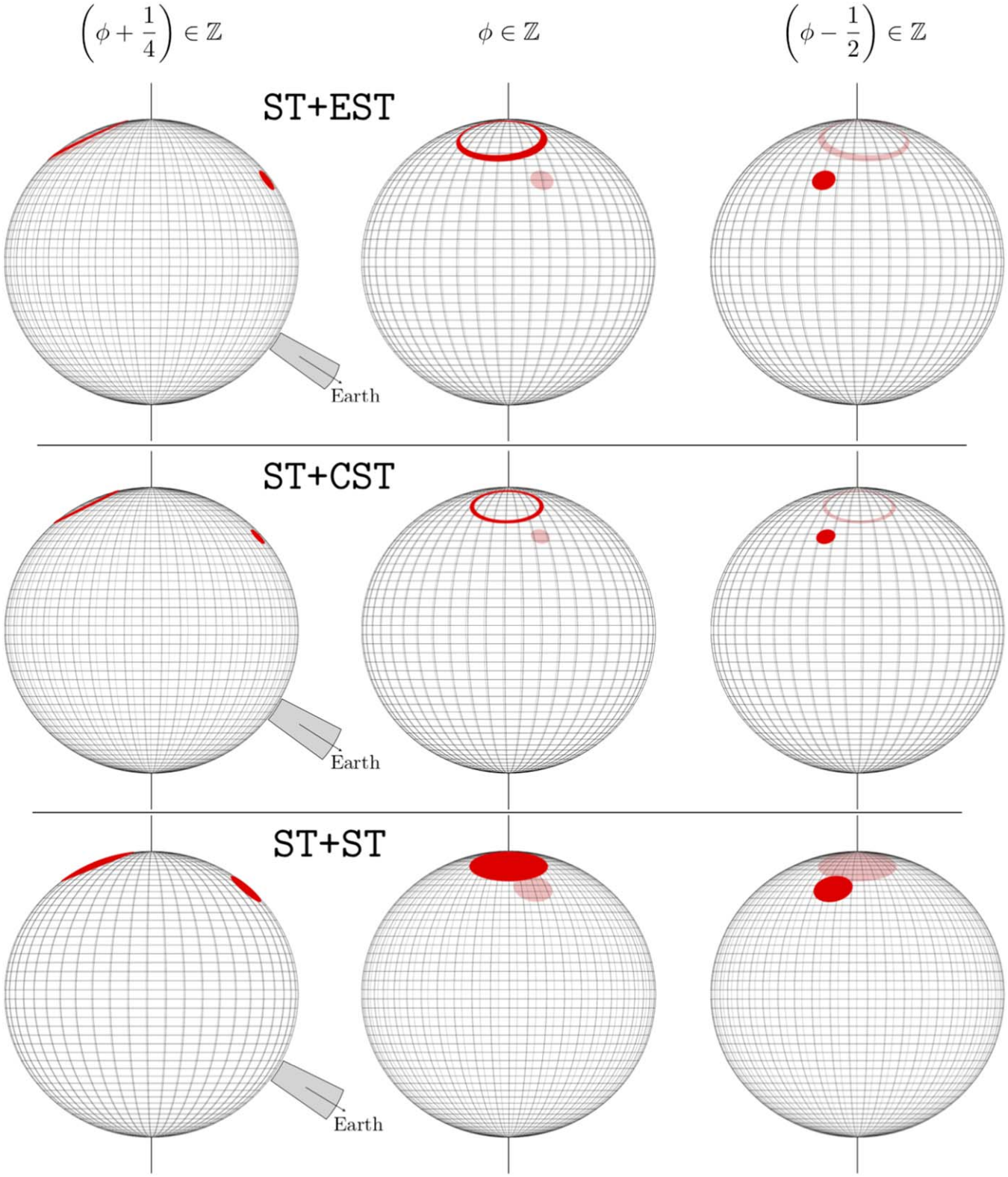

Figure 18. Surface heating configurations conditional on ST+ST (equivalent to ST-U), ST+CST, and ST+EST. For each model the hot regions correspond to the sample assigned the greatest posterior weight and is thus representative, being a draw from the posterior typical set. For each model we display the Earth inclination (but not azimuth) as an angular interval bounded by the 16% and 84% quantiles in marginal posterior mass.

fail to yield improvement; indeed, one could even consider models such as PST-U given that our modeling route for one region was based on CDT-U.

Telescope cross-checking. In Figure 15 we display the posterior information for the phase-integrated spectrum generated by the ST+PST hot regions; corresponding figures for the other models may

Table 2
Summary Table for ST+PST

Parameter	Description	Prior (Density and Support)	$\widehat{\mathrm{CI}}_{68\%}$	$\widehat{D}_{\mathrm{KL}}$
P (ms)	Coordinate spin period	$P = 4.87$,[a] fixed	⋯	⋯
M ($M_\odot$)	Gravitational mass[b]	$M \sim U(1, 3)$[c]	$1.34^{+0.15}_{-0.16}$	$1.26^{+0.02}_{-0.02}$
R_{eq} (km)	Coordinate equatorial radius[d]	$R_{\mathrm{eq}} \sim U[3r_g(1), 16]$[e]	$12.71^{+1.14}_{-1.19}$	$0.78^{+0.02}_{-0.02}$
	Compactness condition	$R_{\mathrm{eq}}/r_g(M) > 3$		
	Compactness condition[f]	$R_{\mathrm{polar}}(M, R_{\mathrm{eq}}, \Omega) \geqslant 3r_g(M)$		
	Enforce elliptical two-surface cross-section	function of $(M, R_{\mathrm{eq}}, \Omega)$		
Θ_p (radians)	ST region center colatitude[g]	$\Theta_p \sim U(0, \pi)$	$2.23^{+0.10}_{-0.10}$	$2.98^{+0.02}_{-0.02}$
ϕ_p (cycles)	ST region initial phase (from Earth)	$\phi_p \sim U(a, a + 0.2)$[h]	$0.46^{+0.00}_{-0.00}$	$4.44^{+0.03}_{-0.04}$
ζ_p (radians)	ST region angular radius	$\zeta_p \sim U(0, \pi/2)$	$0.09^{+0.01}_{-0.01}$	$4.62^{+0.03}_{-0.03}$
Θ_s (radians)	PST region superseding center colatitude	$\Theta_s \sim U(0, \pi)$	$2.91^{+0.05}_{-0.05}$	$3.85^{+0.03}_{-0.03}$
ϕ_s (cycles)	PST region initial phase (from Earth antipode)	$\phi_s \sim U(-0.5, 0.5)$, periodic[i]	$-0.59^{+0.04}_{-0.04}$	$2.43^{+0.02}_{-0.02}$
ψ_s^+ (radians)	PST region angular radii difference[j]	$\xi_s \sim U(0, \pi/2)$, $\psi_s^+ = \psi_s^+(\xi_s, f_s)$[k]	$-0.01^{+0.02}_{-0.02}$	$3.16^{+0.03}_{-0.03}$
ψ_s (radians)	PST region superseding angular radius	$f_s \sim U(0, 2)$, $\psi_s = \psi_s(\xi_s, f_s)$	$0.47^{+0.13}_{-0.13}$	$1.52^{+0.03}_{-0.02}$
$\varkappa_s$	PST region ceding fractional angular offset	$\varkappa_s \sim U(0, 1)$	$0.10^{+0.05}_{-0.03}$	$2.67^{+0.05}_{-0.05}$
φ_s (radians)	PST region ceding azimuthal offset	$\varphi_s \sim U(0, 2\pi)$	$0.71^{+0.10}_{-0.10}$	$2.28^{+0.03}_{-0.03}$
	Non-overlapping hot regions[l]	function of Θ_p through φ_s above		
$\log_{10}(\mathcal{T}_p\,(\mathrm{K}))$	ST region NSX effective temperature	$\log_{10}\mathcal{T}_p \sim U(5.1, 6.8)$, NSX limits	$6.11^{+0.01}_{-0.01}$	$5.29^{+0.03}_{-0.03}$
$\log_{10}(\mathcal{T}_s\,(\mathrm{K}))$	PST region NSX effective temperature	$\log_{10}\mathcal{T}_s \sim U(5.1, 6.8)$, NSX limits	$6.11^{+0.01}_{-0.01}$	$5.75^{+0.03}_{-0.02}$
i (radians)	Earth inclination to rotation axis	$i \sim U(0, \pi/2)$	$0.94^{+0.11}_{-0.10}$	$1.91^{+0.02}_{-0.02}$
D (kpc)	Earth distance	$D \sim N(0.325, 0.009)$[m]	$0.33^{+0.01}_{-0.01}$	$0.07^{+0.01}_{-0.01}$
N_{H} ($10^{20}\mathrm{cm}^{-2}$)	Interstellar neutral H column density	$N_{\mathrm{H}} \sim U(0, 5)$	$0.78^{+0.17}_{-0.16}$	$2.86^{+0.03}_{-0.03}$
NICER α	Calibrated matrix scaling	$\alpha \sim N(1, 0.1)$, $\alpha \in [0.5, 1.5]$	$0.99^{+0.08}_{-0.09}$	$0.04^{+0.01}_{-0.01}$
NICER β	Reference-to-calibrated matrix weighting	$\beta \sim U(0, 1)$	$0.19^{+0.17}_{-0.12}$	$0.96^{+0.02}_{-0.03}$
NICER γ	Reference matrix scaling	$\gamma \sim N(1, 0.1)$, $\gamma \in [0.5, 1.5]$	$0.98^{+0.08}_{-0.08}$	$0.08^{+0.01}_{-0.01}$
	Sampling process information[n]			
	Number of free parameters:[o] 19			
	Number of runs:[p] 1			
	Number of live points: 10^3			
	Inverse hypervolume expansion factor:[q] 0.8			
	Termination condition: 10^{-1}			
	Evidence:[r] $\widehat{\ln \mathcal{Z}} = -36368.28^{+0.49}_{-0.46}$			
	Global KL divergence: $\widehat{D}_{\mathrm{KL}} = 68.9^{+0.9}_{-0.8}$ bits			
	Number of core[s] hours: 42453			
	Likelihood evaluations: 78343018			
	Nested replacements: 57972			
	Weighted posterior samples:[t] 20177			

Notes. We provide (i) the parameters that constitute the sampling space, with symbols, units, and short descriptions; (ii) any notable derived or fixed parameters; (iii) the joint prior distribution, including hard truncation bounds and constraint equations that define the hyperboundary of the support; (iv) one-dimensional (marginal) 68.3% credible interval estimates symmetric in posterior mass about the *median* ($\widehat{\mathrm{CI}}_{68\%}$); and (v) KL divergence estimates in *bits* ($\widehat{D}_{\mathrm{KL}}$) representing prior-to-posterior information gain (a scalar with an associated error calculated as the 68.3% distributional interval about the median with respect to sampling process realizations; see Appendix A.2.4 for high-level description of the divergence). Constraint equations in terms of two or more parameters result in *marginal* distributions that are not equivalent to those inverse-sampled.

[a] Arzoumanian et al. (2018).
[b] Interpreted as a rotationally perturbed mass monopole moment (e.g., Hartle 1967), but the perturbation is small for the spin frequency of PSR J0030+0451 (see Section 4.2).
[c] Hard lower bound based loosely on plausible astrophysical formation channels (see, e.g., Strobel et al. 1999).
[d] An alternative two-dimensional space of $g(M, R_{\mathrm{eq}})$ and $h(M, R_{\mathrm{eq}})$—where $g \neq h$—on which we could choose to specify a joint flat prior density distribution is that of $g := M$ and $h := R_{\mathrm{eq}}/r_g(M)$.
[e] The function $r_g(M)$ denotes the gravitational radius explicitly in dimensions of length.
[f] The coordinate polar radius of the source two-surface, R_{polar} $(M, R_{\mathrm{eq}}, \Omega)$, is a quasi-universal function adopted from AlGendy & Morsink (2014), where $\Omega := 2\pi/P$ is the coordinate angular rotation frequency. This compactness condition, together with the elliptical surface requirement below are unimportant for the $P \sim 5$ ms spin period of PSR J0030+0451, and therefore we ignore these constraints in Tables 3 through 7.
[g] Note that for parameters where a lower bound of zero would correspond to absence of pulsations, we use some small finite number as a lower bound.
[h] Where $\phi_p = a$ is an arbitrary phase dependent on event data pre-processing. We set $a = 0.35$.
[i] The periodic boundary is admitted and handled by MultiNest. However, this is an unnecessary measure because we straightforwardly define the mapping from the native sampling space to the space of ϕ_s such that the likelihood function maxima are not in the vicinity of this boundary.
[j] The difference is defined as $\psi_s^+ := \zeta_s - \psi_s$, where ζ_s is the angular radius of the ceding member that radiates where it is not superseded.
[k] See Section 2.5.6.
[l] Refer to Appendix B.2.4.
[m] Constructed to approximate the information in the measurement (statistical information) reported by Arzoumanian et al. (2018). The support $D \in [0.235, 0.415]$ is equivalent to the interval $\mu \pm 10\sigma$.
[n] Refer to Appendix A for definitions.
[o] In the sampling space; the number of background count-rate parameters is equal to the number of channels defined by the data set.
[p] The mode-separation MultiNest variant was deactivated, meaning that isolated modes are not evolved independently and nested sampling threads contact multiple modes.
[q] For this sampling process, nor any such process reported in this work, we did *not* activate constant-efficiency MultiNest active-point bounding variant.
[r] Defined as the prior predictive probability $p(\boldsymbol{d} \mid \mathtt{ST} + \mathtt{PST})$. We report the interval about the median containing ±45% of 10^3 joint bootstrap-weight replications for the combined run. Note, however, that in order to complete the reported evidence for comparison to models other than those defined in this work, upper bounds for the background parameters need to be specified as described in Section 2.4.3.
[s] Intel® Xeon E5-2697Av4 (2.60 GHz; Broadwell) processors on the SURFsara Cartesius supercomputer. Note that these are physical cores—i.e., hyper-threading technology is not invoked.
[t] Excludes samples with important weight smaller than 10^{-6} times the largest such weight among samples.

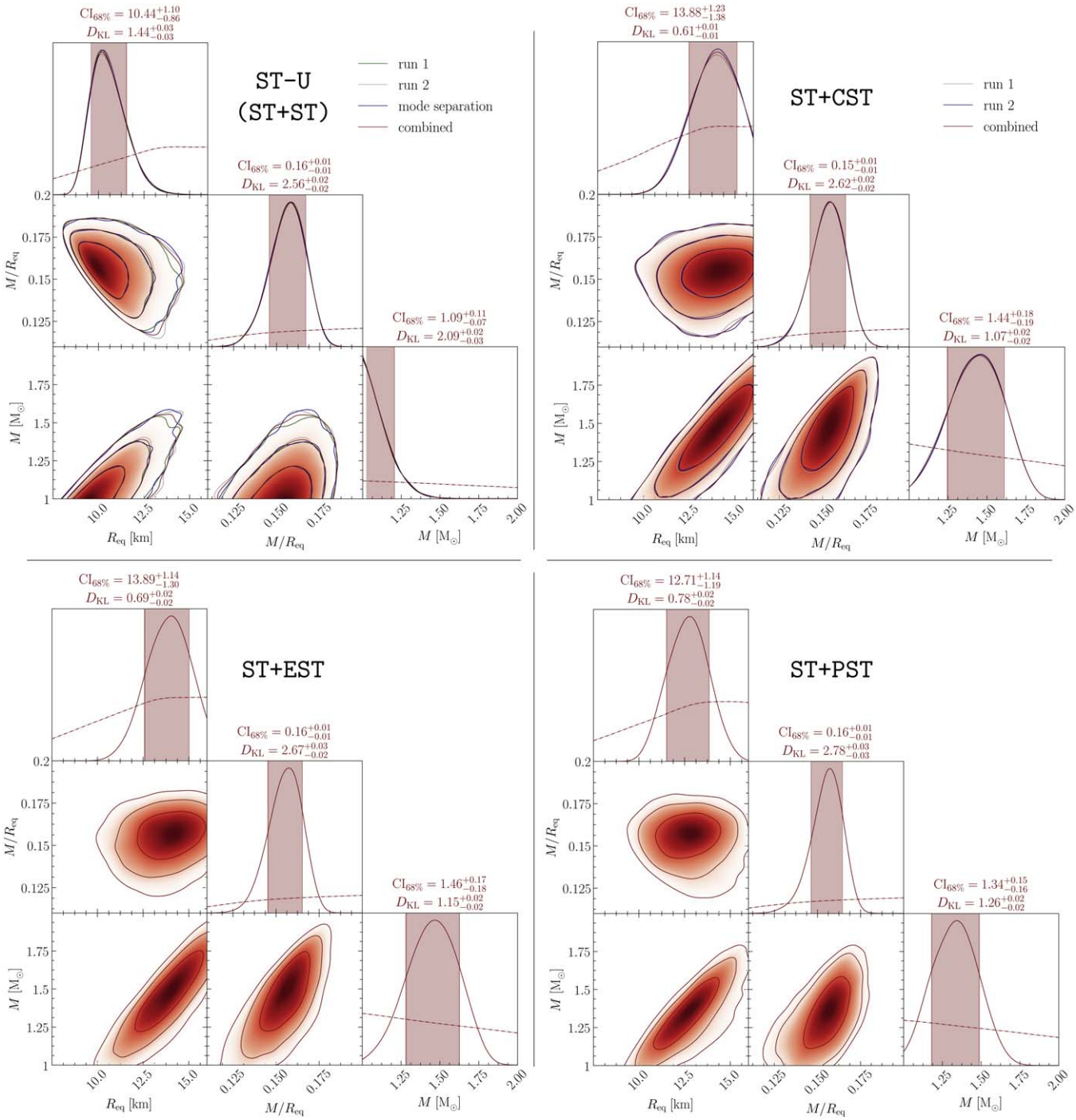

Figure 19. One- and two-dimensional marginal posterior density distributions for the MSP spacetime parameters conditional on the each in the sequence ST−U, ST+CST, ST+EST, and ST+PST. From leftmost to rightmost in each panel, the parameters are equatorial radius, (equatorial) compactness, and gravitational mass. We display the marginal prior density distributions for each parameter as the dashed–dotted functions. For the less expensive ST−U and ST+CST models we executed two production runs and combined them, while for each of ST+EST and ST+PST we executed a single production run. We also display an ST−U run in which the mode-separation MultiNest variant was activated, but because neither the theory nor the software exists for combining such a sampling process with runs in default mode (where sampling threads can migrate between posterior modes), it is not included in the combined run. The mode-separation run allocated ∼one-third of the sampling resolution (i.e., sampling threads, or active points) to a second posterior mode with negligible local mass; this mode corresponds to a distinct phase configuration, with the hot regions—which are distinguished by their order in colatitude—transposed in their coupling to the pulse components visible in the phase-folded event data. We report the KL divergence, D_{KL}, from prior to posterior in bits for each parameter, together with an error interval containing 68.3% of $\widehat{D}_{\mathrm{KL}}$ estimates based on simulated nested sampling process realizations. The shaded credible intervals $\mathrm{CI}_{68\%}$ for each parameter are symmetric in marginal posterior mass about the median, containing 68.3% of the mass; the (barely discernible) darker intervals at the $\mathrm{CI}_{68\%}$ boundaries contain 68.3%, respectively, of the 15.85% and 84.15% quantiles in posterior mass, again based on simulated nested sampling process realizations. The credible regions in the off-diagonal panels, on the other hand, are uniquely the highest-density—and thus the smallest possible—credible regions, containing 68.3%, 95.4%, and 99.7% of the posterior mass. In Appendix A we provide additional information regarding posterior kernel density estimation, error analysis, and the estimators displayed here. The complete figure set (7 images) is available in the online journal.

(The complete figure set (7 images) is available.)

be found in the online figure set. We overlay a model PSR J0030+0451 spectrum, derived via analysis of low-background low-signal-to-noise ratio phase-integrated *XMM* observations (see Bogdanov & Grindlay 2009); this spectrum is considered as a guiding upper bound for all emission in the *NICER* waveband from the PSR J0030+0451 system. The ST−U, CDT−U, ST+CST, ST+EST, and ST+PST models do not violate this condition in channels [25, 300) spanned by the event data considered, nor at incident photon energies that couple strongly to this channel subset; the ST−S model clearly violates this condition. The *XMM*-derived model is more uncertain than shown at low energies in Figure 15, due to unresolved likelihood function degeneracy between $N_{\rm H}$ and the thermal components. The *XMM* spectral analysis also included a power-law component, which is subsumed into the phase-invariant (background) terms in our analysis of *NICER* data (see the topmost *black* count-rate step function in the *bottom* panel of Figure 15); the discrepancy at higher energies is thus accounted for. Nevertheless, a pertinent question remains regarding how much of the signal generated by surface emission is captured by the phase-invariant count-rate terms; these terms combine linearly with the signals from the hot regions and are intended to parameterize the background contribution. We refer the reader to Section 4.1 for more detailed discussion on this topic.

Marginalization. For parameters that are shared between discrete models $\{\mathcal{M}_m\}$, where $\mathcal{M}_m \subset \mathscr{M}$, we could in principle marginalize over the discrete parameter m (see Appendix C of Riley & Watts 2019, for a formulation consistent with Section 2.3), provided we accept a marginal prior mass distribution of m. If two or more models are competitive, the marginal joint posterior distribution of the shared parameters is not dominated by the information from a single model. As stated above, we cannot distinguish between four of the highest-evidence models (ST+PST, ST+EST, ST+CST, and CDT−U). Moreover, for ST+CST, ST+EST, and ST+PST, the marginal joint posterior distribution of the shared spacetime parameters of interest is only mildly sensitive to model choice—i.e., the compactness constraints are commensurate, while the mass and radius are only weakly constrained individually, and the joint credible regions exhibit high partial overlap. For ST+PST, the marginal joint posterior distribution of the shared spacetime parameters discernibly evolves, but the mass and radius remain weakly constrained individually. Given that m labels models that are often (approximately) nested and which differ only in phenomenological complexity, we opt not to marginalize shared parameters over those models; instead, we report headline parameter estimates for ST+PST, which exhibits the largest background-marginalized likelihood function values in any posterior typical set.

4. Discussion

In this section we highlight how our inferences may be sensitive to the modeling assumptions made, and discuss the implications of our inferences for both dense matter physics and NS astrophysics.

4.1. Modeling Assumptions

The inferences that we report are conditional upon a number of modeling assumptions. These assumptions were physically motivated, and reassuringly there are no obvious large discrepancies or structures in the pulse-profile residuals to indicate a major problem. Nonetheless, the sensitivity of our results to these assumptions should be explored in future work, given additional computational resource allocations. With regard to the MSP, the biggest model-dependencies are (i) the atmosphere (see Section 4.1.1); (ii) the treatment of phase-invariant components of the total signal (see Section 4.1.2); and (iii) the assumption there exist two disjoint hot regions, each of which is radiatively contiguous and has a temperature field that is adequately represented by one of the models described in Section 2 (see 4.1.2). The instrument model (see Section 4.1.3) is of less concern.

4.1.1. Atmosphere

Two properties that could affect the atmosphere models used in our analyses are chemical composition (hydrogen as opposed to helium) and ionization state (fully versus partially ionized). In this Letter we have considered only a hydrogen composition: hydrogen would dominate the composition of matter accreted from the interstellar medium (Blaes et al. 1992), while matter accreted from a binary companion star would be predominantly hydrogen or helium. There are other processes that may drive changes in composition. For instance: hydrogen would result from spallation (Bildsten et al. 1992), diffusive nuclear burning could convert hydrogen to helium (Chang & Bildsten 2003, 2004), and significant pulsar wind excavation could make visible an underlying heavy element layer (Chang & Bildsten 2004). If the atmosphere were in fact dominated by helium, we could expect changes because helium atmospheres radiate differently from those of hydrogen. For example, hydrogen and helium model specific intensities, at an atmosphere effective temperature of 10^6 K (the approximate inferred temperature for the hot regions for this source), have fractional differences of at most 2%–5% at 0.5–1 keV in the (maximal) forward direction (Bogdanov et al. 2019c).

Regarding ionization state, our atmosphere models are constructed assuming the atmospheric plasma is fully ionized, such that the dominant opacity in regimes of interest is that due to electron free–free absorption (Ho & Lai 2001). While opacity tables for partially ionized matter exist (Iglesias & Rogers 1996; Badnell et al. 2005; Colgan et al. 2016), they do not cover the full range of energies and temperatures needed for our analysis. However, the hydrogen neutral fraction in the atmosphere at 10^6 K is low, and a comparison of our fully ionized hydrogen atmosphere model with that constructed using the OP (Badnell et al. 2005) opacity table yields specific intensity fractional differences of at most 1%–2% at 0.5–1 keV in the (maximal) forward direction at an atmosphere effective temperature of 10^6 K (Bogdanov et al. 2019c). The importance of including partial ionization, in comparison to developing other aspects of the model (e.g., Section 4.1.2), is not clear; but partially ionized models will be part of future re-analysis as updated opacity tables become available.

4.1.2. Surface Heating Estimation and Treatment of Phase-invariant Components

Predicting the MSP surface temperature field from ab initio[68] calculations of energy deposition by magnetospheric currents is, as described in Section 2, notoriously challenging. In this work we have assumed that there are only two distinct hot regions, motivated by the fact that there appear to be two pulsed components in the pulse profile. If the (surface) field

[68] Terminology adopted from the series of studies by Philippov et al. (2015a, 2015b) and Philippov & Spitkovsky (2018) on pulsar magnetospheric simulations.

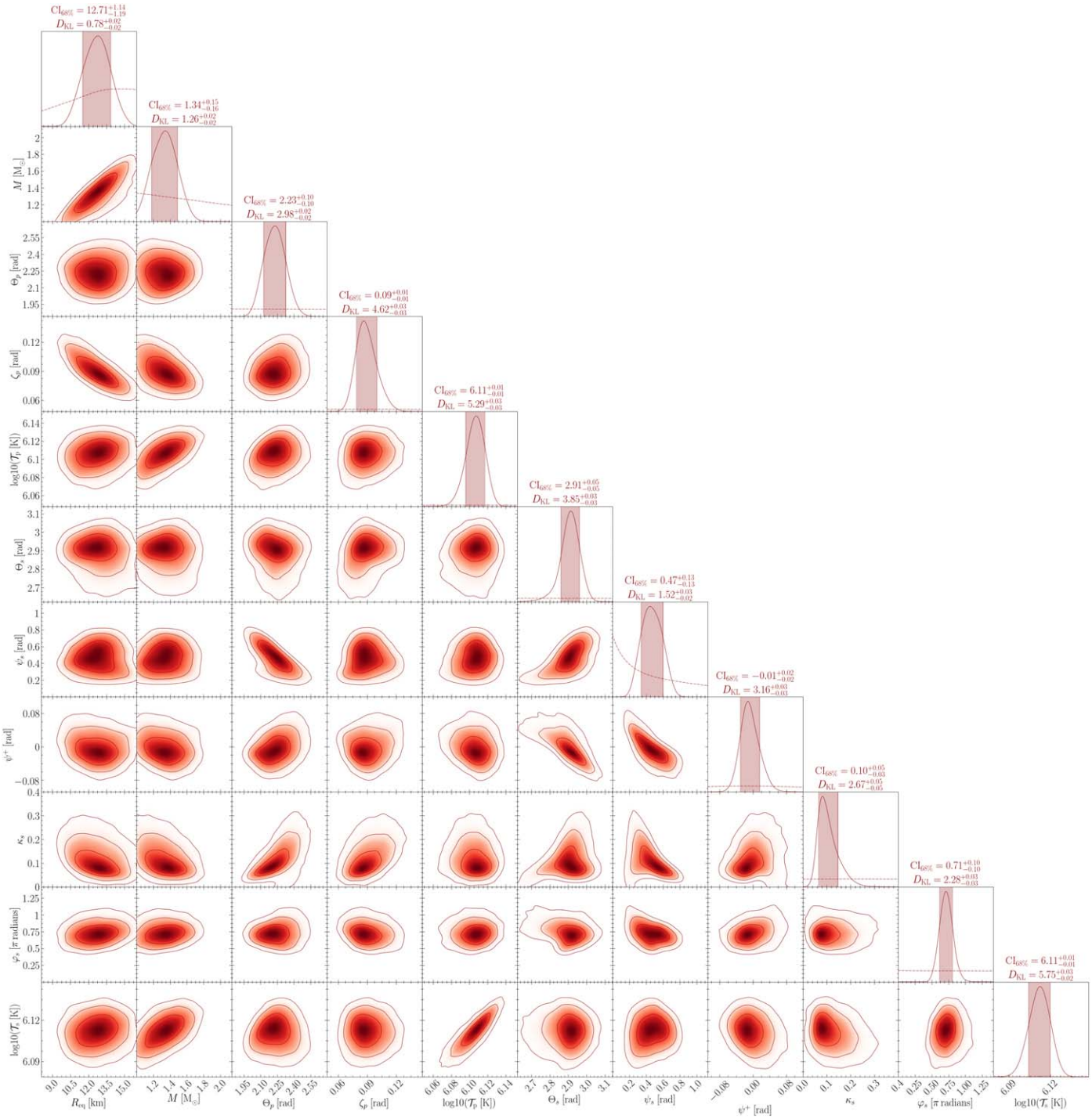

Figure 20. One- and two-dimensional marginal posterior density distributions for the MSP parameters conditional on ST+PST. From leftmost to rightmost: equatorial radius R_{eq}; gravitational radius M; ST center colatitude Θ_p; ST angular radius ζ_p; ST NSX effective temperature $\log_{10}\mathcal{T}_p$; PST superseding member center colatitude Θ_s; PST superseding angular radius ψ_s; PST superseding member angular radius difference $\psi_s^+ = \zeta_s - \psi_s$; where ζ_s is the angular radius of the ceding member; PST ceding member fractional angular offset $\varkappa_s$ (labeled as κ_s); PST ceding member azimuthal offset φ_s; and PST ceding NSX effective temperature $\log_{10}\mathcal{T}_s$. For descriptions of the information displayed, refer to Figure 19; note that here we display the marginal posterior density distribution for each parameter as a single solid function due to the number of panels. We choose not to display joint posterior distributions for all pairs of model parameters because the number of panels is prohibitive; moreover, the posterior azimuthal separation of the ST and PST regions is displayed in Figure 21. The complete figure set (6 images) is available in the online journal.

(The complete figure set (6 images) is available.)

structure involves higher-order multipoles,[69] additional polar caps—and thus additional disjoint surface heating—may be possible, and/or more complicated polar cap topologies (such as ring-like) may be possible. We have also assumed specific forms for each hot region; the true temperature field is more complex, and both the physical complexity and our statistical sensitivity to such complexity should be investigated further. In lieu of a physical emission model for the stellar surface exterior

[69] With respect to some coordinate system that simplifies vector spherical harmonic field expansion; this coordinate system will generally be rotated and displaced from the system with stellar spin axis defined as the polar axis.

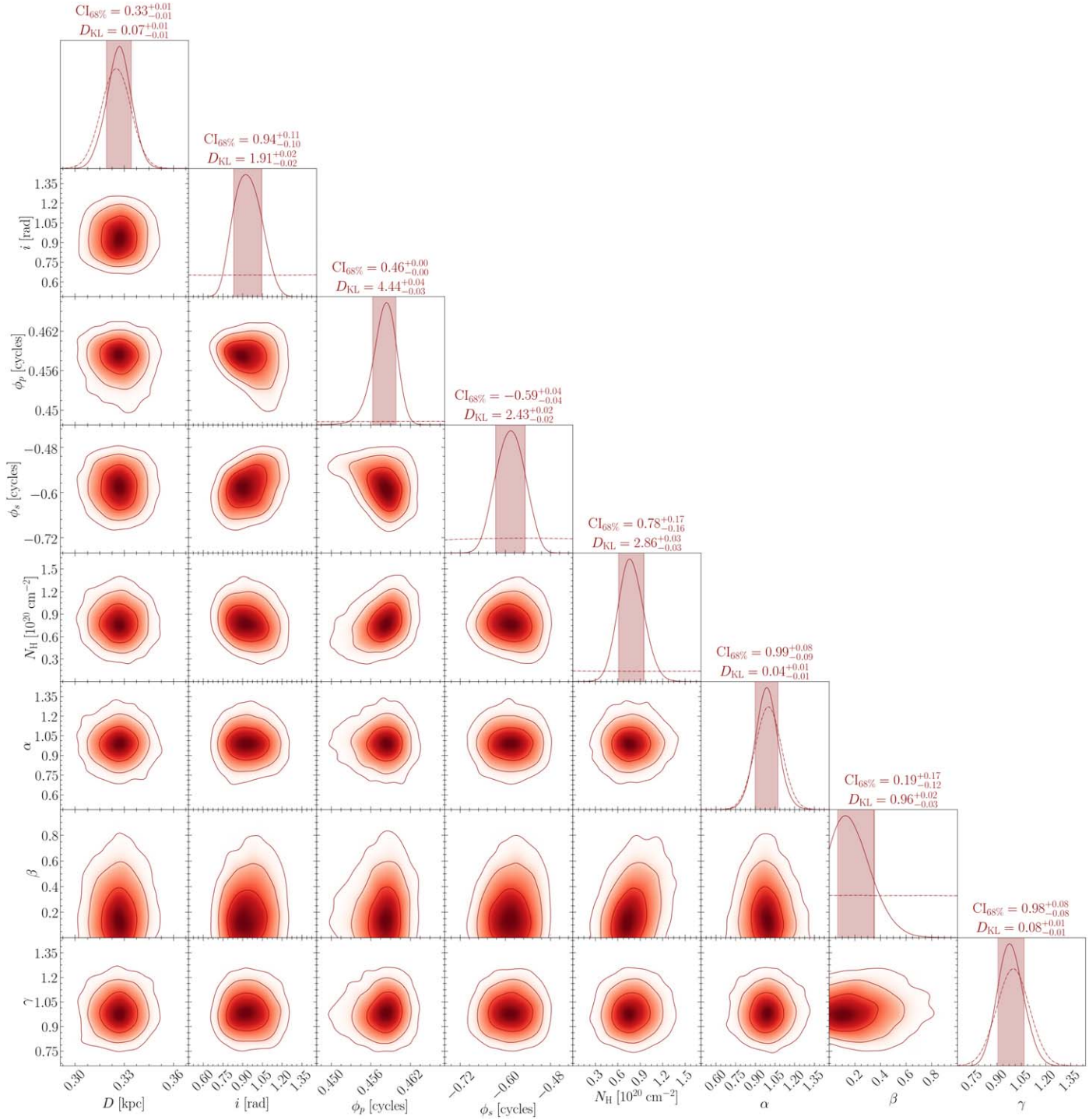

Figure 21. One- and two-dimensional marginal posterior density distributions for parameters pertaining mostly to the MSP observation, conditional on ST+PST. From leftmost to rightmost: distance D; Earth inclination i; ST center azimuth (phase) relative to Earth direction ϕ_p; PST superseding member center azimuth (phase) relative to Earth antipode ϕ_s; interstellar neutral hydrogen column density N_{H}; and *NICER* instrument parameters α, β, and γ. For descriptions of the information displayed, refer to Figure 19; note that here we display the marginal posterior density distribution for each parameter as a single solid function due to the number of panels. The complete figure set (6 images) is available in the online journal.

(The complete figure set (6 images) is available.)

of the hot regions, we subsumed the non-pulsed component of any such emission (which is expected a priori to be dominant if hot regions with smaller angular extent are favored) within phase-invariant count-rate terms; however, this would be a minor concern on the premise that outside of the footpoints of open magnetic field lines at the polar caps, there is no energetically comparable heating to which we are sensitive when observing with *NICER*.

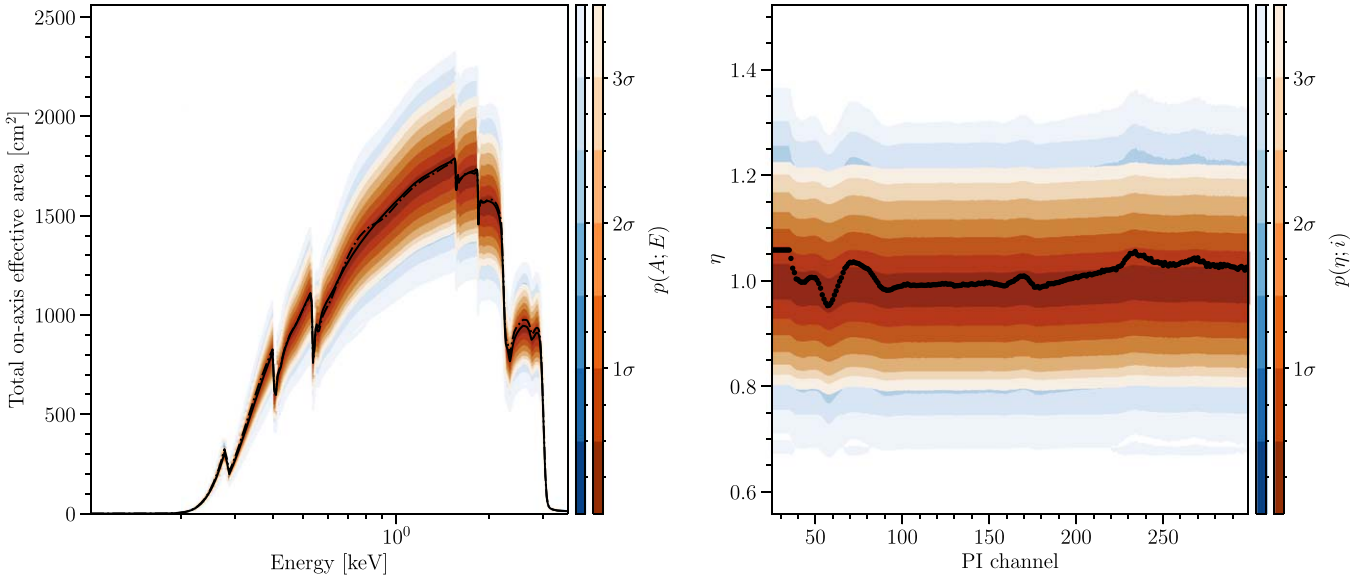

Figure 22. We update the instrument prior (displayed in Figure 2) conditional on ST+PST. The conditional posterior distributions are represented by the orange bands; the conditional prior distributions are displayed in blue to indicate information gain about the instrument, and the black lines and points are equivalent to those in Figure 2. The complete figure set (6 images) is available in the online journal.

(The complete figure set (6 images) is available.)

All emission from sources other than the hot regions—a combination of astrophysical and instrumental—was left free in our models in the form of a set of phase-invariant count-rate terms (background parameters), one per channel, which we collected under the envelope of *background contribution* (refer to Section 2.4.3). Moreover, an improper joint flat prior was implemented that was separable with respect to these background parameters; no upper bounds (or lower bounds) were defined for the prior support, in lieu of a physical (generative) model for the total contribution from the hot regions (and thus surface if emission exterior of the regions is considered unimportant). Such a model would need to account for the combined (phase-invariant) signal attributed to off-surface emission and any astrophysical backgrounds in the field of view, in the *NICER* waveband, based on previous observations of PSR J0030+0451 and/or theoretical modeling.

Remarkably, by neglecting any physical (generative) model for the total counts attributed to *all* surface emission, the phase-invariant terms can even capture emission from the hot regions. One can reason that by permitting the phase-invariant terms to capture all or most of the phase-invariant signal components, there may then exist *background-marginalized* likelihood function maxima corresponding to signals that are (i) dominated by phase-invariant terms over a fraction of a rotational cycle, and (ii) elsewhere found to describe the pulse-profile adequately (in combination with the phase-invariant terms). Near pulse minima, the hot region contribution can be entirely dominated in linear combination with the phase-invariant terms, if the regions are both close to the visible limb of the star, or even partially or wholly non-visible. Such a heating configuration may exhibit systematic bias in the sense that (i) it is not considered an adequate approximation of the configuration inferred when physical limits are modeled for the contribution from the stellar surface (or specifically the hot regions), or (ii) that a configuration alluded to in point (i) is not encompassed by a given posterior credible region boundary. On the other hand, it could be viewed that our treatment of the phase-invariant component of the total signal is in some aspects conservative: posterior credible regions may be appreciably larger than if such physical limits on surface contribution are imposed. In the *X-PSI* documentation (Riley & Watts 2019) a simple parameter estimation workflow is demonstrated using the same default background treatment implemented in this present work; while there is no evidence for systematic bias nor credible region inflation in that specific case (where the true data-generating process is known), guarantees cannot be made universally.

Upon examination of Figure 14 we see that the combined signal from the hot regions falls to near zero at its minimum—such that the fractional amplitude of the signal is near unity—and is thus *not inconsistent* with a phase-invariant component of the combined signal being subsumed in the background. Further, examination of Figure 15, in which we cross-check the ST+PST region spectrum with a PSR J0030+0451 spectrum inferred by Bogdanov & Grindlay (2009), further suggests that a fraction of the contribution from the hot regions is subsumed in the background. Interestingly, Bogdanov & Grindlay (2009) inferred a spectrum with two thermal components of different temperatures.[70]

We attribute the discrepancy to a number of factors. First, the *XMM* photon event set had size $\mathcal{O}(10^4)$, nearly two orders of magnitude smaller than the number of *NICER* events used in this work. Second, it is known that Bogdanov & Grindlay (2009) did not fully resolve degeneracy between the thermal components and the neutral hydrogen column density N_{H}. Lastly, *XMM* is an imaging telescope: the background signal

[70] Note that Bogdanov & Grindlay (2009) explicitly calculate the phase-resolved signal generated by rotating single-temperature circular hot spots, and then phase-average; it is thus not the case that the discrepancy is explained by *our* rotating circular hot spots (refer to ST-U), whose effective temperatures are commensurate, mimicking a dual-temperature *incident* spectrum due to relativistic rotation.

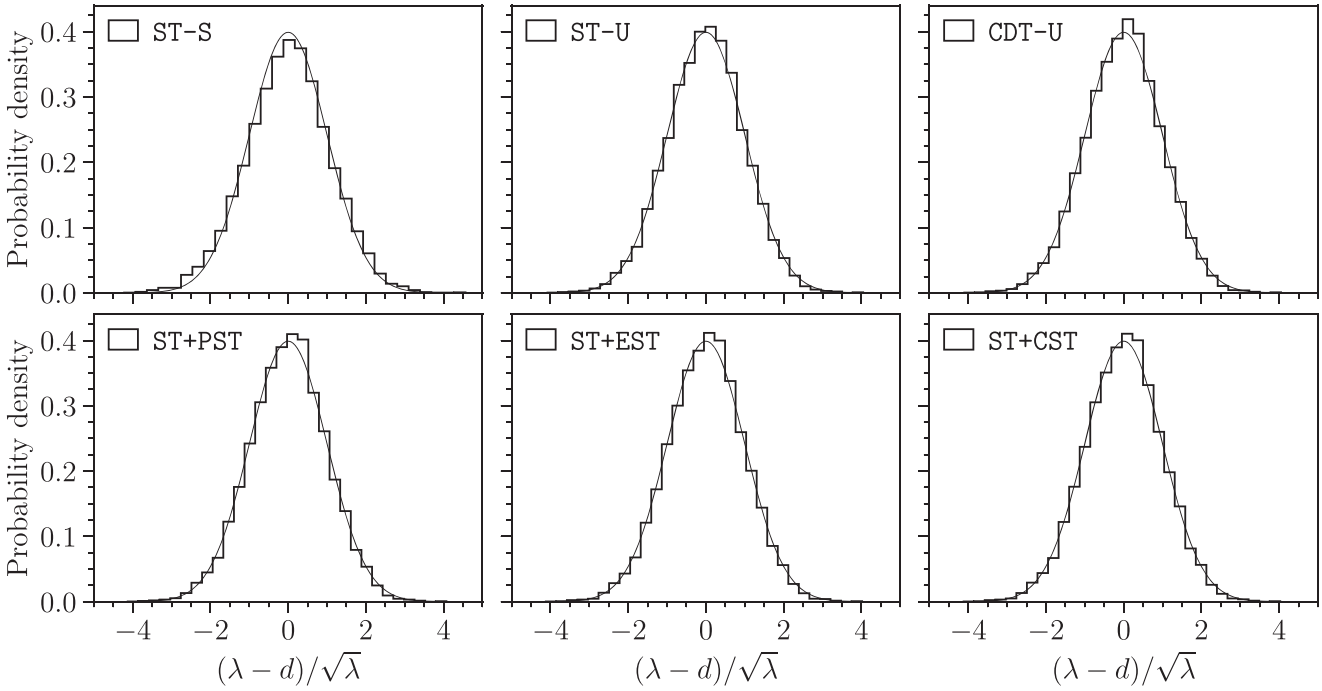

Figure 23. Summary of the two-dimensional residual panels in the online figure set associated with Figure 13. For each model we identify the sample (parameter vector) $\hat{\boldsymbol{\theta}}_{\mathrm{BMML}}$ that reported the highest background-marginalized likelihood function value among identified posterior modes, together with the background count-rate vector $\hat{\boldsymbol{B}}_{\mathrm{CML}}$ that maximizes the conditional likelihood function given $\hat{\boldsymbol{\theta}}_{\mathrm{BMML}}$. Given the expectations $\mathbb{E}[x_{ij}] = \lambda_{ij}(\hat{\boldsymbol{\theta}}_{\mathrm{BMML}}, \hat{\boldsymbol{B}}_{\mathrm{CML}})$, we evaluate the residuals $(\lambda - d)/\sqrt{\lambda}$ for each phase-channel interval, where $\{d_{ij}\}$ are variates of independent Poisson-random variables $\{x_{ij}\}$. We display in each panel the distribution of these random variates (the real count numbers). The minimum number of expected counts in any phase-channel interval is $\sim$30, and thus the $\{x_{ij}\}$ drawn from the sampling distribution are, approximately, identically and normally distributed if transformed as $y \coloneqq (\lambda - x)/\sqrt{\lambda}$. The smooth black distribution in each panel is a normal distribution. The data sampling-distributions for the ST-U model and models higher in complexity do not exhibit systematic structural inaccuracies in the context of the variates: there are no clear signs of noise-model inaccuracy or over-fitting, nor are there clear signs of residual correlations in, e.g., Figure 13. These models can thus perform adequately as point-measure predictors of structure in the real count-number data. The residuals conditional on the ST-S model are distributed with larger variance than unity (normal): more weight is visible in the wings, and less in the near vicinity of zero.

(The complete figure set (3 images) is available.)

was well determined by imaging nearby source-free regions of the sky, and was subsequently used to impose that the surface hot spots generate the remaining signal from the imaged (point-source) MSP. Such a model is distinctly different from those we consider here for *NICER*, a non-imaging telescope; we do not impose (e.g., via some informative background prior) the signal to be generated by the surface hot regions. As a result, we may inaccurately subsume a cooler contribution from the hot regions into the phase-invariant likelihood terms as suggested above; Bogdanov & Grindlay (2009), on the other hand, may miss some non-diffuse radiative component(s) in the near vicinity of the surface of PSR J0030+0451, and thus require their hot spots to generate additional cooler emission than they do in physical reality.

With these considerations in hand, we conclude that without further work, it is unclear which inferred signal is more physically accurate. Indeed, the existence and treatment of additional X-ray emission by PSR J0030+0451—or from its circumstellar vicinity—in the *NICER* waveband is considered an open question for future modeling. Moreover, the sensitivity of parameter estimation to requiring a certain phase-invariant contribution specifically from the hot regions could be investigated with newly allocated computing resources.

A (superficially) straightforward alternative to a full physical generative model of the non-pulsed surface and off-surface emission would be to define bounds—upper in particular—on the prior support of the channel-by-channel background count-rate parameters (refer to related discussion on limits in Section 2.4.3). One must then address the question of how. One option is to move the difficulty of defining a prior density for the background to another level in a Bayesian hierarchy via a hyperparameter (e.g., the upper bound of a flat density function) and a corresponding hyperprior for each channel; each hyperparameter and background parameter pair must then be jointly numerically marginalized over to ensure sampling is tractable in a lower-dimensional parameter space. Another option would be to define a conditional prior distribution for each background parameter, where the upper limit of the support is a function of the source parameters—e.g., some fraction of the phase-average source counts in a channel.

We did not consider at the outset of this work the possibility of jointly modeling *NICER* and *XMM* event data—a strategy both tractable and arguably more rigorous given that we do not have likelihood function information nor posterior information from Bogdanov & Grindlay (2009) that is compatible as a prior information to be updated conditional on the *NICER* data. We could thus consider constructing a joint likelihood function over observations with both telescopes (requiring definition of at least one more nuisance parameter). On the other hand, the

NICER event data comprise far more photons than do the *XMM* data, and so inferences may be dominated by the (background-marginalized) *NICER* likelihood function. There are, however, good prospects for improving our understanding of the *NICER* background (particle radiation and diffuse sky terms; see Section 2.4.3 and Bogdanov et al. 2019a), even without new input from *XMM* observations or other imaging capabilities.

Observations with a future high-time resolution soft X-ray spectroscopic imaging telescope, such as the *Athena X-ray Observatory* (Nandra et al. 2013), would be synergistic with archival *NICER* data—and/or observations with a future large-area soft spectroscopic timing telescope such as the *enhanced X-ray Timing and Polarimetry mission* (*eXTP*; Zhang et al. 2019; Watts et al. 2019) or the *Spectroscopic Time-Resolving Observatory for Broadband Energy X-rays* (*STROBE-X*; Ray et al. 2019). The high-sensitivity (read large-area) imaging capabilities of *Athena* (with $\sim 5''$–$10''$ angular resolution[71]) and its two instruments, the *X-ray Integral Field Unit* (*X-IFU*; 10 μs time resolution[72]) and the *Wide Field Imager* (80 μs time resolution), will permit modeling of the background emission not originating from the line of sight of the PSR J0030+0451[73] system (as permitted by the point-spread function) in a similar manner to Bogdanov & Grindlay (2009) for *XMM* (and also Bogdanov 2013, for PSR J0437−4715). Furthermore, the microcalorimeter spectral resolution of *X-IFU* ($\sim$2.5 eV resolution for photons energies <7 keV) will permit precise measurements of various edges of the interstellar medium (e.g., oxygen K edge at 0.55 keV, iron L edge at 0.72 keV), thereby enabling derivation of independent tight constraints on the hydrogen column density $N_{\rm H}$ for a given MSP. Archival *NICER* MSP data, on the other hand, will remain valuable far into the future because the principal *NICER* mission science objectives focus purely on MSPs; *NICER* will have compiled $\mathcal{O}(10^6)$ s integrated exposures on the primary MSP targets, meaning that despite the smaller total effective area, the size of the event data set will be synergistic with an advanced multi-faceted mission such as *Athena* whose science objectives are broader. Consequently, observations with the *Athena* instruments will enable us to jointly model data sets, and make definitive progress on disentangling the signal generated by surface hot regions from the complicated phase-invariant emission detected by *NICER* (or a future non-imaging telescope).

4.1.3. Instrument

We implemented a specific instrument response model (Section 2.4.2): a simple ad hoc parameterization designed to combine several available calibration products and thus account for both energy-independent (absolute flux calibration) and energy-dependent uncertainty. Sensitivity to our choice of astrophysical calibration source, the Crab, also needs further study; it is known that operation in response to an incident radiation field is a function of its properties (see Section 2.4.2). Future study could explore whether using different calibration sources affects inferences. Another interesting question may be posed as to how the posterior information about this instrument model evolves as we update our knowledge via analysis of other sources (MSPs and otherwise).

As our understanding of the *NICER* instrument improves, it should be possible to improve on this model. One option might be to adopt a more sophisticated approach based on Lee et al. (2011) and Xu et al. (2014) to estimating the uncertainty in instrumental response for *Chandra*. At present such a sophisticated approach may be unjustified because computational expense would be amplified by increasing the complexity of the instrument model instead of—or in addition to—the complexity of model astrophysical sources whose nature is more uncertain.

4.2. Mass and Radius Constraints in Context

For ST+PST, the inferred mass M and equatorial radius $R_{\rm eq}$ are $1.34^{+0.15}_{-0.16}\,M_\odot$ and $12.71^{+1.14}_{-1.19}$ km, where the credible interval bounds are approximately the 16% and 84% quantiles in marginal posterior mass, given relative to the median. The marginal credible intervals may thus be considered as 1σ intervals (containing 68.3% of the posterior mass), where $\sigma_M/M \sim 11\%$ and $\sigma_R/R \sim 9.2\%$. For completeness, and to assist comparison with results derived using other methods, we also give the values for the 90% credible interval ($M = 1.34 \pm 0.24\ M_\odot$, $R_{\rm eq} = 12.71^{+1.83}_{-1.85}$ km) and the 95% credible interval ($M = 1.34^{+0.28}_{-0.27}\,M_\odot$, $R_{\rm eq} = 12.71^{+2.15}_{-2.14}$ km). The (equatorial) compactness $GM/R_{\rm eq}c^2 = 0.156^{+0.008}_{-0.010}$ is more tightly constrained than both M and $R_{\rm eq}$ individually, at the $\sim$6% level; the 90% credible interval is $GM/R_{\rm eq}c^2 = 0.156^{+0.013}_{-0.017}$, and the 95% credible interval is $GM/R_{\rm eq}c^2 = 0.156^{+0.015}_{-0.021}$.

The effect of a rotation rate of 200 Hz on NSs is a small deformation of the star into an oblate spheroid. The deformation enters in two ways. A star rotating at 200 Hz will have an equatorial radius that is at most 2% larger than a nonrotating star with the same mass for the stiffest equations of state, with smaller increases in radius for soft equations of state (Cook et al. 1994).[74] The polar radius will be smaller than the equatorial radius, an effect that depends on the equatorial radius, mass, and spin, with very little dependence on the EOS (AlGendy & Morsink 2014, and references therein); for the a posteriori most probable exterior spacetimes inferred in this present work, the polar radius is $\sim$1% smaller than the equatorial radius. Rotation also increases the mass of a star compared to a nonrotating star with the same number of baryons, but this is at most a 0.2% effect for the rotation-powered pulsars observed by *NICER*. The effect of rotation on the location of the innermost stable circular orbit (ISCO) is larger (van Doesburgh et al. 2018): the (prograde) ISCO typically ranges from $5.6r_g$–$5.8r_g$ for stars spinning at 200 Hz for most theoretical equations of state, compared to the Schwarzschild ISCO radius of $6r_g$. The 95% compactness

[71] Refer to Bogdanov et al. (2019a) for an *XMM* image of the PSR J0030+0451 field; *Athena* promises to improve on the *XMM* point-spread function half-energy width by a factor of $\sim$1.5–3.

[72] Similar to that expected for the *eXTP Spectroscopic Focusing Array* ($\sim$10 μs), but not as good as spectro-timing dedicated missions, like *NICER* or *STROBE-X* with $\sim$0.1 μs time resolution. Note that event time-tagging resolution offered by *X-IFU* will still achieve a level of $\sim$1/500 of the spin period of PSR J0030+0451.

[73] And other MSPs targeted by a mission such as *NICER*.

[74] Note that a one-parameter sequence of nonrotating stars deforms into a sequence of stars rotating at some rate, but a unique deterministic map between the central densities that parameterize those sequences does not exist. This reflects the absence of a unique physical mode for stars to evolve in rotation rate. For comparative purposes, one often considers sequences—parameterized by rotation rate—that conserve a quantity such as total mass, total baryon number, or central density. In the accompanying Letter of Raaijmakers et al. (2019), it is the central density that is explicitly defined as a model parameter (to be marginalized out), and which is held constant for comparison of nonrotating stars to rotating stars in the context of likelihood function evaluation.

credible interval corresponds to equatorial radii in the range of $5.84r_g$–$7.41r_g$, meaning that the most probable stellar surfaces[75] have radii close to or larger than the radius of the ISCO, and thus that the innermost permitted stable orbit is typically at (or just exterior to) the surface.

In this subsection we proceed to discuss how the constraints derived in this work compare to existing constraints on mass, radius, and compactness derived using other methods and independent observations. We also consider prospects for improving *NICER* constraints on PSR J0030+0451. When comparing radii, it should be remembered that most of the published radius determinations using other methods have assumed that the star is spherical. For luminosity radius determinations of rapidly rotating NSs (spins of a few hundred Hz), systematic errors of 5% could be introduced (Bauböck et al. 2015), although other systematic errors may well dominate over an inaccurate treatment of rotational surface and metric deformation.

The present constraints on the radius are consistent with the previous radius lower limit for this pulsar, $R_{\rm eq} > 10.7$ km (95% confidence, assuming a 1.4 $M_\odot$ NS), obtained from early pulse-profile modeling of *XMM* data (Bogdanov & Grindlay 2009). It is also in agreement with the radii inferred from *XMM* observations of other MSPs: $R_{\rm eq} > 11$ km (3σ confidence) for PSR J0437−4715 (Bogdanov 2013); and $R_{\rm eq} > 7.8$ km (68% confidence) for PSR J2124−3358 (Bogdanov et al. 2008). However, these early light-curve models only considered uniform-temperature circular hot spots, neglected stellar oblateness, and we consider the statistical computation described in this Letter as more advanced.

Early constraints on the NS radius were obtained from the X-ray spectroscopic modeling of the thermal emission originating from the entire surface of isolated NS RX J1856.5−3754 (Burwitz et al. 2001; Drake et al. 2002; Pons et al. 2002). However, uncertainties due to the distance and due to the calculations of radiative transfer in the magnetized atmosphere hindered an accurate radius constraint for this NS (Ho et al. 2007). More robust constraints, on the other hand, can be extracted from the X-ray spectra of quiescent low-mass X-ray binaries (qLMXBs) hosted in globular clusters. Not only are their distances known to better than ∼10% precision, but their purely thermal emission is thought to emerge from non-magnetic NS atmosphere models (similar to the NSX model used in the present work) in order to estimate the apparent radius $R_\infty = R_{\rm eq}(1+z) = R_{\rm eq}(1-2GM/R_{\rm eq}c^2)^{-1/2}$. However, the degeneracy between $R_{\rm eq}$ and M in the estimation of R_∞ precluded obtaining useful constraints on the EOS (e.g., Webb & Barret 2007; Heinke et al. 2006, 2014; Guillot et al. 2011), due to the typical shapes of the M–$R_{\rm eq}$ confidence contours that made them compatible with many families of EOS. These results prompted the simultaneous analysis of a set of sources.

In these combined analyses, the degeneracy between M and $R_{\rm eq}$ was lifted by assuming a parameterized shape for the EOS, either a toy-model (constant-radius EOS, as a simplistic representation of nucleonic EOS) or an analytical representation using polytropes. A handful of qLMXBs in globular clusters (up to seven) and of Type-I X-ray bursters (four or five) have been combined to produce constraints on the EOS and/or the radii of NSs (given the assumed EOS shape). Early works using only qLMXBs produced rather small NS radii, $R_{\rm eq} \approx 9$–10 km with ∼10%–15% uncertainties (90% credible interval; Guillot et al. 2013; Guillot & Rutledge 2014; Guillot 2016), but the addition of new data and the use of more recent globular cluster distance measurements resulted in higher values, $R_{\rm eq}$ in the 9.9–11.2 km range for a 1.5 $M_\odot$ NS (2σ credible interval; Bogdanov et al. 2016). Combining qLMXBs and Type-I X-ray bursts, other works have found radii in a wide range of values: $R_{\rm eq} \approx 10.4$–12.9 km (95% credible interval; Steiner et al. 2010, 2013), $R_{\rm eq} \approx 10.5$–12.7 km (90% credible interval; Lattimer & Steiner 2014), and $R_{\rm eq} \approx 9.8$–11.4 km (95% credible interval, Özel et al. 2016), for a 1.4 $M_\odot$ NS. However, these analyses may have been affected by systematics, such as those due to the modeling of piled-up X-ray photons in the *Chandra* data[76]—which was not considered for all qLMXBs in these early analyses—or those due to the choice of atmospheric composition for qLMXBs (generally H versus He). The mass–radius constraints extracted from Type-I X-ray bursts are dependent on the color-correction factors used (between the measured blackbody temperatures and the modeled effective temperatures of the burning atmospheres), which have been debated in the literature (e.g., Suleimanov et al. 2011a; Güver et al. 2012; Güver & Özel 2013; Kajava et al. 2014). These issues cast doubt on the robustness of the error intervals reported in these early works.

Nättilä et al. (2017) recently analyzed Type-I X-ray bursts from 4U 1702−429 by fitting bursting atmosphere models directly to spectra during the cooling tail of the bursts, hence avoiding the use of color-correction factors. These authors found $R_{\rm eq} = 12.4 \pm 0.4$ km and $M = 1.9 \pm 0.3\,M_\odot$ (68% credible interval), although the posterior distributions also allow smaller radii ∼10 km for higher masses ∼2.1 $M_\odot$.

Steiner et al. (2018) considered the effects of pile-up on qLMXB spectra, the possibility of pure helium atmospheres (instead of pure hydrogen atmospheres), as well as non-uniform surface temperature distributions. They obtained constraints on polytropic EOS via Bayesian inference of mass–radius probability distributions of seven NS qLMXBs. These constraints translate to a NS radius in the 10.0–14.4 km range (95% credible interval, assuming a 1.4 $M_\odot$ NS), when considering all of the models tested.

More recently, a physically justified parameterization of the EOS was proposed as an alternative to polytropes (Margueron et al. 2018a, 2018b). In that work, the EOS is a meta-model expressed as a Taylor expansion of nuclear physics parameters, and was applied to a combined spectral analysis of seven qLMXBs to directly extract values of nuclear physics parameters $L_{\rm sym}$, $K_{\rm sym}$, and $Q_{\rm sat}$ (Baillot d'Etivaux et al. 2019). Radius estimates were also derived: $R_{\rm eq} = 12.35 \pm 0.37$ km (2σ credible interval) assuming a 1.45 $M_\odot$ NS.

The cold emission from the MSP PSR J0437−4715 is detectable in the far ultraviolet (Durant et al. 2012) and in the soft X-ray band (0.1–0.3 keV; Guillot et al. 2016), and its mass and distance are known precisely from radio timing (Reardon et al. 2016). Applying NS atmosphere models[77] has permitted estimation of this pulsar's radius: $R_{\rm eq} = 13.1 \pm 0.8$ km (68%

[75] Which in this present work are only embedded in the ambient spacetime while neglecting rotational metric deformation (see Section 2.4.1) and are thus not self-consistently computed with global numerical solutions to the field equations given interior conditions (including an EOS).

[76] It was shown that an unmodeled pile-up fraction as low as ∼1% could affect the radius estimated via modeling of spectral data by as much as 10% (Bogdanov et al. 2016).

[77] Similar but not identical to the atmosphere models used in our *NICER* analysis—the models used in Gonzalez-Caniulef et al. (2019) incorporated various effects that are important at lower temperatures such as partial ionization and plasma frequency effects.

credible interval; Gonzalez-Caniulef et al. 2019). We note that this was a phase-averaged spectral analysis, wherein the emission is assumed to originate from the $\sim 10^5$ K stellar surface exterior to the (heated) polar cap regions.

An indirect method to constrain NS radii is to use emission features from an inner accretion disk in accreting LMXBs. Narrow emission lines, such as Fe K, arising from this rotating material are asymmetrically broadened—to which there is a strong relativistic contribution based upon proximity of the inner disk to the NS surface (Fabian et al. 2000). The accretion disk must truncate at the stellar surface or at a larger radius: spectral modeling of these emission lines enables derivation of a statistical constraint on the inner radius of the (prograde) disk *in units of* the gravitational radius or the spin-dependent ISCO radius (Cackett et al. 2008, 2010; Miller et al. 2013; Degenaar et al. 2015; Ludlam et al. 2017). To do so, an approximative ambient spacetime solution (e.g., Schwarzschild or Kerr) is typically invoked—as in this present work—but without the embedding of a NS surface.[78] The constrained inner radius can then be translated into a lower limit on the stellar compactness for that particular NS.

Ludlam et al. (2017) inferred, for two LMXB systems, inner radii that are consistent with the disks extending down to the ISCOs of their respective ambient spacetimes. In the absence of a constraint on the NS mass that is independent[79] of the inner radius of the disk, a mass merely has to be assumed to obtain an upper limit on the stellar radius. For example, suppose that the 4U 1636−53 system contains a nonrotating[80] minimally compact NS (i.e., disk truncation by surface): for a mass of 1.4 $M_\odot$, the upper limit on the stellar radius would lie close to the Schwarzschild ISCO, at 12.4–13.1 km (1σ confidence interval; Ludlam et al. 2017), which is not inconsistent with the constraint derived conditional on the *NICER* data.[81] Lastly, if the surface of an NS in the 4U 1636−53 system does *not* approximately truncate the disk—and is thus more compact than the ISCO—and/or the (prograde) ISCO is more compact due to NS rotation, the NS could only be viewed as inconsistent with our compactness estimate for PSR J0030+0451 if the following are true: (i) the NS masses are both tightly constrained and happen to be highly commensurate, and (ii) the NSs do not occupy a segment of the mass–radius sequence along which the radius is highly sensitive to increasing mass due to EOS softening or phase transitions (e.g., Drago et al. 2014; Alford & Han 2016; Alford & Sedrakian 2017). However, one mass is unconstrained (disk modeling) while the other is constrained at the ∼10%-level (*NICER* pulse-profile modeling).

Overall, these recent publications have estimated NS radii in the range of 12–14 km, which is compatible with our PSR J0030+0451 radius estimate. We note that equating accurately measured radii of distinct NSs should elicit agreement if (i) the EOS is shared from core to crust, meaning that perturbative effects attributed to crust composition, temperature, and magnetic field strength are sufficiently small in the context of measurement precision; (ii) the EOS is of a nucleonic composition that supports NSs with similar radii over a wide range in mass (0.8 to $\sim 2.0\ M_\odot$); and (iii) differences due to spin-dependent rotational deformations are accounted for or are small enough to justify neglecting (see the discussion in the second and third paragraphs of Section 4.2, and in Raaijmakers et al. 2019). For other families of EOS, such as those involving quarks or hyperons in hybrid stars (e.g., Zdunik & Haensel 2013) or baryon resonances (Drago et al. 2014), the radius may be (highly) sensitive to increasing mass. In such cases, we would expect to find NSs whose radii differ by several km, reinforcing the importance of jointly estimating both the radius and the mass of each member of a population of NSs.

Constraints on NS masses and tidal deformabilities are now also being reported based on the first binary NS merger gravitational wave event, GW170817 (Abbott et al. 2018, 2019). These can be translated into constraints on mass and radius. The inferred values vary somewhat depending on modeling and prior assumptions. Abbott et al. (2018), for example, employed two methods: the first, which did not assume that both stars had the same EOS, yielded radii of $10.8^{+2.0}_{-1.7}$ and $10.7^{+2.1}_{-1.5}$ km for the two stars, with masses in the range 1.16–1.62 $M_\odot$; the second, which assumed a common EOS, yielded a radius of 11.9 ± 1.4 km for both stars and masses in the range 1.18–1.58 $M_\odot$ (all results 90% credible intervals). De et al. (2018), who also assume a common EOS, report radii of $10.7^{+2.1}_{-1.6} \pm 0.2$ km and masses in the range 1.12–1.67 $M_\odot$ (90% credible interval). Constraints taking into account additional information tend to support slightly larger mean values for the radius, and smaller uncertainties, e.g., those derived from the electromagnetic counterpart ($12.4^{+1.1}_{-0.4}$ km, 2σ confidence interval; Most et al. 2018); and those invoking a theoretical minimum (Tews et al. 2017) for neutron matter pressure ($11.4^{+1.9}_{-0.8}$ km, 90% credible interval marginalizing over mass; T. Zhao & J. M. Lattimer 2019, in preparation). The results of Abbott et al. (2018) and De et al. (2018), which employ the assumption that both stars share the same EOS, suggest that the radii of the two stars are nearly equal despite the fact that the mass ratio of the stars could lie between 0.7 and 1.0 with almost uniform probability. Let us make the assumption that PSR J0030+0415 also has the same radius as the binary members to well within the posterior uncertainty on each star: the radius reported in this Letter, inferred from *NICER* data, is more consistent with values in the upper ranges emerging from the gravitational wave analysis of GW170817. Larger radii would only be consistent with mass ratios closer to unity. It follows that if the three stars have nearly the same radius, and if the mass ratio of GW170817 was near its lower limit of 0.7, the common radius should be at the lower end of the *NICER* range. Electromagnetic observations might suggest relatively large amounts of dynamical ejecta, which would favor mass ratios considerably less than unity (Radice et al. 2018).

What are the prospects for improving constraints on mass and radius for PSR J0030+0451? Unfortunately we cannot obtain an independent constraint on the mass for PSR J0030+0451 because it is not in a binary, unlike some of *NICER*'s other MSP targets. Our model of the *NICER* background (particle radiation and diffuse sky terms) is, however, expected to improve without needing to wait for input from other telescope missions; understanding this background accurately will prove crucial, as discussed in Section 4.1.2, and may well

[78] The real surface could thus in principle enclose the ISCO associated with the ambient spacetime solution.

[79] For example, via a binary mass function together with classification of the companion star.

[80] Or at least a small dimensionless spin, despite the 581.0 Hz spin frequency (Ludlam et al. 2017, and references therein).

[81] Where as discussed above the rotational deformation of the surface is small for PSR J0030+0451.

impact mass–radius estimation. A longer total exposure time could also certainly be accumulated (and indeed the *NICER* team anticipate doing this). Previous studies that have examined how posterior estimation of mass and radius is sensitive to factors such as geometry, spin rate, and the number of source counts in the event data, indicate that constraining power increases as the square root of the number of counts (Lo et al. 2013; Psaltis et al. 2014). However, those studies all assumed a single circular single-temperature hot spot, not a more complex hot region configuration such as those we have considered and inferred here. While it is likely that gathering more data will improve the joint constraint on mass and radius (without reference to an EOS model), the precise observing time required to achieve a given level of precision cannot be estimated robustly without further study.

Our report here is encouraging in terms of prospects for other *NICER* targets such as PSR J0437−4715: the pulsar mass is constrained independently to within a few percent via radio pulsar timing because it is in a binary system (Reardon et al. 2016). If such a constraint had been available for PSR J0030 +0451, it is clear that we could have obtained a comparable posterior uncertainty on the inferred radius of a few percent.

4.3. EOS Implications

One of the primary goals of mass–radius inference is to use posterior information[82] to infer the properties of the dense matter EOS, if such information is deemed sufficiently likelihood-dominated to warrant the study. Studies have utilized joint mass–radius posteriors inferred from X-ray spectral modeling of bursting and quiescent NSs (Steiner et al. 2010, 2013; Özel et al. 2016; Raithel et al. 2017). Moreover, studies have utilized mass and tidal deformability constraints derived from analysis of the NS binary merger event GW170817 (Abbott et al. 2018; Annala et al. 2018; Lim & Holt 2018; Malik et al. 2018; Most et al. 2018; Tews et al. 2018b; Carson et al. 2019; Li & Sedrakian 2019; Montaña et al. 2019), and consideration is already being given to combining constraints from electromagnetic and gravitational wave analysis (Kumar & Landry 2019; McNeil Forbes et al. 2019; Weih et al. 2019).

Given a suitable model for the EOS (see, e.g., Read et al. 2009; Raithel et al. 2016; Lindblom 2018; Tews et al. 2018a) there are two approaches to EOS inference: one is to jointly infer the EOS parameters (and central densities) directly from the data (e.g., pulse-profile data); the other is to jointly infer EOS parameters from per-source nuisance-marginalized likelihood functions of exterior-spacetime parameters (e.g., gravitational mass and equatorial radius). The former approach is *at least* as computationally intensive as the direct mass–radius inference reported in this Letter; the latter approach is less computationally intensive *given* archival likelihood function information about exterior spacetime parameters (Riley et al. 2018). In any case, care is required in both overall approach and the selection of (interior source matter and exterior spacetime) model parameterization and priors (Carney et al. 2018; Raaijmakers et al. 2018; Riley et al. 2018; Greif et al. 2019; Landry & Essick 2019). For our analysis here we deliberately defined a joint flat prior density function for M and $R_{\rm eq}$, with the intention that the posterior density function can be invoked as a likelihood function marginalized over all nuisance parameters. We explore the dense matter EOS implications of the inferred mass, radius, and compactness for PSR J0030 +0451 in an accompanying Letter (Raaijmakers et al. 2019), following the approach to EOS inference outlined in Greif et al. (2019).

4.4. Implications of the Surface Heating Configuration

We constructed a sequence of simple models for the properties of the two hot regions, nevertheless motivated by (numerical) pulsar theory. We considered models in which the regions were related via antipodal reflection symmetry with respect to the stellar origin, and models that do not impose such symmetry, meaning that their properties and location were described with distinct parameters (with the restriction that the regions cannot overlap). The models included simply connected circular and crescent regions, and rings (whose hole and annulus are concentric or eccentric), each filled with single-temperature material. The models also included annular (ring) regions whose concentric hole is filled with material of finite temperature distinct from that of the material in the annulus.

We were able to rule out the hot regions being antipodal and identical based on clear systematic structure in the residuals between data and model a posteriori; moreover, a model wherein the regions are both assumed to be simply connected circular single-temperature spots was strongly disfavored. We inferred that the regions are configured to exist in the same rotational hemisphere: one region subtends an angular extent of only a few degrees (in spherical coordinates with origin at the stellar center) but whose other structural details we are insensitive to; the other region is far more azimuthally extended,[83] in the form of a narrow hot crescent. The inferred effective temperature of the `NSX` atmosphere was remarkably consistent across all models considered—for both regions—at $\sim 1.3 \times 10^6$ K. The `ST+PST` model exhibited the largest background-marginalized likelihood function values in the typical set of a posterior mode. Figure 17 renders a representative configuration from the posterior mode, and Table 2 reports the marginal credible intervals for the hot-region parameters. Note that `ST+PST` includes within prior support, configurations wherein the regions are similar or even congruent in shape, and a priori favors (albeit weakly) smaller angular extents—the heating asymmetry is emergent in spite of this.

One of the principal astrophysical questions arising is how such a heating configuration can occur. It appears to be incompatible with magnetospheric current heating at the footpoints of a simple near-centered dipole magnetic field, and is likely to require some higher-order multipole structure (Barnard & Arons 1982; Gralla et al. 2017; Lockhart et al. 2019). We now need to determine the type of field configuration required, the magnitude of the different moments, and whether this is feasible on physical grounds. Consideration will also need to be given to how magnetospheric currents actually map to temperature fields on the stellar surface. We note that there are clear similarities between the inferred `ST+PST` configuration and the current heating distribution contemporaneously derived by Lockhart et al. (2019, see their Figure 6 in particular) via quadrupolar extension of the magnetic field, considering that the heating ring is asymmetric

[82] Strictly, likelihood information (Riley et al. 2018).

[83] In spherical coordinates with polar axis coincident with the stellar rotation axis.

with respect to the dipole axis, and closely resembles a large-scale arc-like hot region. There are, however, also some differences: Lockhart et al. (2019) restricted their study to configurations where the center of the heated ring is antipodal to the heated spot (both of which emit as approximate blackbodies), an assumption that would need to be relaxed to recover our preferred configuration.

There are also implications for pulsar emission in wavebands other than the X-ray (see the reviews by Grenier & Harding 2015; Cerutti & Beloborodov 2017). *If a multipolar field structure is required to explain the surface temperature field, how does this affect radio and gamma-ray emission generated further out in the magnetosphere? Could the multipole structures persist out to the point where emission in these wavebands is thought to happen?* Quadrupole fields fall off faster than dipole fields, as the inverse fifth power of the radius, but the radius beyond which the field is predominantly dipolar would depend on the ratio of quadrupole to dipole components. Most current models of radio and gamma-ray emission assume that the field structure is a centered dipole (e.g., Radhakrishnan & Cooke 1969; Gil et al. 1984; Kijak & Gil 2003; Dyks & Harding 2004; Johnson et al. 2014), and this would need to be revisited.

There are also questions pertaining to stellar evolution. NSs are born with a field structure that could be quite complex as a result of the supernova process (Ardeljan et al. 2005; Obergaulinger & Aloy 2017) but various diffusive evolutionary processes can subsequently modify field structure even for isolated NSs (Reisenegger 2009; Viganò et al. 2013; Mitchell et al. 2015; Gourgouliatos & Hollerbach 2018). In addition, rotation-powered MSPs are thought to go through an extended period of accretion-induced spin-up to reach the observed spin rates. The accretion process may also act to modify the field structure (Romani 1990; Melatos & Phinney 2001; Payne & Melatos 2004). It remains to be determined whether a complex multipolar field structure could emerge and survive from birth, or be generated during the accretion process. If such a field structure is present or evolves during the accretion phase, there will also be implications for the spin-up process and for X-ray emission during that phase of NS evolution. If the magnetic field were to channel accreting material onto two magnetic polar caps on the same hemisphere, for example, this would certainly affect the emission from accreting MSPs (Long et al. 2007, 2008; Patruno & Watts 2012). Whether the star is even visible as an accreting pulsar will depend not only on the geometry of the hot regions where accreting material impacts the star, but also on whether the observer views the hemisphere containing the polar caps or the other one. The flow at the inner-edge of the accretion disk, a strong source of potentially variable X-ray emission, would also be affected by a multipolar field structure. Finally, extremely off-center dipoles or strong non-centered multipole fields, for instance, will produce asymmetries in the Poynting flux of low-frequency radiation parallel to the spin axis, with consequences that could include a large space velocity (Harrison & Tademaru 1975). Lommen et al. (2006) found that PSR J0030+0451 has a relatively low transverse space velocity—a property that is potentially in contention with a field far from that of a centered dipole—although since PSR J0030+0451 is isolated its space velocity would also depend on how the binary was disrupted after the spin-up phase. Discussion on some of these issues is reserved for an accompanying Letter (Bilous et al. 2019). However, further work on the implications of the inferred configuration for pulsar field structure, emission mechanisms, and stellar evolution is certainly required. It is clear that the mass and radius inferred for PSR J0030+0451 depend strongly on the surface radiation field models (including prior support) that we have explored. Further study may show that our models are either not general enough or too general (for example with respect to the prior support). For example, we may find that temperature gradients in the hot regions cannot be neglected; or that field stability considerations impose a minimum angular separation between the polar caps—and possibly, by extension, the hot regions—larger than found in our analysis; or that no magnetospheric model can generate a heating configuration in which one region is a small-scale spot while the other an azimuthally extended crescent.

There are a number of computational aspects for pulsar theorists to consider when developing surface heating models suitable for statistical inference. We were only able to consider the configurations offered uniquely by `ST+PST` as we approached the limit of our computational resource allocation; we can therefore provide stronger guarantees about the accuracy of the posterior computation for the lower-complexity models with more than one run (`ST-U` and `ST+CST`). In the future, we suggest that more resources be devoted to models at the `ST+PST`-level of complexity, in particular for work on (i) parameterization of hot regions with more complex topologies and/or boundaries, (ii) their efficient numerical resolution for likelihood function evaluation, and (iii) exploration of the associated parameter space via sampling or other methods.

We also suggest that additional resources be devoted to research avenues such as self-consistent theory and computation of surface heating by magnetospheric currents, for the purpose of statistical computation. A question may be posed as to the generation of highly non-dipolar surface temperature fields—e.g., arc- or ring-like heating distributions—which can be (i) parameterized such that approximate representations find compromise between accuracy and complexity, capturing the facets considered most crucial to signal generation, perhaps as a sequence of models increasing in complexity; and (ii) built into efficient software implementations. Progress on such fronts should encourage a bridge to form between phenomenological efforts and more self-consistent theory for the purpose of efficient statistical computation, and may also offer a way in which to connect distinct theoretical models (in an approximative manner) on a continuous space.

4.5. Pulse-profile Modeling for Other Types of NS

NICER is the first mission designed specifically to use the pulse-profile modeling technique to infer the mass and radius of NSs. As a soft X-ray telescope with an effective area of less than a square meter, it is optimized for applying the technique to MSPs, which have soft, stable pulse profiles meaning we can use multiple exposures taken over a long baseline to accumulate a sufficient number of events to statistically probe MSP physics. However, the technique can also be applied to other NSs with emission modulated by rapid rotation: accretion-powered pulsars and thermonuclear burst oscillation sources.

In accretion-powered pulsars (see Patruno & Watts 2012, for a review), accreting material is channeled by the magnetic field onto the magnetic polar caps and the pulsed emission has two main components: thermal emission from the heated vicinity of

the accretion impact zones, and nonthermal emission from the shock in the accretion funnel (Poutanen & Gierliński 2003). A third pulsed component may arise due to reflection from the accretion disk (Wilkinson et al. 2011). Thermonuclear burst oscillations (see Watts 2012, for a review) are generated by rotational modulation of global asymmetries that form in a surface radiation field during thermonuclear (Type-I X-ray) bursts; such a burst occurs in the ocean of an accreting NS, driven by unstable burning of accreted hydrogen, helium, or carbon (see, e.g., Galloway et al. 2008). The precise mechanism driving the *detectable* asymmetry (oscillations) is not clear: possibilities include disrupted flame spread (Spitkovsky et al. 2002; Cavecchi et al. 2013), large-scale waves in the burning ocean (Heyl 2004; Piro & Bildsten 2005; Chambers et al. 2019), or patterns triggered by convection (Garcia et al. 2018).

Accretion-powered pulsations and thermonuclear burst oscillations are radiatively harder ($\sim$1–30 keV) than the pulsations of MSPs. Accumulating the requisite number of photons for tight constraints, in a realistic observing time, also requires a telescope with an effective area of several square meters (Watts et al. 2016; Watts 2019). Several mission concepts are currently being developed for large-area broadband X-ray timing telescopes that would access a larger, fainter population of MSPs than we can observe with *NICER*, and fuel pulse-profile modeling for accretion-powered pulsars and thermonuclear burst oscillators: these include the *eXTP* (Watts et al. 2019; Zhang et al. 2019), and the *STROBE-X* (Ray et al. 2019). For an idea of the constraints that can be delivered by pulse-profile modeling and inference using existing data, see the Salmi et al. (2018) analysis of *Rossi X-ray Timing Explorer* observations of the accretion-powered MSP SAX J1808.4−3658.

Our uncertainty toward surface heating physics affects pulse-profile modeling not only for rotation-powered MSPs, but also accretion-powered pulsars and thermonuclear burst oscillators. Although the general mechanism that gives rise to the pulsed components in accretion-powered pulsars is clear, the surface *and off-surface* temperature field and local comoving beaming function—particularly from the accretion-funnel shock—are a priori highly uncertain for any given source. Thermonuclear burst emission has a well-understood (local comoving) beaming function due to the sub-surface thermal origin (Suleimanov et al. 2011b), but the mechanism for generating asymmetries in the global surface radiation field, which in turn generate such rotational oscillations, remains highly uncertain. Reducing the remaining theoretical uncertainties, and developing physically motivated parameterized models of the surface temperature field, will be important. However, our analysis of *NICER* data provides an important real-world demonstration that pulse-profile modeling is a viable technique for constraining masses and radii of NSs, and that the analysis machinery can operate on somewhat flexible models, with weakly informative priors, for both source and background emission.

5. Conclusion

We reported on pulse-profile modeling efforts for the rotation-powered millisecond X-ray pulsar PSR J0030+0451, conditional on *NICER* data. We focused on PSR J0030+0451, a challenging source due to the absence of an independent constraint on the gravitational mass (compared to *NICER*'s other primary target PSR J0437−4715). Nevertheless, PSR J0030+0451 was selected as the optimal source to demonstrate simultaneous inference of gravitational mass and equatorial radius given weakly informative priors, and to develop our analysis procedures.

The mass and radius each have marginal posterior 68% credible interval half-widths at the $\sim$9%–11% level, conditional on the *NICER XTI* event data. These constraints are consistent with those emerging from both gravitational wave analysis and X-ray spectral modeling, and are expected to improve with further exposure. The compactness is constrained more tightly, at the $\sim$6% level. Prospects for *NICER* delivering tight constraints for rotation-powered MSPs where the mass is known independently to uncertainties of a few percent are clearly excellent.

In addition to inferring properties of the spacetime (mass, radius, and compactness), we were also able to infer the properties of the thermally emitting hot regions that we assume generate the pulsations. For the specific set of models that we considered, the inferred configuration has both hot regions in the same rotational hemisphere, with one hot region being a small spot and the other an azimuthally extended narrow crescent. Models wherein the hot regions are antipodal are strongly disfavored, implying a complex offset dipolar and multipolar field structure that, if accurate, has major implications for both pulsar emission and stellar evolution.

We thank the anonymous referees for their suggested improvements to this work. This work was supported in part by NASA through the *NICER* mission and the Astrophysics Explorers Program. T.E.R., A.L.W., and A.V.B. acknowledge support from ERC Starting grant No. 639217 CSINEUTRONSTAR (PI: Watts). A.L.W. would also like to thank Ralph Wijers for environmental support. This work was sponsored by NWO Exact and Natural Sciences for the use of supercomputer facilities, and was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. This research has made extensive use of NASA's Astrophysics Data System Bibliographic Services (ADS) and the arXiv. R.M.L. acknowledges the support of NASA through Hubble Fellowship Program grant *HST*-HF2-51440.001. S.G. acknowledges the support of the Centre National d'Études Spatiales (CNES). W.C.G.H. appreciates use of computer facilities at the Kavli Institute for Particle Astrophysics and Cosmology. S.M.M. thanks NSERC for support. J.M.L. acknowledges support from NASA through Grant 80NSSC17K0554 and the U.S. DOE from grant DE-FG02-87ER40317.

Facility: NICER (Gendreau et al. 2016).

Software: Python/C language (Oliphant 2007), GNU Scientific Library (GSL; Gough 2009), NumPy (van der Walt et al. 2011), Cython (Behnel et al. 2011), SciPy (Jones et al. 2001), OpenMP (Dagum & Menon 1998), MPI (Forum 1994), *MPI for Python* (Dalcín et al. 2008), Matplotlib (Hunter 2007; Droettboom et al. 2018), IPython (Perez & Granger 2007), Jupyter (Kluyver et al. 2016), TEMPO2 (`photons`; Hobbs et al. 2006), PINT (`photonphase`; https://github.com/nanograv/PINT), MultiNest (Feroz et al. 2009), PyMultiNest (Buchner et al. 2014), *GetDist* (https://github.com/cmbant/getdist), *nestcheck* (Higson 2018; Higson et al. 2018, 2019), *fgivenx* (Handley 2018), *X-PSI* (`v0.1`; https://github.com/ThomasEdwardRiley/xpsi; Riley & Watts 2019).

Appendix A
Posterior Computation

Here we describe how we derived posterior inferences conditional on the models defined in Section 2.

A.1. Nested Sampling

We implemented nested sampling using the open-source software MultiNest[84] (Feroz & Hobson 2008; Feroz et al. 2009, 2013; Buchner et al. 2014). For a subset of model nodes in Figures 10 and 12 we executed two production runs—i.e., we generated two realizations of a particular stochastic sampling process—with the following resolution settings: the number of *active* (or *live*) points was 10^3; the bounded-hypervolume expansion factor was (at least[85]) 0.3^{-1}; the termination condition as a function of iteration number i was, schematically,

$$\ln(\widehat{\mathcal{Z}}_i + \widehat{\Delta\mathcal{Z}}_i) - \ln(\widehat{\mathcal{Z}}_i) < x \Longrightarrow \frac{\widehat{\mathcal{Z}}_i}{\widehat{\mathcal{Z}}_i + \widehat{\Delta\mathcal{Z}}_i} > e^{-x}, \quad (10)$$

where $\widehat{\mathcal{Z}}_i$ is an estimator for the evidence integral up to iteration i over estimated prior mass $(1 - \widehat{X}_i)$, and $\widehat{\Delta\mathcal{Z}}_i$ is an estimator for the *maximum remaining* contribution to the evidence over the complementary estimated prior mass $\widehat{X}_i$. As $x \to 0$, $e^{-x} \to 1^-$ and the estimated amassed evidence entirely dominates the estimated remaining evidence; we generally set $x = 10^{-1}$, which is five times smaller than suggested by Feroz et al. (2009).[86] The number of active points is chosen to be a number that is larger than will be typically reported in the literature for similar dimensional problems (perhaps with simpler distributional structure), or recommended by the authors of MultiNest, for a compromise between resource consumption and accuracy. The hypervolume expansion factor is, for all but one run,[87] greater than the number recommended by the authors of MultiNest for accurate evidence estimation. The combination of a number of active points—which is between 40 and 80 times larger than the dimensionality of the sampling space—and the recommended expansion factor, targets posterior computation with an implementation-specific error[88] that is smaller than the error due to the inherent stochasticity of Monte Carlo sampling.

The *constant-efficiency* MultiNest bounding variant should in general be avoided where tractability is not compromised to reduce risk of under-sampling when a hyper-ellipsoidal decomposition does not conform well to likelihood level hypersurfaces; we did not use this bounding variant. However, in the higher-dimensional contexts that we are approaching in this work, it could be useful to activate this bounding variant if integration is forecasted to consume too many computing resources (e.g., the acceptance fraction drops too low below 10^{-3} when using a high-performance system and an expensive likelihood function). Imposing a target acceptance fraction may be useful for initial exploratory runs in order to probe for configurations with high (marginal) likelihood—a weak indicator of potential model performance if future efforts achieve greater computational efficiency. When *constant-efficiency* mode is activated the evidence estimates should be assumed to be positively biased unless importance nested sampling is also activated (which can consume an appreciable fraction of the memory on a typical supercomputer node for $\sim 10^8$ likelihood function evaluations). The acceptance fraction decaying to such levels is indicative that the minimum-bounding hyper-ellipsoidal decomposition does not conform sufficiently well to the nested likelihood function level hypersurfaces, resulting in (i) a large excess of hypervolume exterior to the likelihood surface but within the bounding union of ellipsoids, and/or (ii) the union of bounding ellipsoids exhibits a large fractional overlap. In such cases one could construct an alternative parameterization of the problem that is more tractable with the MultiNest bounding algorithm (in the native sampling space), or use an alternative (nested) sampling algorithm.

The mode-isolation sampling variant was *not* activated unless stated otherwise. This variant isolates the evolution of local modes whose hyper-ellipsoidal clustering decompositions are mutually non-overlapping; upon isolation, the active points of sampling threads constituting each mode cannot migrate between modes at subsequent iterations—they are locked in, which can alleviate premature mode deactivation. Sampling resolution, however, can be absorbed by unimportant modes, which is undesirable. The mode separation sampling variant generates distinct statistics for each mode and labels actives points according to mode association.

With the above settings, among ulterior numerical settings for marginal likelihood function evaluation, *X-PSI* typically executes $\mathcal{O}(10^7)$ marginal likelihood function evaluations in $O(10^4)$ core hours via MultiNest.

A.2. Summary of Information Presentation

We now summarize how we opt to present probabilistic information in the form of figures and tables. We condition on a set of model variants that are treated in effectively the same manner, and the information presentation is consistent for these models; we thus here describe the information once. Figures and tables are displayed in the main body—see Section 3—for the model we deem to perform best considering both prior and posterior predictive measures and checks; we refer the reader to Table 2 and Figures 13 through 21 for reference.

A.2.1. Graphical Posterior-predictive Checking

In a figure such as Figure 13 we display salient information for assessing performance of a model in isolation. Our generative modeling process is fundamentally built on the statement that we do not believe the true data-generating process exists within the model space considered (refer to Section 2.3); however, the models may, for the purpose of generating data, be deemed adequate approximations. We aim

[84] MultiNest `v3.11` can be located at https://github.com/farhanferoz/MultiNest with SHA1-hash 4b3709c. PyMultiNest `v2.6` can be located at https://github.com/JohannesBuchner/PyMultiNest SHA1-hash 5d8c103.

[85] A *lower bound* for reasons pertaining to the implementation of the joint prior distribution; the lower bound itself is numerically transformed into an appropriate MultiNest setting in order to achieve at least this desired expansion factor. Further detail is beyond the scope of this description (refer to Riley & Watts 2019).

[86] The minimum tolerance used for any process was $x = 10^{-3}$ to confirm that termination was not premature.

[87] The `ST+PST` model was implemented as we were exhausting our computational resources and thus we lowered resolution—increased the nested-sample acceptance fraction—to ensure completion of an exploratory run—see Table 2.

[88] The error due to not sampling from the entire prior hypervolume subject to a given likelihood function constraint.

to graphically approximate an answer to the question: *does the model generate,* a posteriori, synthetic data that emulates structure in the real event data?

Graphical posterior predictive checking here relies on the power of human identification of systematic structural differences, which if physically characterized can drive future model development toward better-performing approximations. Structural differences include spectro-temporal correlations between random variates assumed to be statistically independent (e.g., the ST-S version of Figure 13 in the online figure set), and inaccurate noise modeling leading to under- or over-estimation of the variance of random variables; the latter also may manifest due to excessive, non-physical predictive complexity (over-fitting). In the *top* panel of such a figure we display the data set for convenience. In the middle panel we choose to display the posterior-mean Poisson expected count numbers:

$$\lambda_{ij} \coloneqq \mathbb{E}_{\pi(\boldsymbol{\theta})}[c_{ij}(\boldsymbol{\theta})] = \int_{\mathcal{S}} c_{ij}(\boldsymbol{\theta}, B_i)\pi(\boldsymbol{\theta}, B_i \mid \boldsymbol{d})d\boldsymbol{\theta} \approx \sum_k w_k c_{ij}(\boldsymbol{\theta}_k), \tag{11}$$

where $\boldsymbol{\theta}_k \sim \pi(\boldsymbol{\theta} \mid \boldsymbol{d})$ are samples with normalized importance weights w_k drawn from the background-marginalized posterior density $\pi(\boldsymbol{\theta} \mid \boldsymbol{d})$, and $\mathcal{S}$ is the prior support. Crucially, we do not sample the joint posterior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{B} \mid \boldsymbol{d})$ because of prohibitive scaling of expense with dimensionality;[89] therefore, in order to compute a data-space posterior-mean quantity, we opt to maximize the *conditional* likelihood function $L(B_i; \boldsymbol{\theta}_k)$ with respect to each background count-rate parameter B_i, generating an estimator $\widehat{B}_i$, such that $c_{ij}(\boldsymbol{\theta}_k) \coloneqq c_{ij}(\boldsymbol{\theta}_k, \widehat{B}_i)$. In the *bottom* panel we display standardized residuals between data count numbers and the quantities λ_{ij}, where we define Poisson-random variables $x_{ij} \sim p(x_{ij} \mid \lambda_{ij})$. We consider the set of figures associated with Figure 13 as *graphical posterior-checking* plots. If no clear systematic structure manifests during posterior predictive checking and the posterior-predictive distribution is well approximated by a sampling distribution conditional on a parameter vector, we consider the model to be appropriate for predicting future observations against which the model may be falsified.

Bayesian global performance measures—such as the evidence—are useful for determining the utility of increments in the complexity of a generative model. However, while being a target of our posterior computation, prior predictive probabilities are known to not be universally robust and can be sensitive to prior choices and parameterization, especially in phenomenological contexts. Thus in order to assess performance and identify model features that conform well to data structure, it is necessary—and natural—to also visually inspect predictions a posteriori. We consider other measures such as the expected posterior utility (i.e., power for future statistical falsification) beyond the scope of this work (Vehtari & Ojanen 2012).

In this work we refrain as much as possible from invoking *point measures*—based on parameter vector point estimates—to quantitatively summarize and compare models. The reason we do this is twofold: (i) a point estimate is usually the parameter vector that is estimated to globally optimize some quantity, and subsequent to estimation all information encoded by the posterior distribution is not explicitly regarded; and (ii) the target of our statistical computation (nested sampling) is not any particular point measure, but instead to draw samples from the posterior *typical set* for the purpose of estimating posterior integrals. It follows that stronger guarantees can be made about the accuracy of statistical estimators that are posterior integrals than can be made about those which are point estimates; we therefore view integral estimators as generally more helpful and robust than point estimators.

As an example, consider the canonical estimation of the parameter vector that globally maximizes the likelihood function conditional on some model. In our case, our posterior computation does not target accurate maximization, and we are forced to marginalize over the phase-invariant (background) count-rate parameters in order to define a nested-sampling space whose dimensionality is not prohibitive. We thus do not guarantee that the sample (parameter vector) $\hat{\boldsymbol{\theta}}_{\mathrm{BMML}}$ that reported the highest *background-marginalized* likelihood function value among identified posterior modes, together with the background count-rate vector $\hat{\boldsymbol{B}}_{\mathrm{CML}}$ that maximizes the *conditional* likelihood function given $\hat{\boldsymbol{\theta}}_{\mathrm{BMML}}$, is an adequate estimator of the joint vector $(\boldsymbol{\theta}, \boldsymbol{B})_{\mathrm{ML}}$ that globally maximizes the likelihood function. In Figure 23 we consider the joint vector $(\hat{\boldsymbol{\theta}}_{\mathrm{BMML}}, \hat{\boldsymbol{B}}_{\mathrm{CML}})$: empirically we see that all but one model (ST-S) has, within prior support, parameter vectors that perform adequately as point-measure predictors of structure in the real count-number data.

A.2.2. Parameter Kernel Density Estimation and Credible Regions

We applied the post-processing module of the *X-PSI* package (v0.1; Riley & Watts 2019). *X-PSI* wraps—combines and adds functionality—to several other packages for statistical computation: relevant here for Gaussian kernel density estimation (KDE) is *GetDist*.[90] We do not use *GetDist* to compute numerical one-dimensional credible intervals on each parameter, but *GetDist* is used whenever KDE is necessary for post-processing (posterior) samples into estimators: for calculation of the plotted one- and two-dimensional density functions, for calculation of the plotted two-dimensional (joint) credible regions, and for one-dimensional KL divergence estimation where both the marginal posterior and marginal prior density functions need to be estimated from samples.

GetDist can execute smoothing using either a manual Gaussian kernel bandwidth or an automatic optimized bandwidth. For all parameters and models, for simplicity and consistency, we invoke a manual bandwidth of 0.4 times the

[89] If our set of samples was drawn from the joint posterior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{B} \mid \boldsymbol{d})$ we could better approximate the expectation integral in Equation (11), or at greater cost, approximate the posterior-predictive probability mass distribution in data space:

$$p(x_{ij} \mid \boldsymbol{d}) = \int p(x_{ij} \mid \boldsymbol{\theta}, B_i)\pi(\boldsymbol{\theta}, \boldsymbol{B} \mid \boldsymbol{d})d\boldsymbol{B}d\boldsymbol{\theta} \approx \sum_k w_k p(x_{ij} \mid \boldsymbol{\theta}_k, B_{i,k}),$$

whose expectation $\mathbb{E}[x_{ij}]$ and variance $\mathbb{V}[x_{ij}]$ are indicators of residual structure in the data relative to the model a posteriori. The posterior predictive distribution may also be constructed by jointly generating samples from the Bayesian joint distribution: $\boldsymbol{x}_k \sim p(\boldsymbol{x} \mid \boldsymbol{\theta}_k, \boldsymbol{B}_k)$ and $(\boldsymbol{\theta}_k, \boldsymbol{B}_k) \sim \pi(\boldsymbol{\theta}, \boldsymbol{B} \mid \boldsymbol{d})$.

[90] https://github.com/cmbant/getdist. Specifically, tag v0.3.1, with some minor plotting customization locatable at https://github.com/ThomasEdwardRiley/getdist/tree/customization, with SHA1-hash 61f69d0. The technical KDE notes for *GetDist* are located at https://cosmologist.info/notes/GetDist.pdf.

estimated parameter standard deviations, based on the *GetDist* guidelines for choosing analysis settings. Moreover, *GetDist* performs various smoothing corrections to avoid smoothing bias as one attempts to mitigate finite-sample noise: (i) smoothed-density correction near hard one-dimensional parameter bounds defined through the prior support, with capability to estimate a finite local density function gradient; and (ii) iterative multiplicative bias-correction to the estimated density function to nullify over-smoothing. Density estimation near non-trivial prior support boundaries in two parameters does not account for the local boundary, but this is not problematic for our work here. We roll with the default bias-correction settings: a linear boundary kernel (in both one- and two-dimensional spaces) and zeroth-order multiplicative bias-correction (a single application with no iterations).

A.2.3. Error Analysis for Statistical Estimators

Two frameworks now require distinction: (i) the parametric probabilistic framework of the generative model (a Bayesian context) for the data; and (ii) a non-parametric probabilistic framework that operates with realizations of a stochastic sampling process, which in turn operates on a deterministic target probability distribution defined in the parametric framework given data and a generative model (e.g., Skilling 2006; Higson et al. 2018). The sampling process has a mixture of known and approximated properties, including fixed non-physical settings. Estimators derived from the process output are stochastic and we are interested in their distributions.

The purpose of executing some number of computationally expensive *repeats* is for error analysis—specifically the estimation of implementation-specific error (Higson et al. 2019) pertaining to sampling from the joint prior subject to a likelihood function constraint (a more thorough review of error analysis techniques may be found in Riley & Watts 2019). One way to approach this problem is to compute posterior marginal density functions for many bootstrapped realizations of each stochastic run and then graphically probe for the manifestation of implementation-specific error (Higson et al. 2019).

X-PSI wraps the package *nestcheck*[91] (Higson 2018; Higson et al. 2018, 2019) to access existing error analysis routines. As an example of application in this work, consider Figure 20: in the associated online figure set, where we supply a higher-resolution version of each panel, we display the marginal posterior density distributions for each parameter as a set of shaded error bands. The shaded bands represent the distribution of posterior density, at each parameter value, based on simulated nested sampling process realizations; the posterior density function is estimated for each realization with *GetDist* as described above in Appendix A.2.2. The colorbar denotes the percentage of realizations spanned by a band with a given shade, where each band connects intervals (at each parameter value) containing the highest *realization* density of *posterior density*. Note that the colorbar is not associated with the shaded joint density distributions in the off-diagonal plots. The contours in the on-diagonal panels thus encode information on the variation of the parameter kernel density estimator due to the inherent stochasticity of each of sampling processes; the estimator distributions are connected as a function of each respective parameter to delineate the behavior of the probability mass. If the member processes are deemed to exhibit consistency under visual inspection, the combined process may be invoked to estimate distributions of estimators.

We apply *nestcheck* routines to bootstrap re-sample threads and simulate weights for the following: the one-dimensional quantiles in posterior mass for each parameter, which are in turn used to report the credible intervals; the global and parameter-by-parameter KL divergences; and the evidence. The numerical values we report are in some cases (`ST-S`, `ST-U`, and `ST+CST`) derived by combining two realizations into a single realization with $\mathcal{O}(10^3)$ active points (sampling threads), if the runs are considered sufficiently consistent and exhibit sufficient resolution for our purposes here. However, due to computational expense we could only afford *at most* two runs, and for both `ST+EST` and `ST+PST` we were limited to a single run with $\mathcal{O}(10^3)$ active points. Moreover, note that processes executed by the mode-separation algorithm are incompatible with the notion of process combination: the theory nor software implementation exists at the time of writing, and therefore we only display such a run—when relevant and available—in the posterior figures, but do not use it to calculate numerical estimators.

A.2.4. Estimating Posterior Information Gain

In a Bayesian context, the KL divergence (Kullback & Leibler 1951) can be applied as a non-negative real scalar[92] measure of posterior information *gain* about a parameter—or jointly about parameters—of a generative model, conditioned on the data set; it is also known via information-theoretic interpretation as *relative entropy*. Equivalently it is the *posterior-expected additional* number of bits necessary to encode the value of a parameter sample for lossless communication between agents, *if* the (marginal) prior distribution is invoked to design an optimal encoding. Alternate interpretations exist to satiate a variety of readers—e.g., Shlens (2014).

KL-divergence maximization is central to an information-theoretic—but often in practice intractable—definition of a minimally informative prior via reference to the generative model, but without reference to the data (via data-space marginalization). While we cannot feasibly determine the maximal KL divergence with respect to the space of all proper prior density functions, the number of bits of information gain is a useful indicator of the degree to which the likelihood function dominates the information encoded in the posterior. The KL divergence has, as an example, recently been applied by The LIGO Scientific Collaboration et al. (2019) to probe posterior information gain and sensitivity to prior assumptions.

For continuous symbols, the KL divergence is defined in the limit that a discrete symbol becomes a continuous subset of $\mathbb{R}^n$; in practice, of course, computer representation of the reals is discrete. Given the existence of a known optimal prior encoding, the number of additional bits for lossless communication of a posterior sample at a given precision is, in the context of our inference problem, far less than the number of bits required to store the sample. Mathematically, in units of

[91] Specifically, tag `v0.2.0`, with customization to support use of *GetDist* KDE, locatable at https://github.com/ThomasEdwardRiley/nestcheck/tree/feature/getdist_KDE with SHA1-hash 4555df0.

[92] And thus parameterization invariant.

bits:

$$D_{\mathrm{KL}}(\pi \,||\, p) = \int_{\mathcal{S}} \pi(\boldsymbol{\theta} \mid \boldsymbol{d}) \log_2 \left[\frac{\pi(\boldsymbol{\theta} \mid \boldsymbol{d})}{p(\boldsymbol{\theta})} \right] d\boldsymbol{\theta} \approx \sum_k w_k \log_2 \left[\frac{\pi(\boldsymbol{\theta}_k \mid \boldsymbol{d})}{p(\boldsymbol{\theta}_k)} \right], \tag{12}$$

where $\boldsymbol{\theta}_k \sim \pi(\boldsymbol{\theta} \mid \boldsymbol{d})$ are samples with normalized importance weights w_k drawn from the background-marginalized density $\pi(\boldsymbol{\theta} \mid \boldsymbol{d})$. Computation of a single scalar divergence for the n-dimensional joint posterior is straightforwardly given by

$$\begin{aligned} D_{\mathrm{KL}}(\pi \,||\, p) &= \int_{\mathcal{S}} \pi(\boldsymbol{\theta} \mid \boldsymbol{d}) \log_2 \left[\frac{\pi(\boldsymbol{\theta} \mid \boldsymbol{d})}{p(\boldsymbol{\theta})} \right] d\boldsymbol{\theta} \\ &= \int_{\mathcal{S}} \pi(\boldsymbol{\theta} \mid \boldsymbol{d}) \log_2 \left[\frac{1}{\mathcal{Z}} \frac{p(\boldsymbol{d} \mid \boldsymbol{\theta})\, \cancel{p(\boldsymbol{\theta})}}{\cancel{p(\boldsymbol{\theta})}} \right] d\boldsymbol{\theta} \\ &= -\log_2 \mathcal{Z} + \int_{\mathcal{S}} \pi(\boldsymbol{\theta} \mid \boldsymbol{d}) \log_2 L(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\approx -\log_2 \mathcal{Z} + \sum_k w_k \log_2 L(\boldsymbol{\theta}_k). \end{aligned} \tag{13}$$

It is more useful, however, to compute a marginal divergence for each parameter, yielding a handle on which parameters the (marginal) likelihood function is most insensitive to in the context of the (marginal) prior, and thus to which prior assumptions global posterior inferences may be most sensitive. The parameters that exhibit the lowest marginal posterior information gain are those for which prior assumptions are generally more important to be aware of and should thus be an accurate representation of prior belief. Divergence estimation for each parameter (or jointly for $m < n$ parameters) requires more involved post-processing because kernel density estimation is performed for evaluation of the quotient of marginal densities appearing in the integrand.

Note that if the marginal KL divergence for some parameter θ is small (relative to the divergences of other parameters) but the divergence is close to the theoretical maximum expected divergence, the prior exists in the minimally informative limit while being relatively informative. For example, if the Fisher information for θ is everywhere relatively small—meaning the experiment is at most a weak probe of the parameter—the marginal posterior for θ should be dominated even by a minimally informative prior, but posterior inferences about other parameters should be insensitive to all information about θ. If, on the other hand, the likelihood function is a useful probe of θ while the marginal divergence is relatively small, the global posterior may be sensitive to prior information about θ. In practice, we determine that the parameters that typically exhibit the smallest divergences—pulsar distance D and *NICER* instrument parameters α and γ—are assigned marginal priors that are not weakly informative in the context of the likelihood function; moreover, we do not need to calculate Fisher information to understand that the likelihood function itself is a useful probe of these parameters, which exhibit degeneracies with other pulsar parameters. Therefore, we conclude that small divergences in these cases do indicate that our global posterior inferences are strongly conditional on these prior assumptions.

Appendix B
Prior Transforms

For nested sampling we aim to transform from a native sampling space—a unit hypercube $\mathcal{H} = [0, 1]^n$—to a (physical) parameter space $\mathbb{R}^n$ according to an inverse transformation of the joint prior density distribution $p(\boldsymbol{\theta})$ defined on $\mathbb{R}^n$ (typically with some compact support). We thus require implementation of a mapping $\boldsymbol{x} \mapsto \boldsymbol{\theta}$ where $\boldsymbol{x} \in \mathcal{H}$ and $\boldsymbol{\theta} \in \mathbb{R}^n$. In this appendix we give the prior transforms implemented in order to facilitate reproduction of the sampling processes, and also to provide a demonstration of some of the necessary architectural work for parameter estimation via nested sampling.

In this work we provide a summary table for each model; as an example, refer to Table 2. Within each table the joint prior density function $q(\boldsymbol{\theta})$ and its support $\mathcal{S}$ are reported. Here we give a prescription for drawing a sample from a given prior: (i) draw a sample for each parameter according to the listed sampling distribution $p(\boldsymbol{\theta})$ with support $\mathcal{S}^\dagger \subset \mathbb{R}^n$—using one-dimensional inverse sampling—to generate a candidate vector $\widetilde{\boldsymbol{\theta}} \in \mathbb{R}^n$; (ii) systematically evaluate the constraint equations to determine whether $\widetilde{\boldsymbol{\theta}} \in \mathcal{S} \subset \mathbb{R}^n$ or whether $\mathcal{S} \not\ni \widetilde{\boldsymbol{\theta}} \in \mathbb{R}^n$; and (iii) accept the candidate sample if $\widetilde{\boldsymbol{\theta}} \in \mathcal{S}$, otherwise reject the sample. The form—i.e., the relative marginal density at two values of parameter θ—is, for a subset of parameters, not given by the one-dimensional distribution $\theta \sim q(\theta)$ (whose definition in some cases, as the distance D, requires explicitly stated truncation bounds) explicitly written in the prior column: the set of constraint equations defining the joint compact support $\mathcal{S}$ often non-trivially modulate the density distribution $q(\theta)$ if θ appears in constraint equations jointly with some subset of the other parameters. If the joint prior is separable with respect θ to then by definition $p(\theta) \equiv q(\theta)$.

B.1. Gravitational Mass and Equatorial Radius

We defined our joint prior distribution of gravitational mass and equatorial radius in Section 2.4.1. We apply a technique from Appendix E of Riley & Watts (2019): inverse sampling of a joint flat density function $q(M, R_{\mathrm{eq}})$ with a trivial rectangular boundary $M \in [M_a, M_b]$ and $R_{\mathrm{eq}} \in [R_a, R_b]$, and subsequent rejection only if $R_{\mathrm{eq}} \notin [3r_g, 16]$ km.[93] In this case, let $\mathcal{H} = [0, 1] \times [0, 1]$ and let the support of $q(\boldsymbol{\theta}) = q(M, R_{\mathrm{eq}})$ be $\mathcal{S}^\dagger \subset \mathbb{R}^2$. The mapping is then $\mathcal{H} \to \mathcal{S}^\dagger$, $\boldsymbol{x} \mapsto \boldsymbol{\theta}$, where $\mathcal{S}^\dagger \supset \mathcal{S}$. For $(M, R_{\mathrm{eq}}) \in \mathcal{S}$, the joint prior density $p(M, R_{\mathrm{eq}}) \propto q(M, R_{\mathrm{eq}})$ because the constraint equation in compactness is dependent only on M and R_{eq}, and not on any ulterior source parameters. Such a procedure is also summarized in the preamble of Appendix B above.

A standard transform (in the context of the nested sampling software) on the other hand would take the form $\mathcal{H} \to \mathcal{S}$, $\boldsymbol{x} \mapsto \boldsymbol{\theta}$. To construct such a map, one may write $p(M, R_{\mathrm{eq}}) = p(R_{\mathrm{eq}} \mid M) p(M)$ where

$$p(M) = \int_{3r_g(M)}^{R_b} p(M, R_{\mathrm{eq}}) dR_{\mathrm{eq}} \propto R_b - 3r_g(M). \tag{14}$$

[93] Strictly speaking, we also impose several ulterior constraint equations. However, for the spin of PSR J0030+0451 these constraint equations are unimportant for defining the prior support on the joint space of M and R_{eq}.

Then define $\boldsymbol{x} := (x_M, x_R)$ where

$$x_M(M) = \int_{M_a}^{M} p(M')dM' \tag{15}$$

and

$$\begin{aligned} x_R(R_{\rm eq}; M) &= \int_{3r_g(M)}^{R_{\rm eq}} p(R'_{\rm eq} \mid M) dR'_{\rm eq} \\ &= \int_{3r_g(M)}^{R_{\rm eq}} \frac{p(M, R'_{\rm eq})}{p(M)} dR'_{\rm eq} = \frac{R_{\rm eq} - 3r_g(M)}{R_b - 3r_g(M)}. \end{aligned} \tag{16}$$

Inverting, one has

$$R_{\rm eq}(x_R; M) = x_R R_b + 3(1 - x_R) r_g(M), \tag{17}$$

and similarly one obtains a nonlinear function $M(x_M)$.

The problem with such a standard transformation is clear from inspection of $x_R(R_{\rm eq}; M)$. The common $(M, R_{\rm eq})$ degeneracy in pulse-profile modeling is linear due to sensitivity to compactness $M/R_{\rm eq}$. Therefore, requiring a constant compactness $r_g(M)/R_{\rm eq} = $ const. implies

$$x_R(R_{\rm eq}; M) = \frac{\text{const.} - 3}{R_b/r_g(M) - 3}, \tag{18}$$

where $M = M(x_M)$, meaning that $x_R = x_R(x_M; M/R_{\rm eq})$ is generally a nonlinear function. An optimal mapping would preserve the linearity of the degeneracy, and thus we do not opt for a standard transformation where $\mathcal{H} \to \mathcal{S}$, $\boldsymbol{x} \mapsto \boldsymbol{\theta}$; instead we inverse sample with rejection as described above.

B.2. Hot Regions

The joint prior distribution for the parameters of the members comprising both hot regions is non-trivial to implement, requiring a number of considerations. Moreover, the difficulty scales with the complexity of the hot regions. Here we break down the implementation into a series of steps.

B.2.1. Parameter Space

The joint parameter space for the members may be denoted $\boldsymbol{\nu} = (\Theta_p, \phi_p, \zeta_p, \Theta_s, \phi_s, \zeta_s, \ldots)$, where the six parameters explicitly written are inherent to every model wherein antipodal reflection symmetry is *not* imposed, and any ulterior parameters depend on the model. For `ST-U`, these six parameters are sufficient.

B.2.2. Region-exchange Degeneracy

If the two hot regions sharing the stellar surface are of equivalent complexity, the prior support for the coordinates of the regions can be defined so as to avoid degeneracy of the likelihood function under exchange of the region positions.

For regions related via antipodal reflection symmetry, the prior support for the colatitude of the regions can be defined such that only at (or near to) the support boundary are there configurations wherein a pair of regions mutually map onto one another via a rotation about the stellar rotation axis. The same condition applies if the regions are *not* related via antipodal reflection symmetry, but have equivalent complexities—e.g., `ST-U` meaning two `ST` regions. For instance, for `ST-U`, we can impose that $\Theta_p \leqslant \Theta_s$, where Θ_p and Θ_s are, respectively, the colatitudes of the centers of the primary and secondary `ST` regions.

If the regions are not related via antipodal reflection symmetry and do *not* have equivalent complexities—e.g., `ST+CST` or `ST+EST`—then by definition it is not true that an arbitrary point in parameter space yields exactly the same system configuration as distinct point in parameter space due to region exchange. Indeed, region exchange degeneracy may only be exist for a subset of parameter space, and thus one need not impose a joint constraint on the prior support for the colatitudes of the regions.

B.2.3. Ceding- and Superseding-member Radii

As highlighted in Section 2.5.3, it is advisable to define one's native nested-sampling space in order to linearize certain continuous degeneracies where possible.

The angular extent of the regions are remarkable in this respect for models constituted by at least one region at the `CST` complexity level or beyond—i.e., when a hot region is constituted by a superseding member and a ceding member, each with an angular radius. While it is not clear how to fully linearize degeneracy of type I illustrated in Figure 5 (see Section 2.5.2), the important[94] degeneracy of type IV can be linearized by working with the joint space of ψ and ζ, the radii of the superseding and ceding members, respectively (see Section 2.5.3). Although such a choice may be appear obvious in isolation, it may not be the space on which one chooses to intuitively define a joint prior density distribution.

Transforms for a CST or an `EST` *region*. If the superseding member subtends smaller angular extent than the ceding member, it is useful to consider $\psi := f\zeta$ where $f \in [\epsilon_f, 1]$ for small (or zero) ϵ_f. Moreover, it is common to invoke uniform prior density distributions for parameters with the intention of choosing a weakly informative prior but without rigorous proof. We therefore consider a flat separable density for f and ζ, such that $q(f, \zeta) = q(f)q(\zeta)$ where $f \sim U(\epsilon_f, 1)$ and $\zeta \sim U(\epsilon_\zeta, \pi/2 - \epsilon_\zeta)$ for small (or zero) ϵ_ζ. More generally, one might choose $\zeta \in [\epsilon_\zeta, b_\zeta]$; in our case $b_\zeta := \pi/2 - \epsilon_\zeta$. We then require the marginal density function $q(\psi)$ and the conditional density function $q(\zeta \mid \psi)$ in order to define a map $\mathcal{H} \to \mathcal{S}^\dagger$, where $\mathcal{H} = [0, 1] \times [0, 1]$ and where $\mathcal{S}^\dagger$ indicates that the support $\mathcal{S}$ of $p(\psi, \zeta)$, after all considerations in this appendix (see Appendix B.2.4), will be such that $\mathcal{S}^\dagger \supset \mathcal{S}$.

We must now consider the size of ϵ_f and ϵ_ζ: these limits determine the boundary of the prior support in the joint space of ψ and ζ. For a single-temperature hot region, we simply choose $\epsilon_f = 0$, such that either the superseding or ceding member can subtend zero angular extent at the boundary of the support; for a dual-temperature region one might choose a small finite value for ϵ_f given that it is filled with material of finite temperature. Given the choice $\epsilon_f = 0$ the choice of ϵ_ζ is unimportant for deriving the prior distributions of interest.

The joint density $q(\psi, \zeta)$ is given by

$$q(\psi, \zeta) = q(f, \zeta) \left| \frac{\partial f}{\partial \psi} \right| = \zeta^{-1} q(\psi/\zeta) q(\zeta). \tag{19}$$

[94] In terms of prior mass in comparison to that associated with other degenerate structures.

The marginal density function $q(\psi)$ is thus given by

$$q(\psi) = \int \zeta^{-1} q(\psi/\zeta) q(\zeta) d\zeta \propto \begin{cases} \int_{\epsilon_\zeta}^{b_\zeta} \zeta^{-1} d\zeta & \text{if } \psi \leqslant \epsilon_\zeta \\ \int_{\psi}^{b_\zeta} \zeta^{-1} d\zeta & \text{if } \epsilon_\zeta < \psi \leqslant b_\zeta; \end{cases} \tag{20}$$

and further that

$$q(\psi) \propto \begin{cases} \ln(b_\zeta/\epsilon_\zeta) & \text{if } \psi \leqslant \epsilon_\zeta \\ \ln(b_\zeta/\psi) & \text{if } \epsilon_\zeta < \psi \leqslant b_\zeta. \end{cases} \tag{21}$$

Now define $\boldsymbol{x} := (x_\psi, x_\zeta)$ where the mass

$$x_\psi(\psi) := \int_0^\psi q(\psi') d\psi' \propto \begin{cases} \psi \ln(b_\zeta/\epsilon_\zeta) & \text{if } \psi \leqslant \epsilon_\zeta \\ \psi - \epsilon_\zeta - \psi \ln(\psi/b_\zeta) & \text{if } \epsilon_\zeta < \psi \leqslant b_\zeta; \end{cases} \tag{22}$$

note that $x_\psi(\psi)$ is continuous at $\psi = \epsilon_\zeta$, and that $x_\psi \to 1^-$ as $\psi \to b_\zeta^-$ because $q(\psi/\zeta)q(\zeta) = (b_\zeta - \epsilon_\zeta)^{-1}$ where the joint density $q(f, \zeta)$ is finite. The function $\psi(x_\psi)$ is not obviously obtainable in closed form for $\psi > \epsilon_\zeta$, and thus we interpolate to perform the transformation $x_\psi \mapsto \psi$.

We now require the conditional density $q(\zeta \mid \psi)$:

$$q(\zeta \mid \psi) = \frac{q(\psi, \zeta)}{q(\psi)} = \begin{cases} \zeta^{-1}/\ln(b_\zeta/\epsilon_\zeta) & \text{if } (\psi \leqslant \epsilon_\zeta) \wedge (\epsilon_\zeta \leqslant \zeta \leqslant b_\zeta) \\ \zeta^{-1}/\ln(b_\zeta/\psi) & \text{if } (\epsilon_\zeta < \psi \leqslant b_\zeta) \wedge (\psi < \zeta \leqslant b_\zeta). \end{cases} \tag{23}$$

Then define the mass

$$x_\zeta(\zeta; \psi) := \int_{a(\psi)}^{\zeta} q(\zeta' \mid \psi) d\zeta' = \begin{cases} \ln(\zeta/\epsilon_\zeta)/\ln(b_\zeta/\epsilon_\zeta) & \text{if } (\psi \leqslant \epsilon_\zeta) \wedge (\epsilon_\zeta \leqslant \zeta \leqslant b_\zeta) \\ \ln(\zeta/\psi)/\ln(b_\zeta/\psi) & \text{if } (\epsilon_\zeta < \psi \leqslant b_\zeta) \wedge (\psi < \zeta \leqslant b_\zeta). \end{cases} \tag{24}$$

The function $\zeta(x_\zeta; \psi)$ is written in closed-form as

$$\zeta(\boldsymbol{x}) = \begin{cases} \epsilon_\zeta \exp(x_\zeta \ln(b_\zeta/\epsilon_\zeta)) & \text{if } (\psi \leqslant \epsilon_\zeta) \wedge (\epsilon_\zeta \leqslant \zeta \leqslant b_\zeta) \\ \psi(x_\psi) \exp(x_\zeta \ln(b_\zeta/\psi)) & \text{if } (\epsilon_\zeta < \psi \leqslant b_\zeta) \wedge (\psi < \zeta \leqslant b_\zeta). \end{cases} \tag{25}$$

Transforms for a PST region. If the superseding member can subtend a *larger* angular extent than the ceding member we consider the parameters $f \in [\epsilon_f, 2 - \epsilon_f]$ for small (or zero) ϵ_f and $\xi \in [\epsilon_\xi, b_\xi]$ for small ϵ_ξ. If $f \leqslant 1$, we define $\zeta = \xi$ and $\psi = f\zeta$, while if $f > 1$ we define $\psi = \xi$ and $\zeta = (2 - f)\psi$. We consider a flat separable joint density for f and ξ: $f \sim U(\epsilon_f, 2 - \epsilon_f)$ and $\xi \sim U(\epsilon_\xi, b_\xi)$. We again require the marginal density function $q(\psi)$ and the conditional density function $q(\zeta \mid \psi)$ in order to define a map $\mathcal{H} \to \mathcal{S}^\dagger$, where $\mathcal{H} = [0, 1] \times [0, 1]$ and where $\mathcal{S}^\dagger$ indicates that the support $\mathcal{S}$ of $p(\psi, \zeta)$, after all considerations in this appendix (see Appendix B.2.4), will be such that $\mathcal{S}^\dagger \supset \mathcal{S}$.

The joint density $q(\psi, \zeta)$ is piecewise with respect to f, given by

$$q(\psi, \zeta) = q(f, \xi) \left| \frac{\partial(f, \xi)}{\partial(\psi, \zeta)} \right| = \begin{cases} \zeta^{-1} q(f) q(\xi) & \text{if } f \leqslant 1 \\ \psi^{-1} q(f) q(\xi) & \text{if } f > 1. \end{cases} \tag{26}$$

We must now consider the size of ϵ_f and ϵ_ξ: these limits determine the boundary of the prior support in the joint space of ψ and ζ. For a PST region, we simply choose $\epsilon_f = 0$, such that either the superseding or ceding member can subtend zero angular extent at the boundary of the support; for a PDT region one might choose a small finite value for ϵ_f given that it is filled with material of finite temperature. Given the choice $\epsilon_f = 0$ the choice of ϵ_ξ is unimportant for deriving the prior distributions of interest.

In order to construct a map $\mathcal{H} \to \mathcal{S}^\dagger$, we aim to obtain the joint density $p(\psi, \zeta)$ in the conditional form $p(\psi, \zeta) = p(\zeta \mid \psi) p(\psi)$. The marginal density function $p(\psi)$ is given by

$$q(\psi) = \int_0^\psi \psi^{-1} q(f) q(\xi) d\zeta + \int_\psi^{b_\xi} \zeta^{-1} q(f) q(\xi) d\zeta. \tag{27}$$

If $\psi < \epsilon_\xi$, then: for $f > 1$, $q(\xi = \psi) = 0$; and for $f \leqslant 1$, $q(\xi = \zeta) = 0$ for $\zeta < \epsilon_\xi$. It follows that

$$q(\psi) \propto \begin{cases} \int_{\epsilon_\xi}^{b_\xi} \zeta^{-1} d\zeta & \text{if } \psi < \epsilon_\xi \\ \int_0^\psi \psi^{-1} d\zeta + \int_\psi^{b_\xi} \zeta^{-1} d\zeta & \text{if } \epsilon_\xi \leqslant \psi \leqslant b_\xi, \end{cases} \tag{28}$$

and further that

$$q(\psi) \propto \begin{cases} \ln(b_\xi/\epsilon_\xi) & \text{if } \psi < \epsilon_\xi \\ 1 + \ln(b_\xi/\psi) & \text{if } \epsilon_\xi \leqslant \psi \leqslant b_\xi. \end{cases} \tag{29}$$

Now define $\boldsymbol{x} := (x_\psi, x_\zeta)$ where the mass

$$x_\psi(\psi) := \int_0^\psi q(\psi') d\psi' \propto \begin{cases} \psi \ln(b_\xi/\epsilon_\xi) & \text{if } \psi < \epsilon_\xi \\ 2(\psi - \epsilon_\xi) - \psi \ln(\psi/b_\xi) & \text{if } \epsilon_\xi \leqslant \psi \leqslant b_\xi; \end{cases} \tag{30}$$

note that $x_\psi(\psi)$ is continuous at $\psi = \epsilon_\xi$, and that $x_\psi \to 1^-$ as $\psi \to b_\xi^-$ because $q(f)q(\xi) = (b_\xi - \epsilon_\xi)^{-1}/2$ where the joint density $q(f, \xi)$ is finite. The function $\psi(x_\psi)$ is not obviously obtainable in closed form for $\psi > \epsilon_\xi$, and thus we interpolate to perform the transformation $x_\psi \mapsto \psi$.

We now require the conditional density $q(\zeta \mid \psi)$:

$$q(\zeta \mid \psi) = \frac{q(\psi, \zeta)}{q(\psi)} = \begin{cases} \zeta^{-1}/\ln(b_\xi/\epsilon_\xi) & \text{if } (\psi < \epsilon_\xi) \wedge (\epsilon_\xi \leqslant \zeta \leqslant b_\xi) \\ \psi^{-1}/[1 + \ln(b_\xi/\psi)] & \text{if } (\epsilon_\xi \leqslant \psi \leqslant b_\xi) \wedge (0 \leqslant \zeta \leqslant \psi) \\ \zeta^{-1}/[1 + \ln(b_\xi/\psi)] & \text{if } (\epsilon_\xi \leqslant \psi \leqslant b_\xi) \wedge (\psi < \zeta \leqslant b_\xi). \end{cases} \tag{31}$$

Then define the mass

$$x_\zeta(\zeta;\,\psi) \coloneqq \int_{a(\psi)}^{\zeta} q(\zeta' \mid \psi)\,d\zeta' = \begin{cases} \ln(\zeta/\epsilon_\xi)/\ln(b_\xi/\epsilon_\xi) & \text{if } (\psi < \epsilon_\xi) \wedge (\epsilon_\xi \leqslant \zeta \leqslant b_\xi) \\ \dfrac{\zeta}{\psi}[1 + \ln(b_\xi/\psi)]^{-1} & \text{if } (\epsilon_\xi \leqslant \psi \leqslant b_\xi) \wedge (0 \leqslant \zeta \leqslant \psi) \\ [1 + \ln(\zeta/\psi)][1 + \ln(b_\xi/\psi)]^{-1} & \text{if } (\epsilon_\xi \leqslant \psi \leqslant b_\xi) \wedge (\psi < \zeta \leqslant b_\xi). \end{cases} \tag{32}$$

The function $\zeta(x_\zeta;\,\psi)$ is written in closed form as

$$\zeta(\boldsymbol{x}) = \begin{cases} \epsilon_\xi \exp(x_\zeta \ln(b_\xi/\epsilon_\xi)) & \text{if } (\psi < \epsilon_\xi) \wedge (\epsilon_\xi \leqslant \zeta \leqslant b_\xi) \\ x_\zeta \psi(x_\psi)[1 + \ln(b_\xi/\psi)] & \text{if } (\epsilon_\xi \leqslant \psi \leqslant b_\xi) \wedge (0 \leqslant \zeta \leqslant \psi) \\ \psi(x_\psi)\exp(x_\zeta[1 + \ln(b_\xi/\psi)] - 1) & \text{if } (\epsilon_\xi \leqslant \psi \leqslant b_\xi) \wedge (\psi < \zeta \leqslant b_\xi). \end{cases} \tag{33}$$

B.2.4. Non-overlapping Hot Regions

We implicitly define the support $\mathcal{S}$ of the joint density $p(\boldsymbol{v})$, where $\boldsymbol{v} = (\Theta_p, \phi_p, \zeta_p, \Theta_s, \phi_s, \zeta_s, \ldots)$ is the vector of parameters controlling both hot regions sharing the stellar surface, by imposing a constraint equation in terms of $\boldsymbol{v}$: we require that the regions are *non-overlapping*. More explicitly: two radiating regions associated with distinct regions cannot overlap—if they were to overlap, additional logical conditions would be required to specify an order of precedence for intensity evaluation at spacetime events at the stellar surface.

For single-temperature regions with a single member—i.e., simply connected and circular—one need only determine whether the ceding members, with their simple (circular) boundaries, overlap; the same condition is true if there exists a superseding member that is a *hole* in a ceding member—i.e., ST, CST, or EST. For dual-temperature hot regions, whose boundary is always constituted by a maximum of two simple (circular) boundaries, one need only determine whether any pair of members—from two distinct regions—overlap. However, for a PST region the boundary of the radiating region is more unwieldy for evaluating whether or not the radiating region of the PST region overlaps with another region sharing the stellar surface; in this case we simply define an overlap condition only in terms of the simple boundary of the ceding member, a subset of which is superseded by non-radiating surface.

To derive the joint prior density $p(\boldsymbol{v})$ and its support $\mathcal{S}$, we (i) define $q(\boldsymbol{v})$ as a product of density functions, (ii) inverse sample as $\mathcal{H} \rightarrow \mathcal{S}^\dagger$, $\boldsymbol{x} \mapsto \boldsymbol{v}$, and (iii) accept the sample $\tilde{\boldsymbol{v}}$ if the regions are determined to be non-overlapping. Algorithmically, we identify a set of pairs of members with simple boundaries whose center coordinates and angular radii are defined by vector $\tilde{\boldsymbol{v}}$, and evaluate via a spherical coordinate transformation whether or not the angular separation of the member centers is at least equal to the sum of the angular radii. If any of the pairs of members overlap, then $\tilde{\boldsymbol{v}} \notin \mathcal{S}$.

As an example, for ST+EST, $\boldsymbol{v} = (\Theta_p, \phi_p, \zeta_p, \Theta_s, \phi_s, \psi_s, \zeta_s, \varepsilon_s, \varphi_s)$ where the coordinates of the center of the ceding member of the EST region is derived from the vector $(\Theta_s, \phi_s, \psi_s, \zeta_s, \varepsilon_s, \varphi_s)$. Overlap is then evaluated for this ceding member in relation to the ST spot whose boundary is derived from the vector $(\Theta_p, \phi_p, \zeta_p)$. The marginal density function for *every* parameter constituting (the space of) vector $\boldsymbol{v}$ is thus modulated by excluding overlaps.

Appendix C
Model Summary Tables

In this appendix we provide posterior summary tables for all models applied to the PSR J0030+0451 event data, other than ST+PST (Table 2). For ST+EST see Table 3. For ST+CST see Table 4. For ST-U see Table 5. For ST-S see Table 6. For CDT-U see Table 7.

Table 3
Summary Table for ST+EST, Introduced in Section 2.5.7 and Illustrated in Figure 11

Parameter	Description	Prior (Density and Support)	$\widehat{\mathrm{CI}}_{68\%}$	$\widehat{D}_{\mathrm{KL}}$
P (ms)	Coordinate spin period	$P = 4.87$, fixed	⋯	⋯
M ($M_\odot$)	Gravitational mass	$M \sim U(1, 3)$	$1.46^{+0.17}_{-0.18}$	$1.15^{+0.02}_{-0.02}$
R_{eq} (km)	Coordinate equatorial radius	$R_{\mathrm{eq}} \sim U[3r_g(1), 16]$	$13.89^{+1.14}_{-1.30}$	$0.69^{+0.02}_{-0.02}$
	Compactness condition	$R_{\mathrm{eq}}/r_g(M) > 3$		
Θ_p (radians)	ST region center colatitude	$\Theta_p \sim U(0, \pi)$	$2.22^{+0.09}_{-0.10}$	$3.01^{+0.03}_{-0.03}$
ϕ_p (cycles)	ST region initial phase (from Earth)	$\phi_p \sim U(a, a + 0.2)$[a]	$0.45^{+0.00}_{-0.00}$	$6.59^{+0.02}_{-0.03}$
ζ_p (radians)	ST region angular radius	$\zeta_p \sim U(0,\pi/2)$	$0.07^{+0.01}_{-0.01}$	$4.78^{+0.03}_{-0.02}$
Θ_s (radians)	EST region hole center colatitude	$\Theta_s \sim U(0, \pi)$	$2.66^{+0.07}_{-0.09}$	$3.36^{+0.02}_{-0.03}$
ϕ_s (cycles)	EST region initial phase (from Earth antipode)	$\phi_s \sim U(-0.5, 0.5)$, periodic	$-0.51^{+0.01}_{-0.01}$	$4.28^{+0.03}_{-0.02}$
ψ_s^+ (radians)	EST region angular radii difference	$\zeta_s \sim U(0, \pi/2)$, $\psi_s^+ := \zeta_s - \psi_s$	$0.03^{+0.01}_{-0.01}$	$2.93^{+0.03}_{-0.03}$
ψ_s (radians)	EST region hole angular radius	$f_s \sim U(0, 1)$, $\psi_s := f_s\zeta_s$	$0.25^{+0.04}_{-0.04}$	$2.29^{+0.02}_{-0.03}$
ε_s	EST region annulus fractional angular offset	$\varepsilon_s \sim U(0, 1)$[b]	$0.45^{+0.45}_{-0.33}$	$0.08^{+0.01}_{-0.01}$
φ_s (radians)	EST region annulus azimuthal offset	$\varphi_s \sim U(0, 2\pi)$	$0.79^{+0.35}_{-0.26}$	$0.64^{+0.06}_{-0.06}$
	Non-overlapping hot-region annuli	function of $(\Theta_p, \Theta_s, \phi_p, \phi_s, \zeta_p, \zeta_s)$		
$\log_{10}(\mathcal{T}_p\,(\mathrm{K}))$	ST region NSX effective temperature	$\log_{10}\mathcal{T}_p \sim U(5.1, 6.8)$, NSX limits	$6.11^{+0.01}_{-0.01}$	$5.03^{+0.03}_{-0.03}$
$\log_{10}(\mathcal{T}_s\,(\mathrm{K}))$	EST region NSX effective temperature	$\log_{10}\mathcal{T}_s \sim U(5.1, 6.8)$, NSX limits	$6.11^{+0.01}_{-0.01}$	$5.60^{+0.03}_{-0.03}$
i (radians)	Earth inclination to rotation axis	$i \sim U(0, \pi/2)$	$1.01^{+0.07}_{-0.07}$	$2.40^{+0.03}_{-0.02}$
D (kpc)	Earth distance	$D \sim N(0.325, 0.009)$	$0.33^{+0.01}_{-0.01}$	$0.02^{+0.01}_{-0.01}$
N_{H} ($10^{20}\mathrm{cm}^{-2}$)	Interstellar neutral H column density	$N_{\mathrm{H}} \sim U(0, 5)$	$0.61^{+0.18}_{-0.16}$	$2.84^{+0.03}_{-0.03}$
NICER α	Calibrated matrix scaling	$\alpha \sim N(1, 0.1)$, $\alpha \in [0.5, 1.5]$	$0.99^{+0.09}_{-0.09}$	$0.02^{+0.01}_{-0.00}$
NICER β	Reference-to-calibrated matrix weighting	$\beta \sim U(0, 1)$	$0.16^{+0.17}_{-0.11}$	$1.07^{+0.03}_{-0.03}$
NICER γ	Reference matrix scaling	$\gamma \sim N(1, 0.1)$, $\gamma \in [0.5, 1.5]$	$0.99^{+0.09}_{-0.09}$	$0.04^{+0.01}_{-0.01}$
	Sampling process information			
	Number of free parameters: 19 Number of runs: 1 Number of live points: 10^3 Inverse hypervolume expansion factor: 0.3 Termination condition: 10^{-1} Evidence:[c] $\widehat{\ln \mathcal{Z}} = -36367.81^{+0.48}_{-0.43}$ Global KL divergence: $\widehat{D}_{\mathrm{KL}} = 62.1^{+0.8}_{-0.8}$ bits Number of core[d] hours: 61210 Likelihood evaluations: 88965106 Nested replacements: 53149 Weighted posterior samples: 17671			

Notes.
[a] Where $\phi_p = a$ is an arbitrary phase dependent on event data pre-processing. We set $a = 0.35$.
[b] If we were to parameterize the eccentricity in terms of the *sum* of angular radii of the superseding (hole) and ceding regions, $\vartheta := \varepsilon(1 + f)\zeta$, a *conditional* prior such as $\varepsilon \mid f \sim U(0, (1 - f)/(1 + f))$ would be necessary, where the upper bound imposes that the radiating region is not simply connected (i.e., is an annulus): $\varepsilon(1 + f)\zeta + f\zeta \leqslant \zeta \Longrightarrow \varepsilon \leqslant (1 - f)/(1 + f)$.
[c] Defined as the prior predictive probability $p(\boldsymbol{d} \mid \mathrm{ST} + \mathrm{EST})$. We report the interval about the median containing $\pm 45\%$ of 10^3 joint bootstrap-weight replications for the combined run.
[d] Intel® Xeon E5-2697Av4 (2.60 GHz; Broadwell) processors on the SURFsara Cartesius supercomputer.

Table 4
Summary Table for ST+CST, Introduced in Section 2.5.7

Parameter	Description	Prior (Density and Support)	$\widehat{\mathrm{CI}}_{68\%}$	$\widehat{D}_{\mathrm{KL}}$
P (ms)	Coordinate spin period	$P = 4.87$, fixed	⋯	⋯
M ($M_\odot$)	Gravitational mass	$M \sim U(1, 3)$	$1.44^{+0.18}_{-0.19}$	$1.07^{+0.02}_{-0.02}$
R_{eq} (km)	Coordinate equatorial radius	$R_{\mathrm{eq}} \sim U[3r_g(1), 16]$	$13.88^{+1.23}_{-1.38}$	$0.61^{+0.01}_{-0.01}$
	Compactness condition	$R_{\mathrm{eq}}/r_g(M) > 3$		
Θ_p (radians)	ST region center colatitude	$\Theta_p \sim U(0, \pi)$	$2.24^{+0.09}_{-0.09}$	$5.14^{+0.03}_{-0.02}$
ϕ_p (cycles)	ST region initial phase (from Earth)	$\phi_p \sim U(-0.5, 0.5)$, periodic	$0.46^{+0.00}_{-0.00}$	$6.66^{+0.03}_{-0.02}$
ζ_p (radians)	ST region angular radius	$\zeta_p \sim U(0, \pi/2)$	$0.07^{+0.01}_{-0.01}$	$4.59^{+0.02}_{-0.02}$
Θ_s (radians)	CST region center colatitude	$\Theta_s \sim U(0, \pi)$	$2.60^{+0.05}_{-0.06}$	$2.79^{+0.02}_{-0.02}$
ϕ_s (cycles)	CST region initial phase (from Earth antipode)	$\phi_s \sim U(-0.5, 0.5)$, periodic	$-0.50^{+0.00}_{-0.00}$	$7.33^{+0.02}_{-0.03}$
ψ_s^+ (radians)	CST region annulus angular width	$\zeta_s \sim U(0, \pi/2)$, $\psi_s^+ := \zeta_s - \psi_s$	$0.04^{+0.01}_{-0.01}$	$2.71^{+0.02}_{-0.02}$
ψ_s (radians)	CST region hole angular radius	$f_s \sim U(0, 1)$, $\psi_s := f_s \zeta_s$	$0.23^{+0.03}_{-0.03}$	$2.52^{+0.02}_{-0.02}$
	Enforce ST and CST colatitude order[a]	$\Theta_s \geqslant \Theta_p$		
	Non-overlapping hot regions	function of $(\Theta_p, \Theta_s, \phi_p, \phi_s, \zeta_p, \zeta_s)$		
$\log_{10}(\mathcal{T}_p\,(\mathrm{K}))$	ST region NSX effective temperature	$\log_{10}\mathcal{T}_p \sim U(5.1, 6.8)$, NSX limits	$6.10^{+0.01}_{-0.01}$	$4.99^{+0.02}_{-0.02}$
$\log_{10}(\mathcal{T}_s\,(\mathrm{K}))$	CST region NSX effective temperature	$\log_{10}\mathcal{T}_s \sim U(5.1, 6.8)$, NSX limits	$6.11^{+0.01}_{-0.01}$	$5.58^{+0.02}_{-0.02}$
i (radians)	Earth inclination to rotation axis	$i \sim U(0, \pi/2)$	$1.02^{+0.07}_{-0.08}$	$2.32^{+0.02}_{-0.02}$
D (kpc)	Earth distance	$D \sim N(0.325, 0.009)$	$0.33^{+0.01}_{-0.01}$	$0.01^{+0.00}_{-0.00}$
N_{H} ($10^{20}\mathrm{cm}^{-2}$)	Interstellar neutral H column density	$N_{\mathrm{H}} \sim U(0, 5)$	$0.62^{+0.19}_{-0.18}$	$2.75^{+0.02}_{-0.02}$
NICER α	Calibrated matrix scaling	$\alpha \sim N(1, 0.1)$, $\alpha \in [0.5, 1.5]$	$0.99^{+0.10}_{-0.10}$	$0.01^{+0.00}_{-0.00}$
NICER β	Reference-to-calibrated matrix weighting	$\beta \sim U(0, 1)$	$0.16^{+0.18}_{-0.11}$	$0.98^{+0.02}_{-0.02}$
NICER γ	Reference matrix scaling	$\gamma \sim N(1, 0.1)$, $\gamma \in [0.5, 1.5]$	$0.99^{+0.09}_{-0.09}$	$0.01^{+0.00}_{-0.00}$
	Sampling process information			
	Number of free parameters: 17			
	Number of runs: 2			
	Number of live points *per run*: 10^3			
	Inverse hypervolume expansion factor: 0.3			
	Termination condition: 10^{-1}			
	Evidence:[b] $\widehat{\ln \mathcal{Z}} = -36368.00^{+0.34}_{-0.33}$			
	Global KL divergence: $\widehat{D}_{\mathrm{KL}} = 62.9^{+0.6}_{-0.5}$ bits			
	Combined number of core[c] hours: 23010			
	Combined likelihood evaluations: 39501475			
	Combined nested replacements: 105264			
	Combined weighted posterior samples: 32839			

Notes.
[a] Based on learning that additional complexity (of the form we consider in our model space) beyond ST is not warranted for the one region, but is warranted for the other region.
[b] Defined as the prior predictive probability $p(\boldsymbol{d} \mid \mathtt{ST} + \mathtt{CST})$. We report the interval about the median containing ±45% of 10^3 joint bootstrap-weight replications for the combined run.
[c] Approximate equal-partition between Intel® Xeon E5-2697Av4 (2.60 GHz; Broadwell) and E5-2690v3 (2.60 GHz; Haswell) processors on the SURFsara Cartesius supercomputer.

Table 5
Summary Table for ST-U, Introduced in Section 2.5.1 and Illustrated in Figure 3

Parameter	Description	Prior (Density and Support)	$\widehat{\mathrm{CI}}_{68\%}$	$\widehat{D}_{\mathrm{KL}}$
P (ms)	Coordinate spin period	$P = 4.87$, fixed	⋯	⋯
M ($M_\odot$)	Gravitational mass	$M \sim U(1, 3)$	$1.09^{+0.11}_{-0.07}$	$2.09^{+0.02}_{-0.03}$
R_{eq} (km)	Coordinate equatorial radius	$R_{\mathrm{eq}} \sim U[3r_g(1), 16]$	$10.44^{+1.10}_{-0.86}$	$1.44^{+0.03}_{-0.03}$
	Compactness condition	$R_{\mathrm{eq}}/r_g(M) > 3$		
Θ_p (radians)	p region center colatitude	$\Theta_p \sim U(0, \pi)$	$2.48^{+0.06}_{-0.06}$	$6.92^{+0.02}_{-0.02}$
Θ_s (radians)	s region center colatitude	$\Theta_s \sim U(0, \pi)$	$2.78^{+0.02}_{-0.02}$	$4.08^{+0.02}_{-0.02}$
ϕ_p (cycles)	p region initial phase (from Earth)	$\phi_p \sim U(-0.5, 0.5)$, periodic	$0.46^{+0.00}_{-0.00}$	$7.51^{+0.02}_{-0.02}$
ϕ_s (cycles)	s region initial phase (from Earth antipode)	$\phi_s \sim U(-0.5, 0.5)$, periodic	$-0.50^{+0.00}_{-0.00}$	$8.05^{+0.02}_{-0.02}$
ζ_p (radians)	p region angular radius	$\zeta_p \sim U(0, \pi/2)$	$0.14^{+0.02}_{-0.02}$	$3.95^{+0.02}_{-0.02}$
ζ_s (radians)	s region angular radius	$\zeta_s \sim U(0, \pi/2)$	$0.29^{+0.04}_{-0.03}$	$2.98^{+0.02}_{-0.02}$
	Eliminate region-exchange degeneracy Non-overlapping hot regions	$\Theta_s \geqslant \Theta_p$ function of $(\Theta_p, \Theta_s, \phi_p, \phi_s, \zeta_p, \zeta_s)$		
$\log_{10}(\mathcal{T}_p\,(\mathrm{K}))$	p region NSX effective temperature	$\log_{10}\mathcal{T}_p \sim U(5.1, 6.8)$, NSX limits	$6.11^{+0.01}_{-0.01}$	$5.53^{+0.03}_{-0.02}$
$\log_{10}(\mathcal{T}_s\,(\mathrm{K}))$	s region NSX effective temperature	$\log_{10}\mathcal{T}_s \sim U(5.1, 6.8)$, NSX limits	$6.10^{+0.01}_{-0.01}$	$5.71^{+0.02}_{-0.02}$
i(radians)	Earth inclination to rotation axis	$i \sim U(0, \pi/2)$	$1.04^{+0.07}_{-0.08}$	$2.38^{+0.02}_{-0.02}$
D (kpc)	Earth distance	$D \sim N(0.325, 0.009)$	$0.33^{+0.01}_{-0.01}$	$0.35^{+0.02}_{-0.02}$
N_{H} ($10^{20}\mathrm{cm}^{-2}$)	Interstellar neutral H column density	$N_{\mathrm{H}} \sim U(0, 5)$	$1.23^{+0.17}_{-0.17}$	$2.82^{+0.02}_{-0.02}$
NICER α	Calibrated matrix scaling	$\alpha \sim N(1, 0.1)$, $\alpha \in [0.5, 1.5]$	$0.96^{+0.10}_{-0.10}$	$0.11^{+0.01}_{-0.01}$
NICER β	Reference-to-calibrated matrix weighting	$\beta \sim U(0, 1)$	$0.23^{+0.23}_{-0.16}$	$0.57^{+0.02}_{-0.02}$
NICER γ	Reference matrix scaling	$\gamma \sim N(1, 0.1)$, $\gamma \in [0.5, 1.5]$	$0.91^{+0.10}_{-0.09}$	$0.56^{+0.04}_{-0.03}$
	Sampling process information			
	Number of free parameters: 16 Number of runs:[a] 2 Number of live points *per run*: 10^3 Inverse hypervolume expansion factor: 0.3 Termination condition: 10^{-3} Evidence:[b] $\widehat{\ln \mathcal{Z}} = -36377.60^{+0.36}_{-0.35}$ Global KL divergence: $\widehat{D}_{\mathrm{KL}} = 63.7^{+0.5}_{-0.5}$ bits Combined number of core[c] hours: 9588 Combined likelihood evaluations: 25346841 Combined nested replacements: 121617 Combined weighted posterior samples: 49481			

Notes.
[a] The mode-separation MultiNest variant was deactivated for these two runs that were combined to compute estimators. Mode separation means that modes are not evolved independently and nested sampling threads contact multiple modes; a mode-separation run was executed and is displayed in Figure 19 and in the figure sets available in the online corresponding to Figures 20 and 21. The theory nor software implementation exists for combining this run with the two reported in the table.
[b] Defined as the prior predictive probability $p(\boldsymbol{d} \mid \mathtt{ST-U})$. We report the interval about the median containing ±45% of 10^3 joint bootstrap-weight replications for the combined run. See the footnote in Table 2.
[c] Approximate equal-partition between Intel® Xeon E5-2697Av4 (2.60 GHz; Broadwell) and E5-2690v3 (2.60 GHz; Haswell) processors on the SURFsara Cartesius supercomputer.

Table 6
Summary Table for ST-S, Introduced in Section 2.5.1 and Illustrated in Figure 3

Parameter	Description	Prior (Density and Support)	$\widehat{\mathrm{CI}}_{68\%}$	$\widehat{D}_{\mathrm{KL}}$
P (ms)	Coordinate spin period	$P = 4.87$, fixed	⋯	⋯
M ($M_\odot$)	Gravitational mass	$M \sim U(1, 3)$	$2.93^{+0.01}_{-0.01}$	$6.75^{+0.03}_{-0.03}$
R_{eq} (km)	Coordinate equatorial radius	$R_{\mathrm{eq}} \sim U[3r_g(1), 16]$	$15.97^{+0.02}_{-0.04}$	$6.11^{+0.04}_{-0.04}$
	Compactness condition	$R_{\mathrm{eq}}/r_g(M) > 3$		
Θ_p (radians)	p region center colatitude	$\Theta_p \sim U(0, \pi/2)$	$1.26^{+0.02}_{-0.02}$	$4.07^{+0.03}_{-0.02}$
ϕ_p (cycles)	p region initial phase (from Earth)	$\phi_p \sim U(-0.5, 0.5)$, periodic	$-0.09^{+0.00}_{-0.00}$	$8.57^{+0.02}_{-0.03}$
ζ (radians)	ST region angular radius (shared)	$\zeta \sim U(0, \pi/2)$	$0.09^{+0.01}_{-0.00}$	$6.55^{+0.02}_{-0.02}$
$\log_{10}(\mathcal{T}(\mathrm{K}))$	ST region NSX effective temperature (shared)	$\log_{10}\mathcal{T} \sim U(5.1, 6.8)$, NSX limits	$6.08^{+0.00}_{-0.00}$	$6.97^{+0.02}_{-0.02}$
Θ_s (radians)	s region center colatitude	$\Theta_s = \pi - \Theta_p$, derived		
ϕ_s (cycles)	s region initial phase	$\phi_s = \phi_p + 0.5$, derived		
i (radians)	Earth inclination to rotation axis	$i \sim U(0, \pi/2)$	$1.23^{+0.02}_{-0.03}$	$3.98^{+0.03}_{-0.02}$
D (kpc)	Earth distance	$D \sim N(0.325, 0.009)$	$0.32^{+0.01}_{-0.01}$	$0.02^{+0.00}_{-0.00}$
N_{H} ($10^{20}\mathrm{cm}^{-2}$)	Interstellar neutral H column density	$N_{\mathrm{H}} \sim U(0, 5)$	$0.02^{+0.03}_{-0.02}$	$5.93^{+0.04}_{-0.04}$
NICER α	Calibrated matrix scaling	$\alpha \sim N(1, 0.1)$, $\alpha \in [0.5, 1.5]$	$1.00^{+0.09}_{-0.09}$	$0.01^{+0.00}_{-0.00}$
NICER β	Reference-to-calibrated matrix weighting	$\beta \sim U(0, 1)$	$0.36^{+0.22}_{-0.19}$	$0.35^{+0.02}_{-0.02}$
NICER γ	Reference matrix scaling	$\gamma \sim N(1, 0.1)$, $\gamma \in [0.5, 1.5]$	$1.01^{+0.09}_{-0.09}$	$0.03^{+0.00}_{-0.00}$
	Sampling process information			
	Number of free parameters: 12			
	Number of runs: 2			
	Number of live points *per run*: 10^3			
	Inverse hypervolume expansion factor: 0.3			
	Termination condition: 10^{-3}			
	Evidence:[a] $\widehat{\ln \mathcal{Z}} = -37211.71^{+0.29}_{-0.31}$			
	Global KL divergence: $\widehat{D}_{\mathrm{KL}} = 51.3^{+0.5}_{-0.5}$ bits			
	Combined number of core[b] hours: 1494			
	Combined likelihood evaluations: 1666483			
	Combined nested replacements: 87488			
	Combined weighted posterior samples: 30202			

Notes.
[a] Defined as the prior predictive probability $p(\boldsymbol{d} \mid \mathtt{ST-S})$. We report the interval about the median containing ±45% of 10^3 joint bootstrap-weight replications for the combined run. See the footnote in Table 2.
[b] Intel® Xeon E5-2697Av4 (2.60 GHz; Broadwell) processors on the SURFsara Cartesius supercomputer.

Table 7
Summary Table for CDT-U, Introduced in Section 2.5.3 and Illustrated in Figure 6

Parameter	Description	Prior (Density and Support)	$\widehat{\mathrm{CI}}_{68\%}$	$\widehat{D}_{\mathrm{KL}}$
P (ms)	Coordinate spin period	$P = 4.87$,fixed	⋯	⋯
M ($M_\odot$)	Gravitational mass	$M \sim U(1, 3)$	$1.44^{+0.17}_{-0.18}$	1.15
R_{eq} (km)	Coordinate equatorial radius	$R_{\mathrm{eq}} \sim U[3r_g(1), 16]$	$13.86^{+1.16}_{-1.26}$	0.68
	Compactness condition	$R_{\mathrm{eq}}/r_g(M) > 3$		
Θ_p (radians)	p region center colatitude	$\Theta_p \sim U(0, \pi)$	$2.24^{+0.08}_{-0.08}$	4.68
ϕ_p (cycles)	p region initial phase (from Earth)	$\phi_p \sim U(-0.5, 0.5)$, periodic	$0.46^{+0.00}_{-0.00}$	6.75
ψ_p^+ (radians)	p region annulus angular width	$\zeta_p \sim U(0, \pi/2)$, $\psi_p^+ := \zeta_p - \psi_p$	$0.08^{+0.16}_{-0.06}$	2.59
ψ_p (radians)	p region hole angu- lar radius	$f_p \sim U(0, 1)$, $\psi_p := f_p\,\zeta_p$	$0.07^{+0.01}_{-0.01}$	5.28
Θ_s (radians)	s region center colatitude	$\Theta_s \sim U(0, \pi)$	$2.61^{+0.05}_{-0.06}$	2.98
ϕ_s (cycles)	s region initial phase (from Earth antipode)	$\phi_s \sim U(-0.5, 0.5)$, periodic	$-0.50^{+0.00}_{-0.00}$	7.38
ψ_s^+ (radians)	s region annulus angular width	$\zeta_s \sim U(0, \pi/2)$, $\psi_s^+ := \zeta_s - \psi_s$	$0.04^{+0.01}_{-0.01}$	5.23
ψ_s (radians)	s region hole angular radius	$f_s \sim U(0, 1)$, $\psi_s := f_s\,\zeta_s$	$0.23^{+0.03}_{-0.03}$	3.66
	Eliminate region-exchange degeneracy Non-overlapping hot-region annuli	$\Theta_s \geqslant \Theta_p$ function of $(\Theta_p,\Theta_s,\phi_p,\phi_s,\zeta_p,\zeta_s)$		
$\log_{10}(\mathcal{T}_p$ (K))	p region annulus NSX effective temperature	$\log_{10}\mathcal{T}_p \sim U(5.1, 6.8)$, NSX limits	$5.43^{+0.23}_{-0.20}$	1.11
$\log_{10}(T_p$ (K))	p region hole NSX effective temperature	$\log_{10}T_p \sim U(5.1, 6.8)$, NSX limits	$6.11^{+0.01}_{-0.01}$	5.06
$\log_{10}(\mathcal{T}_s$ (K))	s region annulus NSX effective temperature	$\log_{10}\mathcal{T}_s \sim U(5.1, 6.8)$, NSX limits	$6.11^{+0.01}_{-0.01}$	5.56
$\log_{10}(T_s$ (K))	s region hole NSX effective temperature	$\log_{10}T_s \sim U(5.1, 6.8)$, NSX limits	$5.47^{+0.19}_{-0.23}$	1.30
i (radians)	Earth inclination to rotation axis	$i \sim U(0, \pi/2)$	$1.02^{+0.07}_{-0.07}$	2.37
D (kpc)	Earth distance	$D \sim N(0.325, 0.009)$	$0.33^{+0.01}_{-0.01}$	0.02
N_{H} (10^{20}cm^{-2})	Interstellar neutral H column density	$N_{\mathrm{H}} \sim U(0, 5)$	$0.70^{+0.19}_{-0.18}$	2.70
NICER α	Calibrated matrix scaling	$\alpha \sim N(1, 0.1)$, $\alpha \in [0.5, 1.5]$	$0.99^{+0.09}_{-0.09}$	0.02
NICER β	Reference-to-calibrated matrix weighting	$\beta \sim U(0, 1)$	$0.16^{+0.17}_{-0.11}$	1.05
NICER γ	Reference matrix scaling	$\gamma \sim N(1, 0.1)$, $\gamma \in [0.5, 1.5]$	$0.99^{+0.09}_{-0.09}$	0.03
	Sampling process information			
	Number of free parameters: 20 Number of runs: 1 Number of live points:[a] 10^3 Inverse hypervolume expansion factor:[b] 0.3 Termination condition: 10^{-1} Evidence:[c] $\widehat{\ln \mathcal{Z}} = -36366.76 \pm 0.21$ Global KL divergence: $\widehat{D}_{\mathrm{KL}} = 64.0$ bits Combined number of core[d] hours: 101917 Likelihood evaluations: 156707329 Nested replacements: 54610 Weighted posterior samples: 17503			

Notes. In this table we only give the numerical details for one completed run. However, we attempted to perform a higher-efficiency higher-resolution second run, without activation of the mode-separation MultiNest sampling algorithm. Our attempt to improve the sampling efficiency by linearizing the degeneracy (type IV in Figures 5 and 7) observed in the first run; however, we failed to fully apply the necessary transformations (later applied to models ST+CST and beyond, as described in Section 2.5.3 and Appendix B.2), and the mapping from the native sampling space to the physical parameter space inadvertently preserved the nonlinearity of the degeneracy. We thus did not attain higher efficiency, which coupled with higher-resolution calculation, meant that this run was nearing—but did not reach—termination according to the standard criterion used for the other sampling processes reported in this work. Due to the low sampling efficiencies being reported, we ceased computation after consumption of ∼160540 core hours in order to preserve resources and redesign our modeling route as described in Sections 2.5.4 and 2.5.7. The highest-likelihood points from the posterior mode in this second run were however utilized, in combination with those from the first run, to map out all of the degenerate posterior structure in Figures 7 and 8. The inefficiency suffered during CDT-U posterior computation served as a stark reminder that we should design our problems as carefully as possible in order to avoid resource wastage, and was the motivation behind considering in detail the question of how much hot-region complexity is helpful. While we could in principle conclude this adjourned run with additional computing resources, we have argued in this work that would not be fruitful do so.

[a] The mode-separation MultiNest variant was activated, meaning that isolated modes are evolved independently and nested sampling threads migrate between multiple modes. A local posterior mode was identified, corresponding to a weaker phase solution in which the primary and secondary hot regions transpose—relative to the global posterior mode—in their coupling to pulse components visible in the phase-folded event data. The number of live points locked into the mode with dominant posterior mass was 637, a number assigned according to the prior mass distribution upon mode separation and under the influence of Monte Carlo noise. The posterior mass ratio (or ratio of local evidences) is estimated to be ∼1100. As stated in Appendix A.2.3, because of activation of mode-separation sampling, we cannot perform error analysis via process bootstrapping, and we would not be able to combine with another run, supposing that one was available. We nevertheless have an error on the log-evidence reported natively by MultiNest. Another consequence of the activation of mode-separation is that sampling resolution was absorbed by the local posterior mode with much lower mass—in other words, the active points were sparser in the dominant mode. In combination with sampling error due to likelihood isosurface nonlinearity, the consequence was that the dominant and degenerate posterior mode—which forms a large connected structure in parameter space as discussed in Section 2.5.4—was not fully resolved. In particular, the type I degeneracy branch (refer to Figures 5 and 7) was not fully resolved, with sampling threads (active points) migrating to and densely populating the type IV branch (again refer to Figure 7) in an unbalanced manner. The second, albeit incomplete, run described in the caption of this table exhibited much improved resolution of the dominant mode (see the points in Figure 7); the resolution remained incomplete, however, which we attribute to sampling error due to the clear nonlinear degeneracy present in the mode.

[b] We decreased the expansion factor to 0.8^{-1} when the acceptance rate decayed to below 5×10^{-4}, which slightly decreased the rate of decay. At this point the process was sampling from the typical set and nearing termination (2×10^4 core hours remaining at low acceptance fraction).

[c] Defined as the prior predictive probability $p(\boldsymbol{d} \mid \mathtt{CDT-U})$.

[d] Intel® Xeon E5-2697Av4 (2.60 GHz; Broadwell) processors on the SURFsara Cartesius supercomputer.

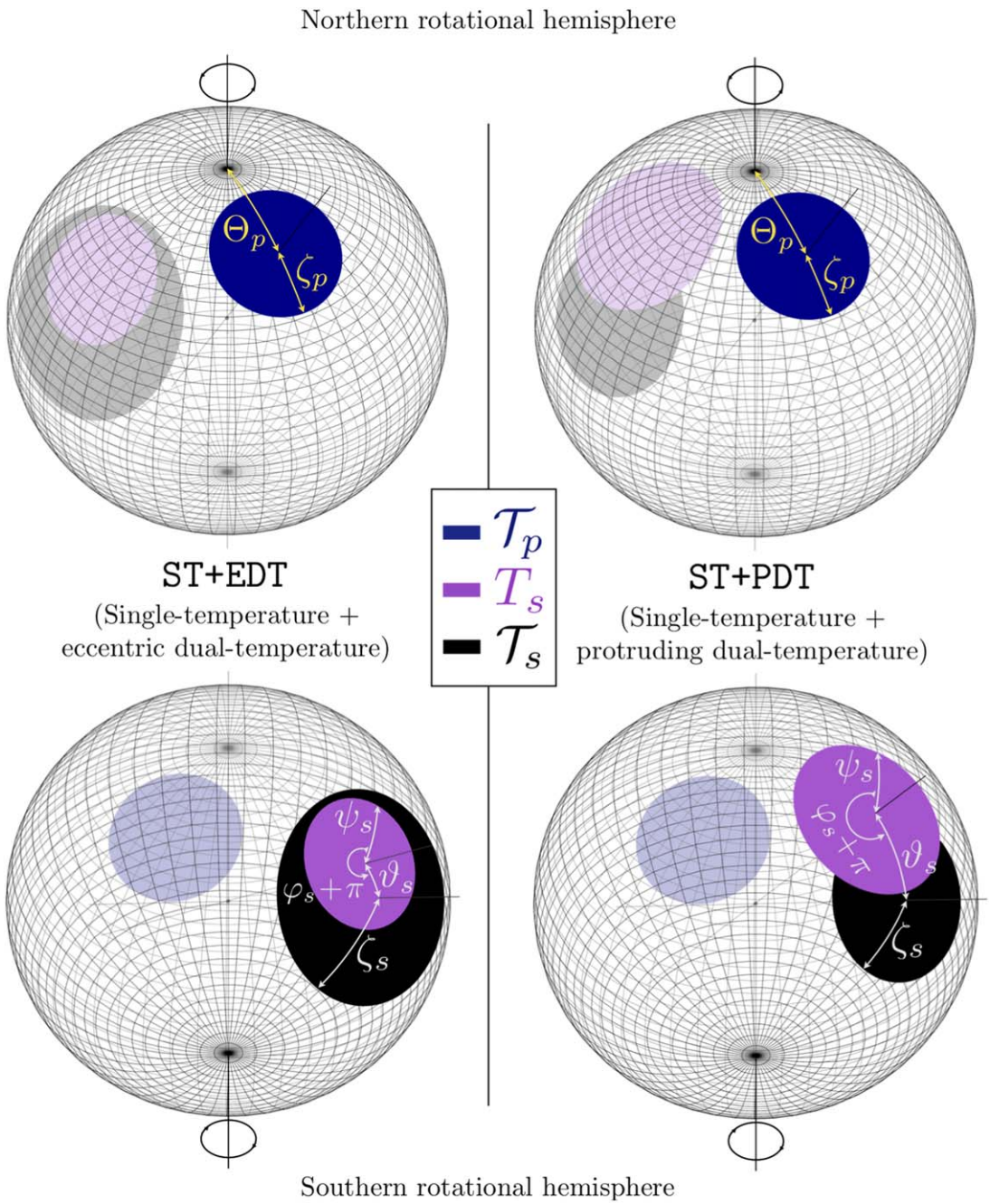

Figure 24. Schematic diagrams of models wherein an ST region shares the stellar surface with a higher-complexity EDT or PDT region. The complete figure set (3 images) is available in the online journal.

(The complete figure set (3 images) is available.)

Appendix D
Supplementary Ideas for Model Extension

We now make note of models within the scope of this work that were either clearly (without need for explicit posterior computation due to the posterior properties of simpler models) not competitive for PSR J0030+0451 or unhelpfully complex in phenomenologically describing the structure of a hot region. Excess complexity does not mean that the modeling has been optimized, but indicates that a particular extension to a model is not warranted because we are insensitive to a subset of parameters (or combinations of parameters), *and* that the signals that maximize the likelihood function are signals that are effectively generated by a simpler (nested) model.

ST+EDT and *ST+PDT*. Obtained via simple extension of ST+EST and ST+PST: let the superseding member (the hole for EST) contain radiating material. Refer to Figure 24.

EDT-S and EDT-U. Refer to the online figure set associated with Figure 24. For EDT-S, the surface radiation field associated with the secondary region is derived exactly by applying antipodal symmetry to the primary region: there are no free parameters associated with the secondary region. Conversely, for EDT-U, the secondary region is endowed with distinct parameters—i.e., it is *not* derived from the primary

region under antipodal symmetry. However, the parameters of the secondary region have an otherwise equivalent meaning—in terms of surface radiation field specification—to their primary-region counterparts. As an example, the azimuth of the center of the secondary ceding member is defined relative to the meridian passing through the center of the secondary hole and through the rotational poles.

*PDT-S*and PDT-U. Refer to the online figure set associated with Figure 24. For PDT-S, the surface radiation field associated with the secondary region is derived exactly by applying antipodal symmetry to the primary region: there are no free parameters associated with the secondary region. Conversely, for PDT-U, the secondary region is endowed with distinct parameters—i.e., it is *not* derived from the primary region under antipodal symmetry. However, the parameters of the secondary region have an otherwise equivalent meaning—in terms of surface radiation field specification—to their primary-region counterparts.

ORCID iDs

T. E. Riley https://orcid.org/0000-0001-9313-0493
A. L. Watts https://orcid.org/0000-0002-1009-2354
S. Bogdanov https://orcid.org/0000-0002-9870-2742
P. S. Ray https://orcid.org/0000-0002-5297-5278
R. M. Ludlam https://orcid.org/0000-0002-8961-939X
S. Guillot https://orcid.org/0000-0002-6449-106X
A. V. Bilous https://orcid.org/0000-0002-7177-6987
D. Chakrabarty https://orcid.org/0000-0001-8804-8946
A. K. Harding https://orcid.org/0000-0001-6119-859X
W. C. G. Ho https://orcid.org/0000-0002-6089-6836
S. M. Morsink https://orcid.org/0000-0003-4357-0575
T. E. Strohmayer https://orcid.org/0000-0001-7681-5845

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, PhRvL, 119, 161101
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2018, PhRvL, 121, 161101
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019, PhRvX, 9, 011001
Alford, M., & Sedrakian, A. 2017, PhRvL, 119, 161104
Alford, M. G., & Han, S. 2016, EPJA, 52, 62
AlGendy, M., & Morsink, S. M. 2014, ApJ, 791, 78
Annala, E., Gorda, T., Kurkela, A., & Vuorinen, A. 2018, PhRvL, 120, 172703
Antoniadis, J., Freire, P. C. C., Wex, N., et al. 2013, Sci, 340, 448
Ardeljan, N. V., Bisnovatyi-Kogan, G. S., & Moiseenko, S. G. 2005, MNRAS, 359, 333
Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2018, ApJS, 235, 37
Arzoumanian, Z., Gendreau, K. C., Baker, C. L., et al. 2014, Proc. SPIE, 9144, 914420
Badnell, N. R., Bautista, M. A., Butler, K., et al. 2005, MNRAS, 360, 458
Baillot d'Etivaux, N., Guillot, S., Margueron, J., et al. 2019, arXiv:1905.01081
Barnard, J. J., & Arons, J. 1982, ApJ, 254, 713
Bauböck, M., Berti, E., Psaltis, D., & Özel, F. 2013, ApJ, 777, 68
Bauböck, M., Özel, F., Psaltis, D., & Morsink, S. M. 2015, ApJ, 799, 22
Bauböck, M., Psaltis, D., & Özel, F. 2019, ApJ, 872, 162
Baym, G., Hatsuda, T., Kojo, T., et al. 2018, RPPh, 81, 056902
Becker, W., Trümper, J., Lommen, A. N., & Backer, D. C. 2000, ApJ, 545, 1015
Behnel, S., Bradshaw, R., Citro, C., et al. 2011, CSE, 13, 31
Bildsten, L., Salpeter, E. E., & Wasserman, I. 1992, ApJ, 384, 143
Bilous, A. V., Watts, A. L., Harding, A. K., et al. 2019, ApJL, 887, L23
Blaes, O. M., Blandford, R. D., Madau, P., & Yan, L. 1992, ApJ, 399, 634
Bogdanov, S. 2013, ApJ, 762, 96
Bogdanov, S., & Grindlay, J. E. 2009, ApJ, 703, 1557
Bogdanov, S., Grindlay, J. E., & Rybicki, G. B. 2008, ApJ, 689, 407
Bogdanov, S., Guillot, S., Ray, P., et al. 2019a, ApJL, 887, L25
Bogdanov, S., Guillot, S., Ray, P., et al. 2019c, Constraining the Neutron Star Mass-Radius Relation and Dense Matter Equation of State with NICER. I. The Millisecond Pulsar X-Ray Data Set, Zenodo, doi:10.5281/zenodo.3524457
Bogdanov, S., Heinke, C. O., Özel, F., & Güver, T. 2016, ApJ, 831, 184
Bogdanov, S., Lamb, F. K., Mahmoodifar, S., et al. 2019b, ApJL, 887, L26
Bogdanov, S., Rybicki, G. B., & Grindlay, J. E. 2007, ApJ, 670, 668
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Burwitz, V., Zavlin, V. E., Neuhäuser, R., et al. 2001, A&A, 379, L35
Cackett, E. M., Miller, J. M., Ballantyne, D. R., et al. 2010, ApJ, 720, 205
Cackett, E. M., Miller, J. M., Bhattacharyya, S., et al. 2008, ApJ, 674, 415
Cadeau, C., Morsink, S. M., Leahy, D., & Campbell, S. S. 2007, ApJ, 654, 458
Carney, M. F., Wade, L. E., & Irwin, B. S. 2018, PhRvD, 98, 063004
Carson, Z., Steiner, A. W., & Yagi, K. 2019, PhRvD, 99, 043010
Cavecchi, Y., Watts, A. L., Braithwaite, J., & Levin, Y. 2013, MNRAS, 434, 3526
Cerutti, B., & Beloborodov, A. M. 2017, SSRv, 207, 111
Chambers, F. R. N., Watts, A. L., Keek, L., Cavecchi, Y., & Garcia, F. 2019, ApJ, 871, 61
Chang, P., & Bildsten, L. 2003, ApJ, 585, 464
Chang, P., & Bildsten, L. 2004, ApJ, 605, 830
Colgan, J., Kilcrease, D. P., Magee, N. H., et al. 2016, ApJ, 817, 116
Cook, G. B., Shapiro, S. L., & Teukolsky, S. A. 1994, ApJ, 424, 823
Cromartie, H., Fonseca, E., Ransom, S. M., et al. 2019, NatAs, in press (arXiv:1904.06759)
Dagum, L., & Menon, R. 1998, CSE, 5, 46
Dalcín, L., Paz, R., Storti, M., & D'Elía, J. 2008, JPDC, 68, 655
De, S., Finstad, D., Lattimer, J. M., et al. 2018, PhRvL, 121, 091102
Degenaar, N., Miller, J. M., Chakrabarty, D., et al. 2015, MNRAS, 451, L85
Demorest, P. B., Pennucci, T., Ransom, S. M., Roberts, M. S. E., & Hessels, J. W. T. 2010, Natur, 467, 1081
Drago, A., Lavagno, A., & Pagliara, G. 2014, PhRvD, 89, 043014
Drake, J. J., Marshall, H. L., Dreizler, S., et al. 2002, ApJ, 572, 996
Droettboom, M., Caswell, T. A., Hunter, J., et al. 2018, matplotlib/matplotlib v2.2.2, Zenodo, doi:10.5281/zenodo.1202077
Durant, M., Kargaltsev, O., Pavlov, G. G., et al. 2012, ApJ, 746, 6
Dyks, J., & Harding, A. K. 2004, ApJ, 614, 869
Fabian, A. C., Iwasawa, K., Reynolds, C. S., & Young, A. J. 2000, PASP, 112, 1145
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, arXiv:1306.2144
Forum, M. P. 1994, MPI: A Message-Passing Interface Standard, Tech. Rep., MPIF, https://www.mpi-forum.org/docs/mpi-1.0/mpi-10.ps
Galloway, D. K., Muno, M. P., Hartman, J. M., Psaltis, D., & Chakrabarty, D. 2008, ApJS, 179, 360
Garcia, F., Chambers, F. R. N., & Watts, A. L. 2018, PhRvF, 3, 123501
Gendreau, K. C., Arzoumanian, Z., Adkins, P. W., et al. 2016, Proc. SPIE, 9905, 99051H
Gil, J., Gronkowski, P., & Rudnicki, W. 1984, A&A, 132, 312
Gonzalez-Caniulef, D., Guillot, S., & Reisenegger, A. 2019, MNRAS, 490, 5848
Gough, B. 2009, GNU Scientific Library Reference Manual (3rd ed.; Network Theory Ltd.)
Gourgouliatos, K. N., & Hollerbach, R. 2018, ApJ, 852, 21
Gralla, S. E., Lupsasca, A., & Philippov, A. 2017, ApJ, 851, 137
Greif, S. K., Raaijmakers, G., Hebeler, K., Schwenk, A., & Watts, A. L. 2019, MNRAS, 485, 5363
Grenier, I. A., & Harding, A. K. 2015, CRPhy, 16, 641
Guillot, S. 2016, MmSAI, 87, 521
Guillot, S., Kaspi, V. M., Archibald, R. F., et al. 2016, MNRAS, 463, 2612
Guillot, S., & Rutledge, R. E. 2014, ApJL, 796, L3
Guillot, S., Rutledge, R. E., & Brown, E. F. 2011, ApJ, 732, 88
Guillot, S., Servillat, M., Webb, N. A., & Rutledge, R. E. 2013, ApJ, 772, 7
Güver, T., & Özel, F. 2013, ApJL, 765, L1
Güver, T., Psaltis, D., & Özel, F. 2012, ApJ, 747, 76
Hamaguchi, K. & NICER Calibration Team 2019, in 14th Workshop of Int. Astronomical Consortium for High Energy Calibration, https://indico2.riken.jp/event/2910/contributions/12756/attachments/8511/10339/NICER-IACHEC-2019.pdf
Handley, W. 2018, JOSS, 3, 849
Harding, A. K., & Muslimov, A. G. 2001, ApJ, 556, 987
Harding, A. K., & Muslimov, A. G. 2011, ApJ, 743, 181
Harrison, E. R., & Tademaru, E. 1975, ApJ, 201, 447
Hartle, J. B. 1967, ApJ, 150, 1005
Hartle, J. B., & Thorne, K. S. 1968, ApJ, 153, 807
Hebeler, K., Lattimer, J. M., Pethick, C. J., & Schwenk, A. 2010, PhRvL, 105, 161102
Heinke, C. O., Cohn, H. N., Lugger, P. M., et al. 2014, MNRAS, 444, 443
Heinke, C. O., Rybicki, G. B., Narayan, R., & Grindlay, J. E. 2006, ApJ, 644, 1090
Heyl, J. S. 2004, ApJ, 600, 939

Higson, E. 2018, JOSS, 3, 916
Higson, E., Handley, W., Hobson, M., & Lasenby, A. 2018, BayAn, 13, 873
Higson, E., Handley, W., Hobson, M., & Lasenby, A. 2019, MNRAS, 483, 2044
Ho, W. C. G., & Heinke, C. O. 2009, Natur, 462, 71
Ho, W. C. G., Kaplan, D. L., Chang, P., van Adelsberg, M., & Potekhin, A. Y. 2007, MNRAS, 375, 821
Ho, W. C. G., & Lai, D. 2001, MNRAS, 327, 1081
Hobbs, G. B., Edwards, R. T., & Manchester, R. N. 2006, MNRAS, 369, 655
Hunter, J. D. 2007, CSE, 9, 90
Iglesias, C. A., & Rogers, F. J. 1996, ApJ, 464, 943
Johnson, T. J., Venter, C., Harding, A. K., et al. 2014, ApJS, 213, 6
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open Source Scientific Tools for Python (http://www.scipy.org/)
Kajava, J. J. E., Nättilä, J., Latvala, O.-M., et al. 2014, MNRAS, 445, 4218
Kass, R. E., & Raftery, A. E. 1995, J. Am. Stat. Assoc., 90, 773
Kijak, J., & Gil, J. 2003, A&A, 397, 969
Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Schmidt (Amsterdam: IOS Press), 87
Kramer, M., & Wex, N. 2009, CQGra, 26, 073001
Kullback, S., & Leibler, R. A. 1951, Ann. Math. Statist., 22, 79
Kumar, B., & Landry, P. 2019, PhRvD, 99, 123026
Kurkela, A., Fraga, E. S., Schaffner-Bielich, J., & Vuorinen, A. 2014, ApJ, 789, 127
Landry, P., & Essick, R. 2019, PhRvD, 99, 084049
Lattimer, J. M., & Prakash, M. 2016, PhR, 621, 127
Lattimer, J. M., & Steiner, A. W. 2014, ApJ, 784, 123
Lee, H., Kashyap, V. L., van Dyk, D. A., et al. 2011, ApJ, 731, 126
Li, J. J., & Sedrakian, A. 2019, ApJL, 874, L22
Lim, Y., & Holt, J. W. 2018, PhRvL, 121, 062701
Lindblom, L. 2018, PhRvD, 97, 123019
Lo, K. H., Miller, M. C., Bhattacharyya, S., & Lamb, F. K. 2013, ApJ, 776, 19
Lockhart, W., Gralla, S. E., Özel, F., & Psaltis, D. 2019, MNRAS, 490, 1774
Lommen, A. N., Kipphorn, R. A., Nice, D. J., et al. 2006, ApJ, 642, 1012
Lommen, A. N., Zepka, A., Backer, D. C., et al. 2000, ApJ, 545, 1007
Long, M., Romanova, M. M., & Lovelace, R. V. E. 2007, MNRAS, 374, 436
Long, M., Romanova, M. M., & Lovelace, R. V. E. 2008, MNRAS, 386, 1274
Ludlam, R. M., Miller, J. M., Arzoumanian, Z., et al. 2018, ApJL, 858, L5
Ludlam, R. M., Miller, J. M., Bachetti, M., et al. 2017, ApJ, 836, 140
Malik, T., Alam, N., Fortin, M., et al. 2018, PhRvC, 98, 035804
Margueron, J., Hoffmann Casali, R., & Gulminelli, F. 2018a, PhRvC, 97, 025805
Margueron, J., Hoffmann Casali, R., & Gulminelli, F. 2018b, PhRvC, 97, 025806
McNeil Forbes, M., Bose, S., Reddy, S., et al. 2019, PhRvD, 100, 083010
Melatos, A., & Phinney, E. S. 2001, PASA, 18, 421
Miller, J. M., Parker, M. L., Fuerst, F., et al. 2013, ApJL, 779, L2
Miller, M. C. 2013, arXiv:1312.0029
Miller, M. C., & Lamb, F. K. 1998, ApJL, 499, L37
Miller, M. C., & Lamb, F. K. 2015, ApJ, 808, 31
Miller, M. C., Lamb, F. K., Dittmann, A. J., et al. 2019, ApJL, 887, L24
Mitchell, J. P., Braithwaite, J., Reisenegger, A., et al. 2015, MNRAS, 447, 1213
Montaña, G., Tolós, L., Hanauske, M., & Rezzolla, L. 2019, PhRvD, 99, 103009
Morsink, S. M., Leahy, D. A., Cadeau, C., & Braga, J. 2007, ApJ, 663, 1244
Most, E. R., Weih, L. R., Rezzolla, L., & Schaffner-Bielich, J. 2018, PhRvL, 120, 261103
Nandra, K., Barret, D., Barcons, X., et al. 2013, arXiv:1306.2307
Nättilä, J., Miller, M. C., Steiner, A. W., et al. 2017, A&A, 608, A31
Nättilä, J., & Pihajoki, P. 2018, A&A, 615, A50
Nättilä, J., Steiner, A. W., Kajava, J. J. E., Suleimanov, V. F., & Poutanen, J. 2016, A&A, 591, A25
Obergaulinger, M., & Aloy, M. Á. 2017, JPhCS, 932, 012043
Oertel, M., Hempel, M., Klähn, T., & Typel, S. 2017, RvMP, 89, 015007
Oliphant, T. E. 2007, CSE, 9, 10
Özel, F., & Freire, P. 2016, ARA&A, 54, 401
Özel, F., & Psaltis, D. 2009, PhRvD, 80, 103003
Özel, F., Psaltis, D., Güver, T., et al. 2016, ApJ, 820, 28
Patruno, A., & Watts, A. L. 2012, arXiv:1206.2727
Pavlov, G. G., & Zavlin, V. E. 1997, ApJL, 490, L91
Payne, D. J. B., & Melatos, A. 2004, MNRAS, 351, 569
Pechenick, K. R., Ftaclas, C., & Cohen, J. M. 1983, ApJ, 274, 846
Perez, F., & Granger, B. E. 2007, CSE, 9, 21
Philippov, A. A., Cerutti, B., Tchekhovskoy, A., & Spitkovsky, A. 2015a, ApJL, 815, L19
Philippov, A. A., & Spitkovsky, A. 2018, ApJ, 855, 94
Philippov, A. A., Spitkovsky, A., & Cerutti, B. 2015b, ApJL, 801, L19
Piro, A. L., & Bildsten, L. 2005, ApJ, 629, 438
Pons, J. A., Walter, F. M., Lattimer, J. M., et al. 2002, ApJ, 564, 981
Poutanen, J., & Beloborodov, A. M. 2006, MNRAS, 373, 836
Poutanen, J., & Gierliński, M. 2003, MNRAS, 343, 1301
Psaltis, D., & Özel, F. 2014, ApJ, 792, 87
Psaltis, D., Özel, F., & Chakrabarty, D. 2014, ApJ, 787, 136
Raaijmakers, G., Riley, T. E., & Watts, A. L. 2018, MNRAS, 478, 2177
Raaijmakers, G., Riley, T. E., Watts, A. L., et al. 2019, ApJL, 887, L22
Radhakrishnan, V., & Cooke, D. J. 1969, ApL, 3, 225
Radice, D., Perego, A., Hotokezaka, K., et al. 2018, ApJL, 869, L35
Raithel, C. A., Özel, F., & Psaltis, D. 2016, ApJ, 831, 44
Raithel, C. A., Özel, F., & Psaltis, D. 2017, ApJ, 844, 156
Rangelov, B., Pavlov, G. G., Kargaltsev, O., et al. 2017, ApJ, 835, 264
Ray, P. S., Arzoumanian, Z., Ballantyne, D., et al. 2019, arXiv:1903.03035
Read, J. S., Lackey, B. D., Owen, B. J., & Friedman, J. L. 2009, PhRvD, 79, 124032
Reardon, D. J., Hobbs, G., Coles, W., et al. 2016, MNRAS, 455, 1751
Reisenegger, A. 2009, A&A, 499, 557
Riley, T. E. 2019, A NICER View of PSR J0030+0451: Nested Samples for Millisecond Pulsar Parameter Estimation, v1.0, Zenodo, doi:10.5281/zenodo.3386449
Riley, T. E., Raaijmakers, G., & Watts, A. L. 2018, MNRAS, 478, 1093
Riley, T. E., & Watts, A. L. 2019, ApJS, submitted
Romani, R. W. 1990, Natur, 347, 741
Salmi, T., Nättilä, J., & Poutanen, J. 2018, A&A, 618, A161
Shaw, A. W., Heinke, C. O., Steiner, A. W., et al. 2018, MNRAS, 476, 4713
Shlens, J. 2014, CoRR, arXiv:1404.2000
Skilling, J. 2006, BayAn, 1, 833
Spitkovsky, A., Levin, Y., & Ushomirsky, G. 2002, ApJ, 566, 1018
Steiner, A. W., Heinke, C. O., Bogdanov, S., et al. 2018, MNRAS, 476, 421
Steiner, A. W., Lattimer, J. M., & Brown, E. F. 2010, ApJ, 722, 33
Steiner, A. W., Lattimer, J. M., & Brown, E. F. 2013, ApJL, 765, L5
Stevens, A. L., Fiege, J. D., Leahy, D. A., & Morsink, S. M. 2016, ApJ, 833, 244
Strobel, K., Schaab, C., & Weigel, M. K. 1999, A&A, 350, 497
Suleimanov, V., Poutanen, J., Revnivtsev, M., & Werner, K. 2011a, ApJ, 742, 122
Suleimanov, V., Poutanen, J., & Werner, K. 2011b, A&A, 527, A139
Tews, I., Carlson, J., Gandolfi, S., & Reddy, S. 2018a, ApJ, 860, 149
Tews, I., Lattimer, J. M., Ohnishi, A., & Kolomeitsev, E. E. 2017, ApJ, 848, 105
Tews, I., Margueron, J., & Reddy, S. 2018b, PhRvC, 98, 045804
The LIGO Scientific Collaboration, the Virgo Collaboration, Abbott, B. P., et al. 2019, PhRvX, 9, 031040
Timokhin, A. N., & Arons, J. 2013, MNRAS, 429, 20
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
van Doesburgh, M., van der Klis, M., & Morsink, S. M. 2018, MNRAS, 479, 426
Vehtari, A., & Ojanen, J. 2012, Statist. Surv., 6, 142
Viganò, D., Rea, N., Pons, J. A., et al. 2013, MNRAS, 434, 123
Vincent, F. H., Bejger, M., Różańska, A., et al. 2018, ApJ, 855, 116
Watts, A. L. 2012, ARA&A, 50, 609
Watts, A. L. 2019, in AIP Conf. Ser. 2127, Xiamen-CUSTIPEN Workshop on the EOS of Dense Neutron-Rich Matter in the Era of Gravitational Wave Astronomy, ed. A. Li, B.-A. Li, & F. Xu (Melville, NY: AIP) 020008
Watts, A. L., Andersson, N., Chakrabarty, D., et al. 2016, RvMP, 88, 021001
Watts, A. L., Yu, W., Poutanen, J., et al. 2019, SCPMA, 62, 29503
Webb, N. A., & Barret, D. 2007, ApJ, 671, 727
Weih, L. R., Most, E. R., & Rezzolla, L. 2019, ApJ, 881, 73
Wilkinson, T., Patruno, A., Watts, A., & Uttley, P. 2011, MNRAS, 410, 1513
Wilms, J., Allen, A., & McCray, R. 2000, ApJ, 542, 914
Xu, J., van Dyk, D. A., Kashyap, V. L., et al. 2014, ApJ, 794, 97
Zavlin, V. E., Pavlov, G. G., & Shibanov, Y. A. 1996, A&A, 315, 141
Zdunik, J. L., & Haensel, P. 2013, A&A, 551, A61
Zhang, S., Santangelo, A., Feroci, M., et al. 2019, SCPMA, 62, 29502