

Systematic Effects of Hydrogen and Helium Atmosphere Mismatch on Radius Inference in PSR J0740+6620-like Synthetic NICER Data

ISIAH M. HOLT ^{1,2}, M. COLEMAN MILLER ^{1,3}, ALEXANDER J. DITTMANN ⁴, AND FREDERICK K. LAMB ^{5,6}

¹*Department of Astronomy, University of Maryland, College Park, MD 20742-2421, USA*

²*NASA Goddard Space Flight Center, Greenbelt, MD, USA*

³*Joint Space-Science Institute, University of Maryland, College Park, MD 20742-2421, USA*

⁴*Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, USA**

⁵*Illinois Center for Advanced Studies of the Universe and Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801-3080, USA*

⁶*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801-3074, USA*

ABSTRACT

Constraints on neutron star radii provide insight into the properties of the cold, dense matter in their interiors. Previous studies using synthetic Neutron star Interior Composition Explorer (NICER) pulse waveform data have demonstrated that radius inferences derived therefrom are robust against several classes of modeling systematics. Here we explore the consequences of assuming the wrong atmospheric composition, using synthetic data based on the $\sim 2.1 M_{\odot}$ pulsar PSR J0740+6620. We find that the assumption of a hydrogen atmosphere when the synthetic data assumed a helium atmosphere, or vice versa, produces little bias in the inferred radius at the spot-to-background ratio of the actual PSR J0740+6620 data. However, when we increase the spot-to-background ratio by a factor of ~ 20 while keeping the total number of counts fixed at the observed $\sim 5.5 \times 10^5$, we find that composition mismatch can produce significantly biased radius estimates while remaining hidden inside a fit with statistically acceptable residuals. Even in these cases, the Bayesian evidence consistently identifies the correct atmospheric model. Our findings reinforce the importance of using the Bayesian evidence for model comparison and goodness-of-fit tests.

Keywords: Millisecond pulsars (1062); X-ray stars (1823); Neutron stars (1108); Neutron star cores (1107)

1. INTRODUCTION

The cores of neutron stars consist of matter that is theorized to be catalyzed to its ground state at densities up to several times nuclear saturation density ($n_s \approx 0.16 \text{ fm}^{-3}$, equivalent to a mass density $\rho_s \approx 2.6 \times 10^{14} \text{ g cm}^{-3}$). Laboratories are unable to replicate the high densities, low temperatures, and extreme neutron–proton asymmetries present in neutron star cores. Thus, observations of neutron stars provide unique insights into the nature of dense matter. Most of the macroscopic properties of a neutron star, such as its mass, radius, and tidal deformability, depend on the equation of state (EOS) of the matter in its interior, i.e., the pressure as a function of the energy density. Over the past decade, numerous ob-

servations have helped constrain the EOS. These include measurements of high neutron star masses such as PSR J1614–2230 ($M = 1.97 \pm 0.04 M_{\odot}$, [Demorest et al. 2010](#)), PSR J0348+0432 ($M = 1.806 \pm 0.037 M_{\odot}$, [Saffer et al. 2024](#)), PSR J0740+6620 ($M = 2.08 \pm 0.07 M_{\odot}$, [Fonseca et al. 2021](#)), and PSR J0952–0607 ($M = 2.35 \pm 0.17 M_{\odot}$, [Romani et al. 2022](#)). They also include constraints on the tidal deformability of neutron stars from gravitational-wave observations of the binary neutron star merger GW170817 ([Abbott et al. 2017](#); [Abbott et al. 2018](#); [De et al. 2018](#)), and estimates of neutron star radii from thermal X-ray observations. Among these, measurements of the radii of neutron stars with precisely determined masses are especially valuable.

Fits of nonmagnetic helium model atmospheres to the spectra of cooling neutron stars in quiescent low-mass X-ray binary systems (qLMXBs; [Guillot et al. 2011, 2013](#); [Lattimer & Steiner 2014](#); [Bogdanov et al. 2016](#)) com-

* NASA Einstein Fellow

monly yield radius estimates $\sim 50\%$ larger than fits of nonmagnetic hydrogen model atmospheres to the same data (Servillat et al. 2012; Catuneanu et al. 2013; Heinke et al. 2014; see also Miller & Lamb 2016 for a review). The origin of this discrepancy could lie in a spectral shape degeneracy. Hydrogen and helium atmospheres produce emergent spectra of different hardness at the same effective temperature, so fitting a hydrogen atmosphere model to helium atmosphere data (or vice versa) forces a trade-off between the inferred temperature and emitting area, and hence the radius. The soft X-ray spectra of these sources are roughly thermal and lack strong composition-dependent features, so the data are often insufficient to distinguish between a smaller, hotter neutron star with one composition and a larger, cooler neutron star with another (see, e.g., Miller & Lamb 2016). Similarly, hydrogen and carbon model atmospheres can give equally good fits to the spectra of isolated cooling neutron stars while returning very different radii (see, e.g., Klochkov et al. 2015). In these spectral fitting methods, the composition ambiguity is difficult to resolve because the data lack the spectral features or independent geometric constraints needed to distinguish atmospheric models.

The Neutron Star Interior Composition Explorer (NICER; Gendreau et al. 2012) conducts phase- and energy-resolved measurements of the thermal X-ray pulses produced by some rotating neutron stars, which can be used to constrain the masses and radii of these stars. Mass and radius constraints have been derived from NICER data, sometimes supplemented with X-ray Multi-Mirror (XMM-Newton) data, for several neutron stars. These include the $\sim 1.4 M_{\odot}$ pulsar PSR J0030+0451 (Miller et al. 2019; Riley et al. 2019; Vinciguerra et al. 2024; Kini et al. 2026), the $\sim 2.1 M_{\odot}$ pulsar PSR J0740+6620 (Miller et al. 2021; Riley et al. 2021; Salmi et al. 2022; Dittmann et al. 2024; Salmi et al. 2024a; this pulsar is especially important because of its high mass and thus high central density), the $\sim 1.4 M_{\odot}$ pulsar PSR J0437–4715 (Choudhury et al. 2024; Miller et al. 2026), the $\sim 1.04 M_{\odot}$ pulsar PSR 1231–1411 (Salmi et al. 2024b; Qi et al. 2025), the $\sim 1.4 M_{\odot}$ pulsar PSR J0164–3329 (Mauviard et al. 2025), and the $\sim 1.8 M_{\odot}$ pulsar PSR J2124–3358 (González-Caniulef et al. 2026). In all such analyses, the atmosphere of the neutron star plays an important role, as it determines the angular distribution of the thermal X-ray radiation leaving the stellar surface.

In pulse-profile modeling (PPM), the predicted X-ray pulse waveform depends not only on the stellar compactness, observer geometry, and emission region properties, but also on the beaming of the emergent radia-

tion, which is influenced by the atmospheric composition (see, e.g., Pechenick et al. 1983; Poutanen & Gierliński 2003; Lo et al. 2013; Miller & Lamb 2015). Published PPM analyses of NICER data have predominantly assumed partially and/or fully ionized hydrogen atmospheres when modeling the X-ray emission (see, e.g., Miller et al. 2019; Riley et al. 2019; Miller et al. 2021; Riley et al. 2021; Salmi et al. 2024a; Dittmann et al. 2024; Choudhury et al. 2024; Miller et al. 2026).

Salmi et al. (2023) investigated atmosphere model systematics in NICER pulse-profile analyses of PSR J0030+0451 and PSR J0740+6620, including fully ionized hydrogen and helium models. They found that none of the atmosphere cases significantly changed the inferred radius of PSR J0740+6620, which they attributed potentially to the source’s X-ray faintness, tighter external constraints, and/or view geometry. Our synthetic experiment is complementary to that result. We confirm their result at low source-to-background ratios, but demonstrate that the modeling assumption of atmospheric composition can introduce overwhelming systematic biases when analyzing higher-fidelity data.

In this paper, we check whether atmospheric composition mismatch can bias the inferred radius of a PSR J0740+6620-like neutron star while remaining hidden inside a statistically acceptable fit. We generate synthetic NICER-like pulse-profile data assuming fully ionized hydrogen or fully ionized helium atmospheres, and fit the hydrogen-generated data with both atmosphere models and the helium-generated data with both models. In each case, we vary the ratio of hot spot counts to background counts while holding the total expected count number fixed at $\sim 5.5 \times 10^5$, which is the number observed from PSR J0740+6620, in order to explore how the detectability of the composition mismatch depends on the relative strength of the pulsed signal. For each configuration, we perform input-parameter evaluations, full posterior sampling, goodness-of-fit assessments, and Bayesian evidence comparisons.

We find that in both directions of the atmospheric mismatch, the incorrect atmosphere model can yield statistically acceptable phase-channel and bolometric χ^2 values after posterior exploration, yet still produce biased radius estimates. The effect is direction-dependent. Fitting hydrogen-generated data with a helium model can increase the inferred radius by more than 2σ , whereas fitting helium-generated data with a hydrogen model produces smaller increases that remain within 2σ for the cases we examined. However, at the actual count level and spot-to-background ratio estimated for PSR J0740+6620, fitting hydrogen-generated data with a helium model does not produce substantial bias in the

inferred radius. Despite the statistical acceptability of these fits, the Bayesian evidence consistently favors the correct atmospheric model. The apparent asymmetry between the two mismatch directions should not be over-interpreted, as it may depend on the assumed geometry, count level, background conditions, and, in particular, we expect it to weaken for less compact stars.

The remainder of this paper is organized as follows. Section 2 describes the effects of atmospheric beaming of the emergent radiation on the pulse profile. Section 3 describes our methods for generating and analyzing the synthetic data. Section 4 presents our results. We discuss the implications of these results in Section 5 and summarize our conclusions in Section 6.

2. ATMOSPHERIC BEAMING AND ITS EFFECTS ON THE PULSE PROFILE

PPM is less susceptible to atmospheric uncertainty than spectral fitting methods because the pulse waveform carries geometric information via phase-dependent flux modulation. The time-varying projection of a heated region (“hot spot”) as the star rotates encodes the observer inclination, the spot colatitude, and the stellar compactness through the relativistic light-bending that maps surface emission angles to the observer’s line of sight. Even if the beaming function is incorrect, the phase at which the spot appears and disappears and the overall depth of the modulation are primarily determined by the spot geometry and the star’s compactness, reducing the freedom available for an incorrect atmosphere to absorb the mismatch. Indeed, analyses of synthetic data have shown that PPM radius estimates are robust against several other classes of systematic error. [Lo et al. \(2013\)](#) and [Miller & Lamb \(2015, 2016\)](#) demonstrated that incorrect assumptions about the shapes and temperature distributions of the hot spots do not significantly bias the inferred radius, provided the fit is statistically acceptable. [Holt et al. \(2025\)](#) showed that in joint NICER/XMM-Newton analyses, even a factor-of-five underestimate of the XMM-Newton background shifts the radius posterior by only $\sim 1\sigma$. These results have strengthened the case for PPM as a method for reliable radius inference. To our knowledge, however, no prior synthetic data study has systematically checked whether atmospheric composition mismatch, specifically, data generated with one atmosphere using a model that assumes another, can produce a comparably small or potentially larger bias.

In fully ionized atmospheres at the effective temperatures relevant to NICER ($kT_{\text{eff}} \sim 0.1$ keV), the dominant sources of continuum opacity are free-free (inverse bremsstrahlung) absorption and electron scat-

tering. The free-free absorption coefficient scales as $\alpha_{\text{ff}} \propto Z^2 n_i n_e / \nu^3$ for ion charge Z , ion and electron number densities n_i and n_e , and photon frequency ν (see, e.g., [Rybicki & Lightman 1979](#)). The composition dependence of this expression is more subtle than the explicit Z^2 factor suggests. At a given mass density ρ , a fully ionized hydrogen atmosphere has $n_i = n_e = \rho/m_{\text{H}}$, whereas a fully ionized helium atmosphere has $n_i = \rho/4m_{\text{H}}$ and $n_e = \rho/2m_{\text{H}}$. A helium nucleus is four times the mass of a proton and supplies only two electrons. As a result, helium offers fewer targets and fewer absorbers per gram. The reduction in n_i cancels the factor $Z^2 = 4$ exactly, and the accompanying reduction in n_e leaves α_{ff} a factor of two lower for helium than for hydrogen at the same density and temperature.

The beaming pattern, i.e., the intensity as a function of the angle θ between the outgoing photon direction and the surface normal, is shaped by how the radiation decouples from the atmosphere, and the difference between the two compositions is therefore not a simple consequence of the ionic charge. On the basis of the scalings above, we expect it to be modest rather than dramatic. Figure 1 shows the emergent beaming patterns for both of the model atmospheres. At every energy, the helium atmosphere retains more intensity toward grazing angles than hydrogen, meaning that its beaming is closer to isotropic, whereas hydrogen is more limb-darkened, consistent with previous calculations (see, e.g., [Bogdanov et al. 2021](#); [Salmi et al. 2023](#)). The two patterns are nonetheless similar in overall shape, and the difference between them is small. Its magnitude depends on the effective temperature, surface gravity, and photon energy.

For PPM, the consequence of this beaming difference is that a helium atmosphere directs relatively more radiation toward grazing angles compared to a hydrogen atmosphere. When the star rotates and the hot spot sweeps across the observer’s line of sight, the angular distribution of the emitted radiation determines how steeply the observed flux rises and falls with pulse phase. A more limb-darkened (hydrogen-like) pattern produces a sharper pulse profile with deeper modulation for a given geometry, whereas a more isotropic (helium-like) pattern produces a broader, shallower pulse. We illustrate this in Figures 2 and 3, which compare synthetic pulse profiles generated with hydrogen and helium atmospheres for PSR J0740+6620-like geometry. At the spot-to-background ratio of PSR J0740+6620, the two compositions yield nearly identical waveforms, with fractional modulations of $f_{\text{H}} = 0.034$ and $f_{\text{He}} = 0.033$ (Figure 2). Amplifying the ratio by a factor of 60 to suppress the background exposes the shape difference, yielding

Table 1. Primary Parameters in the Pulse Waveform Model

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

NOTE—This table gives the pulse waveform model that [Miller et al. \(2021\)](#) fit jointly to the NICER and XMM-Newton data on PSR J0740+6620, the priors on the parameters, and the best-fit values. [Miller et al. \(2021\)](#) used flat priors of $0.0-3.0$ rad for $\Delta\theta_1$ and $\Delta\theta_2$. In the present work, we adopted slightly wider priors of $0.0-3.14$ rad for both parameters. The 12 parameters listed here are the primary parameters in the pulse waveform model used in [Miller et al. \(2021\)](#). In the present work, we focus on the equatorial circumferential radius R_e . We generated synthetic pulse waveform data by Poisson-sampling this best-fit waveform at the same 346.532 Hz rotational frequency assumed by [Miller et al. \(2021\)](#). Unlike [Miller et al. \(2021\)](#), which analyzed synthetic joint NICER and XMM-Newton data, we generate and analyze only synthetic NICER data. The priors listed here are those used when fitting the model to our synthetic datasets. See the text for further details.

fractional modulations, $f_H = 0.484$ and $f_{\text{He}} = 0.471$ (Figure 3). In both regimes, the hydrogen atmosphere produces the slightly deeper, sharper pulse expected from its stronger limb darkening.

This is particularly relevant for PSR J0740+6620, a $\sim 2.1 M_\odot$ pulsar ([Fonseca et al. 2021](#)) whose high mass makes it valuable for constraining the EOS of cold, dense matter at densities above those accessible through observations of $\sim 1.4 M_\odot$ neutron stars (see, e.g., [Miller et al. 2021](#); [Riley et al. 2021](#); [Salmi et al. 2024a](#); [Dittmann et al. 2024](#)). The NICER observations of PSR J0740+6620 are severely background-dominated ([Wolff et al. 2021](#)), with the estimated ratio of hot spot counts to total counts in the NICER band being of order a few percent. This means that the pulsed signal from the spots is superimposed on a large background of counts from other X-ray sources in the field. This low spot-to-background ratio has two competing implications for the composition mismatch problem. It may make the waveform less sensitive to the beaming pattern, because the beaming-dependent modulation constitutes a smaller fraction of the total signal and the statistical uncertainties in the pulse shape are large. On the other hand, if the beaming mismatch does shift the inferred parameters, the large background means that

the shift may be harder to detect through standard goodness-of-fit diagnostics.

3. METHODS

In this section we describe our procedure for generating synthetic NICER-like data for fully ionized hydrogen and helium atmospheres, and for fitting each dataset with both atmosphere models. We first describe the baseline configuration (§3.1) and then the spot-to-background scaling experiment (§3.2), the synthetic data generation procedure (§3.3), the likelihood and normalization treatment (§3.4), the two-stage analysis (§3.5), the fit diagnostics (§3.6), the Bayesian evidence calculation (§3.7), and the radius bias diagnostic (§3.8).

3.1. Baseline J0740-like Configuration

We adopt the two-uniform-temperature spot geometry and best-fit 12-parameter solution from [Miller et al. \(2021\)](#) as our reference configuration. The parameters are the gravitational mass M , equatorial circumferential radius R_e , observer inclination i , hydrogen column density N_H , source distance d , and the colatitudes, angular radii, and effective temperatures of the two hot spots. Table 1 lists these 12 parameters, the ranges of

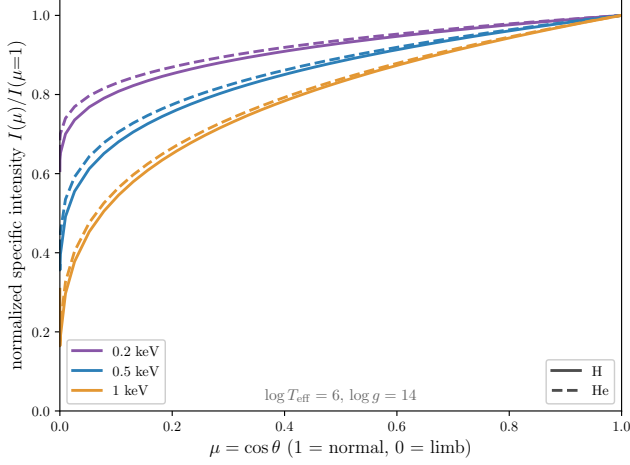


Figure 1. Emergent specific intensity as a function of emission angle (the beaming pattern) for hydrogen (solid) and helium (dashed) atmospheres at $\log T_{\text{eff}} = 6$ and $\log g = 14$. The θ in $\mu = \cos \theta$ is measured from the local surface normal, so $\mu = 1$ is normal emission and $\mu = 0$ is the limb. Each curve is normalized to its emission value $I(\mu=1)$. Line colors denote photon energy (purple: 0.2 keV; blue: 0.5 keV; orange: 1 keV). At every energy, the helium atmosphere retains more intensity towards grazing angles than hydrogen; its beaming is closer to isotropic, while hydrogen is more limb-darkened, and the limb darkening steepens with increasing photon energy.

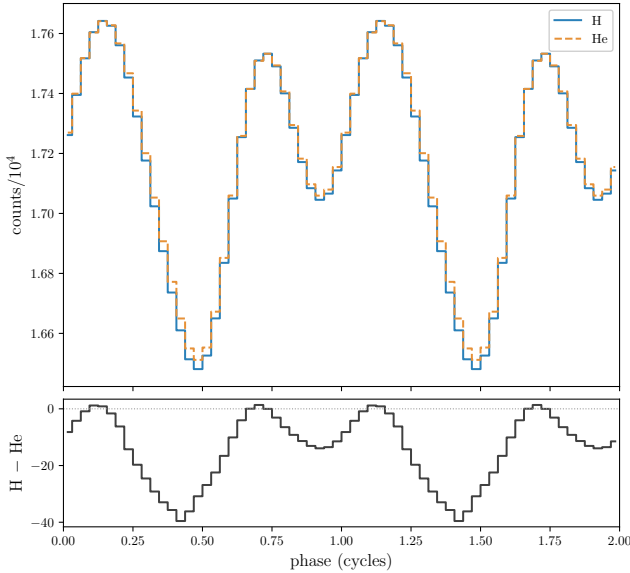


Figure 2. Synthetic pulse profile for a PSR J0740+6620-like geometry with hydrogen (blue, solid) and helium (orange, dashed) atmospheres at the realistic spot-to-background ratio. The lower panel shows $H - \text{He}$. The pulsed signal is a few percent of the total counts and the two compositions are statistically indistinguishable, with fractional modulations $f_H = 0.034$ and $f_{\text{He}} = 0.033$.

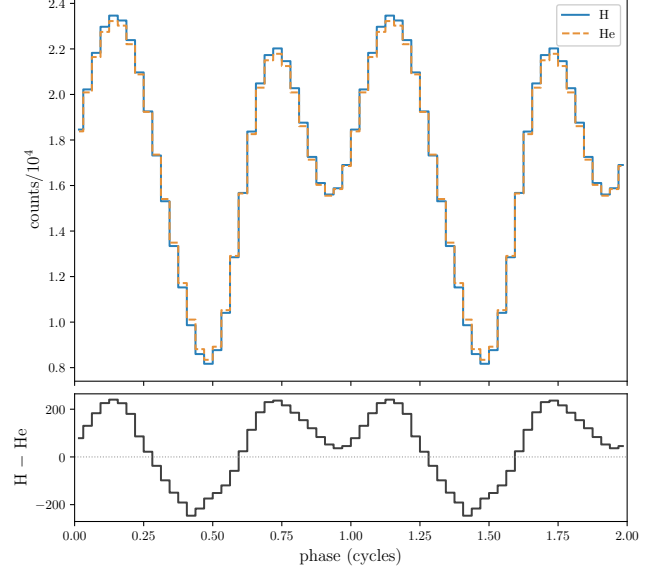


Figure 3. Synthetic pulse profile for a PSR J0740+6620-like geometry with the spot contribution amplified by a factor of 60. The model assuming a hydrogen atmosphere produces a marginally deeper pulse with fractional modulations $f_H = 0.484$ and $f_{\text{He}} = 0.471$, which is consistent with limb darkening for a hydrogen atmosphere.

the priors Miller et al. (2021) used when they fit this pulse waveform model to the actual NICER and XMM-Newton data, and the best-fit values of these parameters. As noted in Table 1, we adopted slightly wider priors for the spot angular radii $\Delta\theta_1$ and $\Delta\theta_2$ than were used by Miller et al. (2021). The purpose of choosing a specific reference configuration is not to exhaust all possible pulsar geometries, but to check the consequences of atmospheric composition mismatch in a realistic setting.

We set the total expected number of counts in the synthetic datasets to $N_{\text{counts}} \sim 5.5 \times 10^5$, consistent with the NICER data on PSR J0740+6620 used in Miller et al. (2021). The baseline ratio of hot-spot counts to background counts, $R_0 \equiv S_0/B_0$, reflects the background-dominated conditions of the actual NICER observations, in which only a few percent of the total counts originate from the pulsed hot-spot emission.

3.2. Spot-to-Background Ratio Experiment

To explore how the consequences of atmospheric composition mismatch depend on the relative strength of the pulsed signal, we systematically vary the ratio of the pulsed signal, we systematically vary the ratio of hot-spot counts to background counts while holding the total expected count number fixed. We define the multiplier m as

$$m \equiv \frac{R_{\text{synth}}}{R_0}, \quad (1)$$

where R_{synth} is the rescaled spot-to-background count ratio and $R_0 \equiv \frac{S_0}{B_0}$ is the baseline ratio. The total number of counts is:

$$N_{\text{counts}} = S_{\text{synth}} + B_{\text{synth}}, \quad (2)$$

where S_{synth} and B_{synth} are the rescaled total spot and background counts, respectively. We require their ratio to satisfy

$$\frac{S_{\text{synth}}}{B_{\text{synth}}} = R_{\text{synth}}. \quad (3)$$

For clarity, the count normalization used in both generation and likelihood evaluations can be written explicitly. For each multiplier we define

$$a_m \equiv \frac{S_{\text{synth}}}{S_0}, \quad b_m \equiv \frac{B_{\text{synth}}}{B_0}. \quad (4)$$

If $S_{jk}^{(0)}$ and $B_{jk}^{(0)}$ are the baseline expected spot and background counts in energy channel j and phase bin k , the mean counts used to generate the synthetic data are

$$\lambda_{jk}^{(m)} = a_m S_{jk}^{(0)} + b_m B_{jk}^{(0)}. \quad (5)$$

The waveform calculation implements the spot scaling through the exposure time: for each synthetic dataset we use $T_{\text{obs}}^{(m)} = a_m T_{\text{obs}}^{(0)}$. Because the predicted spot counts scale linearly with exposure, evaluation of the same 12 physical input parameters for that dataset predicts $a_m S_0 = S_{\text{synth}}$ spot counts, rather than the original S_0 . Thus, the increased spot normalization is not supplied by an additional signal normalization parameter. The factor b_m is used to construct the synthetic background. The treatment of the background during fitting is described in §3.4.

By choosing m values 1, 20, 40, 60, 80, and 100, we explore configurations ranging from the realistic, heavily background-dominated regime of PSR J0740+6620 to strongly spot-dominated scenarios similar to the millisecond X-ray pulsar PSR J0437-4715. The high multiplier cases are controlled experiments and are not intended to represent PSR J0740+6620 itself with otherwise unchanged observational properties. A source with such a large pulsed fraction would generally require the observational setup and background assumptions to be reconsidered. We also explored higher multipliers such as 120, 140, 160 and 180, but these multipliers do not add significant information, given the already very high signal-to-noise.

3.3. Synthetic Data Generation

The hot-spot emission spectra and beaming patterns are computed using the NSX code for fully ionized model

atmospheres (Ho & Lai 2001), following the same approach adopted in Miller et al. (2021). We note that the surface gravity of a neutron star is large enough that the lightest element present is expected to settle to the top of the atmosphere within seconds to minutes (based on extrapolations of calculations in Alcock & Illarionov 1980). As hydrogen is the most abundant element in the universe, and PSR J0740+6620 likely underwent prolonged accretion to reach its current spin frequency, a pure hydrogen atmosphere is a well-motivated assumption (Romani 1987; Bogdanov et al. 2019). However, the true surface composition is uncertain. A helium atmosphere is possible if the accreted material was hydrogen-poor or if subsequent processes altered the composition. We therefore check both fully ionized hydrogen and helium atmospheres.

The NICER data for PSR J0740+6620 are recorded in 94 energy channels and are binned into 32 rotational phase bins. We refer to each of the $32 \times 94 = 3008$ combinations of a phase bin and an energy channel as a phase-channel bin. For the synthetic hydrogen data, we compute the expected phase-channel count distribution using the fully ionized hydrogen atmosphere model. The baseline per-bin spot and background distributions are scaled to the new totals S_{synth} and B_{synth} from §3.2 without changing their shape, and we draw a Poisson realization of the counts in each phase-channel bin. For each multiplier m we perform an independent Poisson realization. The resulting dataset is then fit with both the hydrogen atmosphere model and the helium atmosphere model. For the synthetic helium data, we repeat the procedure using the fully ionized helium atmosphere model to generate the expected counts. Each multiplier again corresponds to an independent Poisson realization, and the resulting dataset is fit with both the helium atmosphere model and the hydrogen atmosphere model.

In both cases, the synthetic data are generated using the reference 12-parameter configuration described in §3.1. The only difference between the two families is the atmosphere model.

3.4. Likelihood and normalization

We evaluate the NICER data using the phase-and energy-resolved Poisson likelihood adopted in previous analyses of PSR J0740+6620 (Miller et al. 2021; Dittmann et al. 2024). For a model predicting m_{jk} counts in energy channel j and phase bin k , with d_{jk} observed counts, the log likelihood is

$$\ln \mathcal{L} = \sum_{j,k} [d_{jk} \ln m_{jk} - m_{jk}], \quad (6)$$

where terms independent of the model have been omitted. As in [Dittmann et al. \(2024\)](#), a phase-independent background contribution is allowed independently in each NICER energy channel and is marginalized over, together with the overall rotational phase.

The spot normalization for each scaled synthetic dataset is fixed by the dataset-specific exposure described in Section 3.2; there is no additional free signal-normalization parameter. The background normalization is handled separately through the NICER background marginalization described above.

3.5. Two-Stage Analysis Procedure

In the first stage, we evaluate the likelihood with the 12 primary parameters held fixed at their input values (‘Best-fit value’ from Table 1). The 94 energy-channel background and the overall phase shift of the waveform are marginalized over analytically. We refer to this throughout as the input-parameter evaluation. It provides an initial diagnostic of how much the incorrect atmosphere degrades the fit before the primary parameters are allowed to adjust.

In the second stage, we sample over all 12 primary parameters by performing a full Bayesian posterior exploration using `pocoMC` (preconditioned Monte Carlo; [Karamanis et al. 2022a,b](#)), allowing all model parameters to vary within their priors. The sampler stochastically explores parameter space for each atmosphere assumption. From the resulting `pocoMC` run, we extract the posterior samples and Bayesian evidence. For the goodness-of-fit diagnostics reported below, we evaluate χ^2 at the maximum-likelihood sample, defined as the stored `pocoMC` sample with the largest $\ln \mathcal{L}$.

We chose `pocoMC` over the nested sampler+Markov Chain Monte Carlo hybrid strategy employed in [Miller et al. \(2019, 2021\)](#); [Dittmann et al. \(2024\)](#) and [Holt et al. \(2025\)](#) for several reasons. First, `pocoMC` is designed to handle the complex, multimodal posterior surfaces that arise in pulse-profile modeling, using a normalizing-flow-based preconditioner to accelerate convergence in high-dimensional parameter spaces. Second, `pocoMC` provides a direct estimate of the Bayesian evidence as a byproduct of the sampling, eliminating the need for a separate nested sampling run. Third, previous work has shown that some popular nested samplers, such as MultiNest, can underestimate posterior uncertainties when the number of live points is insufficient or the sampling efficiency parameter is too large ([Dittmann et al. 2024](#); [Salmi et al. 2024a](#); [Hoogkamer et al. 2025](#)), requiring subsequent MCMC continuation to broaden the credible intervals to their converged values. The `pocoMC` sampler avoids this two-step procedure by producing

converged posteriors directly and returning reliable evidence estimates. For this study, in which we analyze 24 independent configurations (4 test cases $\times$ 6 multipliers), the computational efficiency of a single-stage sampler is a practical advantage.

3.6. Fit Diagnostics

We assess the quality of each fit using two χ^2 statistics. The phase-channel χ^2 uses the standard Pearson χ^2 computed over all phase-channel bins, i.e.,

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

where d_i and m_i are the observed and model-predicted counts in bin i , and the sum runs over all 3008 phase-channel bins. This statistic is sensitive to localized discrepancies in individual bins.

The bolometric χ^2 is computed by first summing the counts over energy channels to produce a phase-only pulse profile, and then computing the χ^2 of this collapsed profile. The bolometric χ^2 checks whether the overall pulse shape is well reproduced when integrated over the photon energy.

Next, we compared the resulting value of χ^2 with the expected χ^2 distribution for the appropriate number of degrees of freedom in the input-parameter and sampled analyses, respectively. We report the probability of obtaining a χ^2 value at least as large as the observed one under the null hypothesis that the model is correct (i.e., the p -value). We note that acceptable χ^2 values may not guarantee that the assumed atmospheric composition is correct in all cases. A model that passes these checks can still return a biased radius posterior, as we demonstrate in §4. Nevertheless, χ^2 remains a valuable first diagnostic. Poor χ^2 values at the input-parameter stage correctly identify composition mismatch at high spot-to-background ratios before posterior exploration (see Tables 2 and 3). Acceptable χ^2 values after sampling confirm that a plausible fit has been found using the wrong atmosphere, and that this fit is worth scrutinizing further.

For the phase-channel χ^2 , the number of degrees of freedom is determined by the number of phase-channel bins minus the number of model parameters that are free to vary. In the input-parameter evaluation, all 12 primary parameters are held fixed, but the 94 energy-channel background levels and one overall rotational phase are marginalized analytically, giving $3008 - 94 - 1 = 2913$ degrees of freedom. In the fully sampled analysis, all 12 primary parameters are also free, giving $3008 - 94 - 1 - 5 = 2908$ degrees of freedom, where the effective number of waveform parameters influencing the

fit is approximately 5 (determined from synthetic data checks following the procedure described in [Dittmann et al. 2024](#)). For the bolometric χ^2 , we sum the counts over energy channels to produce 32 phase bins and adopt an effective number of parameters of approximately 5 plus the overall phase and a single background term, giving approximately $32 - 5 - 2 = 25$ degrees of freedom in the sampled case and $32 - 2 = 30$ in the input-parameter case.

3.7. Bayesian Evidence

To compare the hydrogen and helium atmosphere models for each synthetic dataset, we compute the Bayesian evidence $\mathcal{Z}$ for each model. The evidence is defined as the integral of the likelihood over the prior, i.e.,

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

where $\mathcal{L}(\theta)$ is the likelihood, $\pi(\theta)$ is the prior, and θ denotes the vector of model parameters. We report the log evidence ratio

$$\Delta \ln \mathcal{Z} = \ln \mathcal{Z}_{\text{correct}} - \ln \mathcal{Z}_{\text{incorrect}}, \quad (9)$$

where “correct” and “incorrect” models refer to the atmospheric composition used to generate the synthetic data and the mismatched atmospheric model, respectively. Thus, $\Delta \ln \mathcal{Z} > 0$ favors the correct atmosphere model, while $\Delta \ln \mathcal{Z} < 0$ favors the incorrect model.

We do not adopt fixed qualitative thresholds for interpreting the magnitude of $\Delta \ln \mathcal{Z}$. Such thresholds are arbitrary and do not account for uncertainty in the evidence estimates or for variation introduced by the independent Poisson realizations used for the synthetic datasets. We therefore report the evidence differences and interpret them comparatively across the models and synthetic experiments.

Unlike the χ^2 statistic, which evaluates the quality of the fit at the best-fit point, the evidence integrates over the entire posterior volume, which makes the evidence sensitive to model adequacy. A model that achieves an acceptable fit but requires a fine-tuned region of parameter space will be penalized relative to a model that fits comparably well over a broader parameter range. We use the evidence to check whether the correct atmospheric model can be identified even when both models yield statistically acceptable χ^2 values.

3.8. Radius Bias Diagnostic

To quantify the bias in the inferred radius introduced by assuming the incorrect atmospheric composition, we compute the percentile location of the injected radius

R_{inj} within the radius posterior obtained for the incorrect atmosphere fit. We adopt the convention that a percentile below the ~ 2.3 rd or above the ~ 97.7 th places R_{inj} outside the 2σ credible interval of the posterior, which indicates a potentially worrisome bias. Intermediate percentiles indicate a shift that, while potentially concerning, remains within 2σ and does not constitute a definitive detection of significant bias given the statistical uncertainties inherent in a single Poisson realization.

4. RESULTS

We present our results in three layers. We first describe the input-parameter diagnostics (§4.1), then the full posterior sampling results (§4.2–§4.5), and finally the radius posterior and Bayesian evidence behavior in both mismatch directions (§4.6–§4.7).

4.1. Input-parameter Incorrect Atmosphere Checks

Before allowing the model parameters to vary, we evaluate the likelihood at the 12-parameter solution under both the correct and incorrect atmosphere assumptions for each spot-to-background ratio multiplier. Tables 2 and 3 summarize these input-parameter diagnostics.

For the hydrogen-generated synthetic data (Table 2), the helium atmosphere evaluations produce increasingly poor phase-channel and bolometric χ^2 values at multipliers of $40\times$ and above. The bolometric χ^2 rises from ~ 31 at $1\times$ to 114 at $100\times$, while the correct hydrogen model maintains bolometric χ^2 values in the range ~ 19 – 47 . The phase-channel χ^2 shows a similar pattern, with the incorrect atmosphere values climbing above 3100 while the correct atmosphere values remain near the expected value for 2913 degrees of freedom.

For the helium-generated synthetic data (Table 3), the pattern is qualitatively similar. The hydrogen atmosphere evaluations produce elevated phase-channel and bolometric χ^2 values at multipliers of $60\times$ and above. The bolometric χ^2 under the incorrect hydrogen assumption reaches values up to ~ 88 , compared with ~ 17 – 36 for the correct helium model.

These results show that at the fixed reference configuration, the incorrect atmosphere can appear strongly disfavored, especially at higher spot-to-background ratios where the pulsed signal contributes a larger fraction of the total counts. However, as we show in the following subsections, these initial diagnostics do not predict the outcome of full posterior exploration.

4.2. Full Posterior Sampling Recovers Acceptable Fits

After full posterior exploration with `pocoMC`, the incorrect atmosphere model recovers statistically acceptable phase-channel and bolometric χ^2 values for all

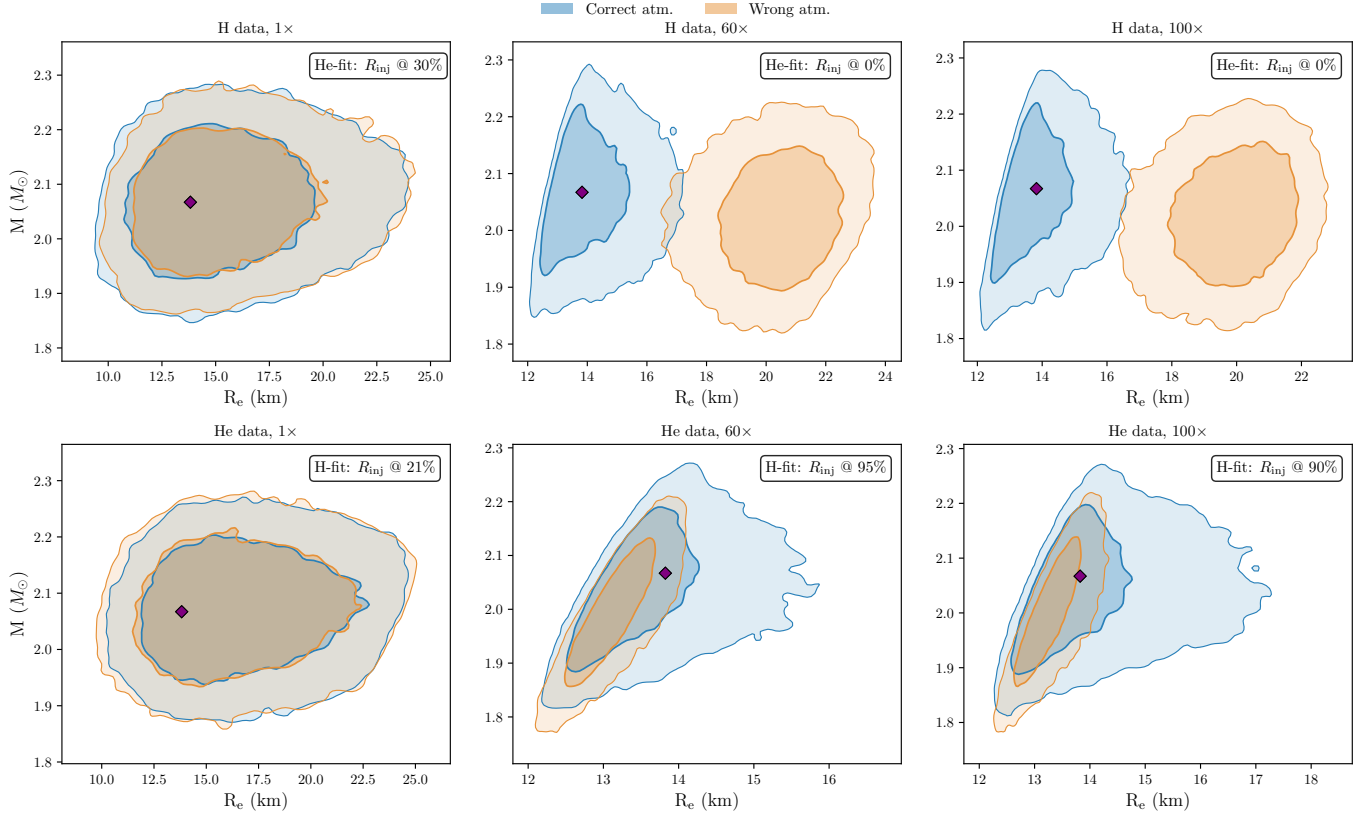


Figure 4. Mass-radius posterior distributions for selected multipliers ($1\times$, $60\times$, $100\times$). Blue contours show the correct atmosphere fit (1σ and 2σ credible regions), and orange contours show the incorrect atmosphere fit. The purple diamond marks the injected values ($R_{\text{inj}} = 13.82$ km, $M_{\text{inj}} = 2.067 M_{\odot}$). *Top row:* hydrogen-generated data, showing the correct H fit and incorrect He fit. The percentile of R_{inj} in the correct H-fit posterior is 33rd, 51st, and 58th at $1\times$, $60\times$, and $100\times$, respectively, confirming that the correct model recovers the injected radius without significant bias. The corresponding incorrect He-fit percentiles are 30th, 0.06th, and 0.07th, showing that the He-fit posterior is displaced to substantially larger radii at $60\times$ and $100\times$. *Bottom row:* helium-generated data, showing the correct He fit and incorrect H fit. The correct He-fit percentiles are 18th, 71st, and 58th, while the incorrect H-fit percentiles are 21st, 95th, and 90th. The H-fit posterior remains closer to the truth for all three multipliers. The percentile of R_{inj} in the incorrect model posterior is annotated in each panel.

checked spot-to-background ratios, in both mismatch directions. Despite the poor input-parameter evaluations documented in §4.1, the sampler identifies parameter combinations that substantially improve the fit to the incorrect model.

Table 4 summarizes the `pocoMC` results for the hydrogen-generated synthetic data. For the hydrogen-generated data, the helium-fit phase-channel χ^2 probabilities range from 6.4×10^{-2} to 8.5×10^{-1} , and the bolometric χ^2 probabilities range from 1.5×10^{-2} to 7.2×10^{-1} , which all remain above statistical rejection thresholds. The correct hydrogen-fit probabilities are comparable.

Table 5 presents the corresponding results for the helium-generated synthetic data. The hydrogen-fit phase-channel χ^2 probabilities range from 4.2×10^{-2} to 5.3×10^{-1} , and the bolometric χ^2 probabilities range from 2.6×10^{-1} to 9.8×10^{-1} . All fits are statistically

acceptable. The correct helium-fit probabilities are similarly acceptable for all multipliers. An incorrect atmosphere model that appears strongly disfavored at the fixed reference configuration can become statistically acceptable once the full parameter space is explored.

4.3. Reliability of the Correct-Atmosphere Radius Posteriors

The acceptable χ^2 values exhibited in §4.2 show that the sampler produces statistically good fits, but they do not by themselves establish that the radius posteriors are converged. Before turning to the atmospheric mismatch results in §4.4 and §4.5, we confirm that the radius posteriors obtained with the correct atmosphere are converged. We verify this directly through convergence checks in which we double the effective-sample-size of the `pocoMC` runs and continue them with `emcee`. The radius posteriors are stable under both checks for all

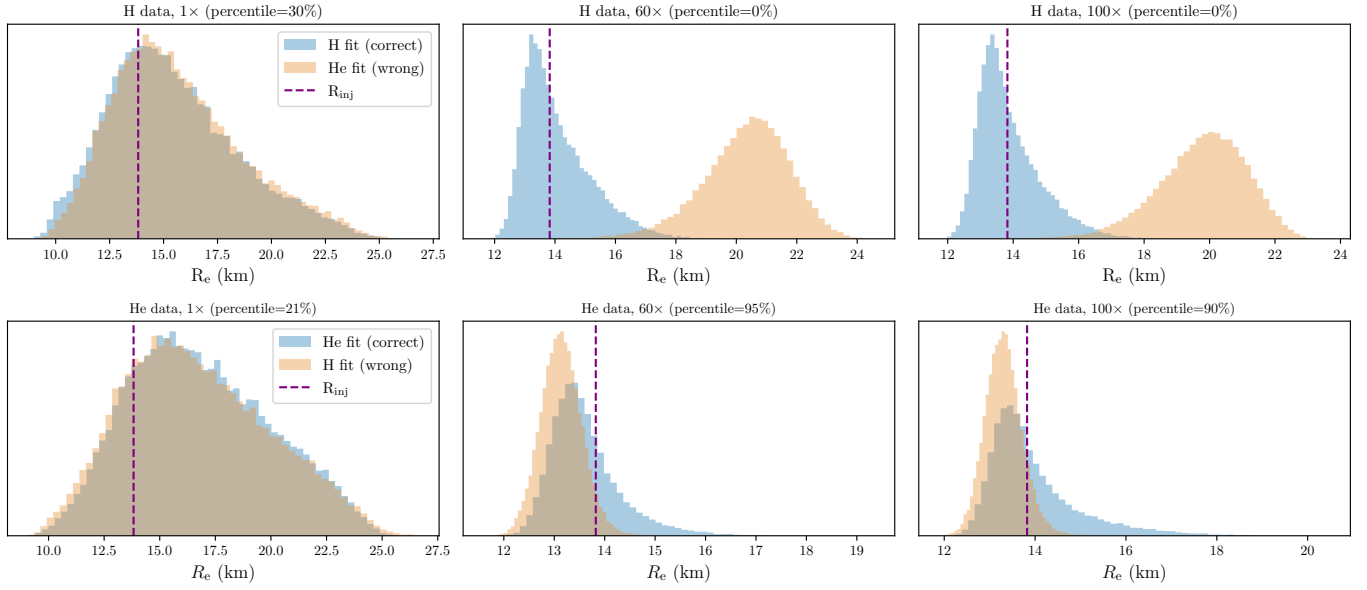


Figure 5. One-dimensional radius posteriors for selected multipliers ($1\times$, $60\times$, $100\times$). The blue histograms show the correct atmosphere fit, and the orange histograms show the incorrect atmosphere fit. The dashed purple line marks the injected radius. *Top row:* hydrogen-generated data. The percentile of R_{inj} in the correct H-fit posterior is 33rd, 51st, and 58th, whereas in the incorrect He-fit posterior it is 30th, 0.06th, and 0.07th. At $60\times$ and $100\times$, the He-fit (incorrect) posterior is shifted well above R_{inj} . *Bottom row:* helium-generated data. The correct He-fit percentiles are 18th, 71st, and 58th, and the incorrect H-fit percentiles are 21st, 95th, and 90th. The bias is smaller when incorrectly fitting H to He data than when incorrectly fitting He to H data.

Table 2. Input-parameter evaluation for hydrogen-generated synthetic data

Scale	R_{synth}	f_{pulse}	H atm. (correct)		He atm. (incorrect)	
			χ^2_{pc} (Prob.)	χ^2_{b} (Prob.)	χ^2_{pc} (Prob.)	χ^2_{b} (Prob.)
$1\times$	0.064	0.034	2878 (.673)	32.3 (.354)	2877 (.679)	31.5 (.390)
$20\times$	1.282	0.334	2824 (.878)	31.1 (.410)	2943 (.345)	48.5 (.018)
$40\times$	2.565	0.435	2913 (.494)	22.3 (.843)	3141 (.002)	60.8 (.001)
$60\times$	3.847	0.484	3025 (.072)	47.2 (.024)	3178 (.000)	64.2 (.000)
$80\times$	5.130	0.513	2817 (.896)	19.1 (.938)	3097 (.009)	75.7 (.000)
$100\times$	6.412	0.532	2996 (.139)	41.0 (.086)	3242 (.000)	114 (.000)

NOTE—Input-parameter χ^2 values and probabilities at the reference 12-parameter solution. “Scale” is the spot-to-background ratio multiplier m (see §3.2). R_{synth} is the corresponding target spot-to-background ratio. f_{pulse} is the fractional modulation of the expected bolometric synthetic profile, $f_{\text{pulse}} \equiv (C_{\text{max}} - C_{\text{min}})/(C_{\text{max}} + C_{\text{min}})$, and is included to show pulse strength with scaled configuration. χ^2_{pc} is the phase-channel χ^2 and χ^2_{b} is the bolometric χ^2 . The values in parentheses are p -values (the probability of obtaining a χ^2 at least this large if the model is correct). Degrees of freedom: 2913 (phase-channel) and 30 (bolometric). The hydrogen model yields acceptable probabilities at all multipliers. The helium model is strongly disfavored for multipliers $\geq 40\times$.

Table 3. Input-parameter evaluation for helium-generated synthetic data

Scale	R_{synth}	f_{pulse}	He atm. (correct)		H atm. (incorrect)	
			χ_{pc}^2 (Prob.)	χ_{b}^2 (Prob.)	χ_{pc}^2 (Prob.)	χ_{b}^2 (Prob.)
1×	0.061	0.033	2950 (.312)	17.2 (.970)	2951 (.308)	18.4 (.952)
20×	1.224	0.325	2890 (.617)	25.6 (.696)	2947 (.326)	41.4 (.080)
40×	2.447	0.423	2910 (.513)	30.5 (.438)	3049 (.039)	65.6 (.000)
60×	3.671	0.471	2996 (.138)	36.5 (.194)	3196 (.000)	77.8 (.000)
80×	4.894	0.499	2896 (.583)	26.9 (.630)	3123 (.004)	87.9 (.000)
100×	6.118	0.517	2906 (.535)	25.1 (.721)	3103 (.007)	77.0 (.000)

NOTE—As in Table 2, but for helium-generated synthetic data. The hydrogen model becomes strongly disfavored for multipliers $\geq 40\times$.

Table 4. Maximum-likelihood diagnostics for hydrogen-generated synthetic data

R_{synth}	H atm. (correct)		He atm. (incorrect)		$\Delta \ln \mathcal{Z}$
	χ^2_{pc} (Prob.)	χ^2_{b} (Prob.)	χ^2_{pc} (Prob.)	χ^2_{b} (Prob.)	
0.064	2876 (.659)	30.6 (.203)	2875 (.664)	30.6 (.202)	-0.9
1.282	2812 (.898)	28.5 (.287)	2829 (.851)	30.7 (.199)	+10.8
2.565	2900 (.536)	20.4 (.727)	2924 (.415)	20.5 (.721)	+14.2
3.847	3011 (.089)	36.2 (.068)	3025 (.064)	42.7 (.015)	+9.7
5.130	2812 (.898)	18.9 (.803)	2852 (.769)	24.1 (.516)	+23.1
6.412	2983 (.162)	32.7 (.138)	2988 (.147)	37.7 (.050)	+4.1

NOTE— χ^2 values and probabilities evaluated at the maximum-likelihood `pocoMC` sample for the hydrogen-generated data. Degrees of freedom: 2908 (χ^2_{pc}) and 25 (χ^2_{b}). For $\Delta \ln \mathcal{Z} = \ln \mathcal{Z}_{\text{H}} - \ln \mathcal{Z}_{\text{He}}$, the positive values favor the hydrogen model. Both models yield acceptable χ^2 values at every multiplier, yet the evidence consistently identifies the correct model in all spot-dominated multipliers.

Table 5. Maximum-likelihood diagnostics for helium-generated synthetic data

R_{synth}	He atm. (correct)		H atm. (incorrect)		$\Delta \ln \mathcal{Z}$
	χ^2_{pc} (Prob.)	χ^2_{b} (Prob.)	χ^2_{pc} (Prob.)	χ^2_{b} (Prob.)	
0.061	2941 (.329)	12.8 (.979)	2942 (.325)	12.8 (.979)	+0.6
1.224	2881 (.634)	20.7 (.710)	2902 (.528)	21.3 (.676)	+12.5
2.447	2906 (.505)	27.5 (.331)	2940 (.334)	27.7 (.323)	+19.8
3.671	2986 (.152)	27.8 (.318)	3041 (.042)	29.2 (.256)	+31.7
4.894	2886 (.609)	23.9 (.523)	2929 (.388)	24.2 (.506)	+26.6
6.118	2888 (.600)	16.8 (.888)	2940 (.333)	17.6 (.859)	+30.0

NOTE— χ^2 values and probabilities evaluated at the maximum-likelihood `pocoMC` sample for the helium-generated data. For $\Delta \ln \mathcal{Z} = \ln \mathcal{Z}_{\text{He}} - \ln \mathcal{Z}_{\text{H}}$, the positive values favor the helium model. The evidence consistently and strongly favors the correct model ($\Delta \ln \mathcal{Z} \sim +12.5$ to $+31.7$).

four check cases (see Appendix A). Since the correct-atmosphere radius posteriors are converged, any systematic displacement of the injected radius from the bulk of the posterior in the incorrect atmosphere fits is attributable to the atmospheric composition mismatch rather than to a sampling issue.

4.4. Hydrogen Data Fit with Helium: The Dangerous Case

When hydrogen-generated synthetic data (H, d) are fit with a helium atmosphere model (He, m), the helium model’s best-fit parameters found by the sampler while exploring parameter space can achieve acceptable residuals as quantified by χ^2 , but in the process, the radius posterior becomes significantly biased.

The bias is apparent in Figure 4 (top row), which compares the correct (blue) and incorrect (orange) atmosphere M – R posteriors for selected multipliers. We show $1\times$ (the baseline ratio) to illustrate the background-dominated regime where mismatch is undetectable, $60\times$ as a representative case in which the $He, m \rightarrow H, d$ bias becomes severe (incorrect model percentile $< 0.1\%$), and $100\times$ as a further spot-dominated case that confirms the persistence of the effect. At $1\times$, the two posteriors overlap substantially, and the injected parameters lie comfortably within both. At $60\times$ and $100\times$, the helium-fit posterior has shifted to significantly larger radii, with the injected value falling outside the 2σ contour. The corresponding 1D radius posteriors are shown in the top row of Figure 5. The incorrect model distribution is displaced well above the injected radius.

Figure 6 shows the M – R posteriors for all six multipliers. The helium fit posterior is shifted to larger radii relative to the hydrogen fit posterior, and the injected radius falls deep into the low-percentile tail of the incorrect model distribution for all spot-dominated cases.

This is the more dangerous mismatch direction in the present configuration. The incorrect model yields a statistically acceptable fit while returning a significantly biased radius.

4.5. Helium Data Fit with Hydrogen: Disfavored but Less Damaging

When helium-generated synthetic data (He, d) are fit with a hydrogen atmosphere model (H, m), the hydrogen model also achieves acceptable χ^2 values after posterior sampling. However, unlike the $He, m \rightarrow H, d$ case, the injected radius remains within 2σ of the incorrect model posterior for the checked multipliers (see the bottom rows of Figures 4 and 5). The incorrect model percentile values range from $\sim 21\%$ at $1\times$ to $\sim 97\%$ at $80\times$, but none crosses the 2σ threshold into the extreme tails.

Figure 7 shows the M – R posteriors for all six multipliers in this direction. The incorrect hydrogen-model contours (orange) are displaced to smaller radii relative to the correct helium-model contours (blue), but the displacement is modest compared to the reverse direction, and the injected parameters remain within the 2σ credible region at every multiplier.

Therefore, although the incorrect hydrogen model is statistically disfavored by the Bayesian evidence (Tables 5), it does not produce the same level of radius bias observed in the $He, m \rightarrow H, d$ direction.

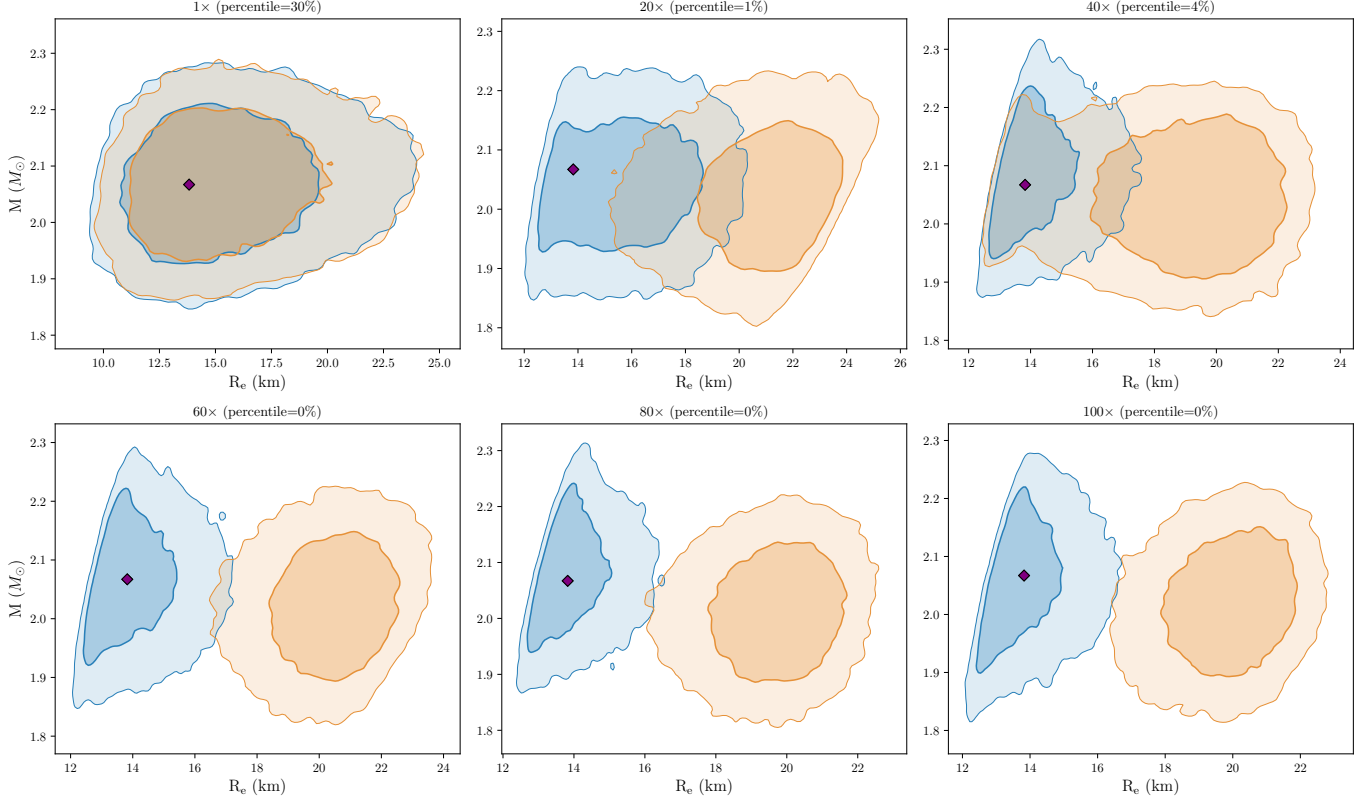


Figure 6. Mass-radius posteriors for all six multipliers with hydrogen-generated synthetic data. The blue contours show the correct H fit, and the orange contours show the incorrect He fit. The purple diamond marks the injected values. The percentile of R_{inj} within the correct H-fit posterior is 33rd, 19th, 45th, 51st, 51st, and 58th for $1\times$ through $100\times$, respectively, confirming that the correct model is unbiased. The corresponding incorrect He-fit percentiles are 30th, 0.8th, 4th, 0.06th, 0.08th, and 0.07th, showing that the He-fit posterior is consistently shifted to larger radii, with R_{inj} falling below the 2σ threshold in most spot-dominated cases.

4.6. Direct Comparison of the Two Mismatch Directions

The correct model percentiles suggest that the correct atmosphere recovers the injected radius without significant bias. The asymmetry between the two incorrect model directions is clear. Fitting hydrogen-generated data with a helium model produces incorrect model percentiles at or below 6% (and frequently below 1%) for all spot-dominated multipliers (Figure 6), while fitting helium-generated data with a hydrogen model produces incorrect model percentiles that remain between $\sim 21\%$ and $\sim 97\%$, consistently within 2σ (Figure 7).

The two directions share one important property. In both cases, the incorrect atmosphere model achieves statistically acceptable χ^2 values after posterior sampling. The difference lies in the consequences of that acceptable fit. The $He, m \rightarrow H, d$ direction produces large radius bias, while the $H, m \rightarrow He, d$ direction does not, in the checked J0740-like configuration.

4.7. Bayesian Evidence as a Complement to χ^2

The $\Delta \ln \mathcal{Z}$ values reported in Tables 4 and 5 show that the Bayesian evidence favors the correct atmospheric model for all spot-dominated multipliers in both mismatch directions. At the realistic $1\times$ baseline, the evidence difference is small: $\Delta \ln \mathcal{Z} = -0.9$ for hydrogen-generated data fit with a helium model and $\Delta \ln \mathcal{Z} = +0.6$ for helium-generated data fit with a hydrogen model. In the former case, despite the slight preference for the incorrect atmosphere, the inferred radius is not significantly biased. Thus, in the J0740-like baseline regime, the two atmosphere models are effectively difficult to distinguish and the small evidence difference does not translate into a consequential radius bias.

For the hydrogen-generated data, $\Delta \ln \mathcal{Z}$ ranges from approximately +4 to +23 in favor of the hydrogen model. For the helium-generated data, $\Delta \ln \mathcal{Z}$ ranges from approximately +12.5 to +31.7 in favor of the helium model.

While both the correct and incorrect atmosphere models can achieve acceptable phase-channel and bolomet-

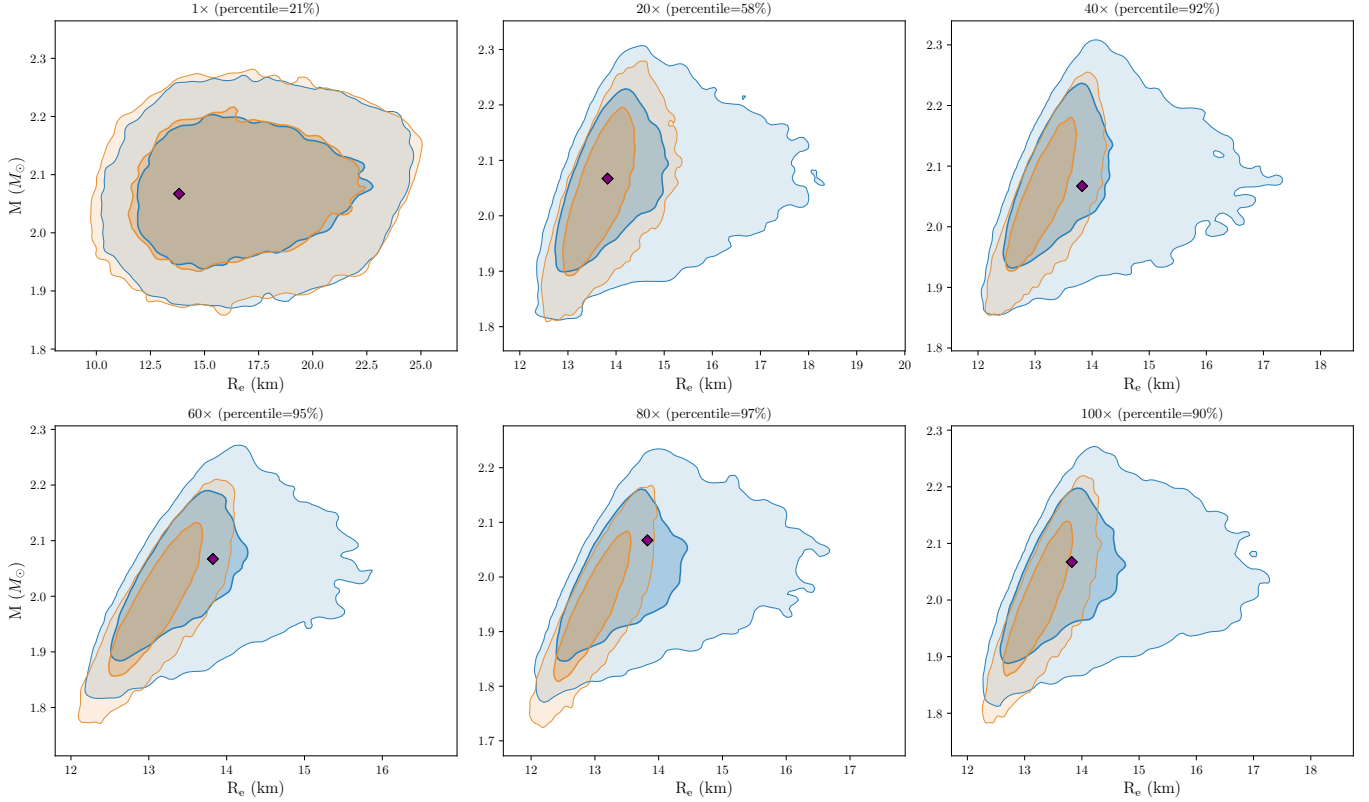


Figure 7. Mass–radius posteriors for all six multipliers with helium-generated synthetic data. The blue contours show the correct He fit, and the orange contours show the incorrect H fit. The purple diamond marks the injected values. The percentile of R_{inj} within the correct He-fit posterior is 18th, 43rd, 68th, 71st, 68th, and 58th for $1\times$ through $100\times$, respectively. The corresponding incorrect H-fit percentiles are 21st, 58th, 92nd, 95th, 97th, and 90th. The H-fit posterior is displaced to smaller radii relative to the He-fit posterior, but the injected radius remains within 2σ for all checked multipliers.

ric χ^2 values, the evidence retains additional sensitivity to the atmospheric composition mismatch. The evidence integrates over the full posterior volume, penalizing models that require fine-tuned parameter combinations to achieve their best fit. The contrast is apparent in Tables 4 and 5. For the spot-dominated cases, rows where the correct and incorrect atmosphere χ^2 probabilities are nearly identical are nevertheless accompanied by positive $\Delta \ln \mathcal{Z}$ values of approximately +4 or larger.

The non-monotonic behavior of $\Delta \ln \mathcal{Z}$ with increasing multiplier is a consequence of the independent Poisson realizations used to generate each synthetic dataset. Because each multiplier corresponds to a separate realization, the specific noise pattern varies from dataset to dataset, producing fluctuations in the evidence difference that do not reflect a smooth trend.

5. DISCUSSION

The results presented in §4 confirm that atmospheric composition mismatch can produce biased radius estimates while remaining hidden inside a statistically acceptable fit. The pulse-profile model has substantial freedom. When the assumed atmosphere is incorrect,

the posterior exploration can compensate for the incorrect beaming by adjusting other parameters, yielding an acceptable χ^2 at the cost of a biased radius posterior. The subsections below discuss the physical interpretation of the observed asymmetry, the insufficiency of χ^2 alone, and the broader implications for NICER-like analyses.

5.1. Why the $He, m \rightarrow H, d$ Direction Is More Dangerous

Our results show that fitting hydrogen-generated data with a helium model produces a larger radius bias in a J0740-like configuration. The origin of this asymmetry may lie in the high compactness of PSR J0740+6620. The amount of light bending grows rapidly as a star approaches the photon sphere but changes only slowly in the Newtonian limit, meaning the bending is far more sensitive to compactness for compact stars than for extended ones. At the compactness of our reference configuration, decreasing the radius therefore increases the light bending sharply, whereas increasing the radius by the same amount reduces it only slightly. Since stronger light bending keeps the hot spot in view over a larger

fraction of the rotation, the fractional modulation of the waveform varies steeply with radius below the injected value and only weakly above it. A model that needs to reduce the modulation can do so with a small decrease in radius, whereas a model that needs to increase the modulation must move to a much larger radius.

A helium beaming pattern is broader and more isotropic than the hydrogen pattern that generated the data (see §2), and produces too little modulation at the injected radius. To recover the modulation present in the hydrogen-generated waveform, the sampler must move toward larger radii, i.e., toward lower compactness, which is the direction in which the light bending responds weakly. A large change in R_e is therefore needed, and the helium-fit posteriors are displaced to substantially larger radii, with the injected radius falling below the 2σ threshold in most spot-dominated cases.

5.2. *Why the $H, m \rightarrow He, d$ Direction Is Less Damaging*

In the reverse direction, the fitting model starts from a beaming pattern that is more narrowly peaked than the true helium beaming, and therefore produces too much modulation at the injected radius. Correcting this requires moving toward a higher compactness. A comparatively small adjustment in R_e suffices, and the incorrect hydrogen model can accommodate the mismatch without being driven into a strongly biased region of the parameter space.

The asymmetry is reflected in the widths of the radius posteriors. At the $180\times$ multiplier (Table 6), the mismatched helium fit to hydrogen data has a $\pm 1\sigma$ radius width of 3.46 km, roughly five times the 0.71 km width of the mismatched hydrogen fit to helium data, and it develops a pronounced tail toward large radii. A broad, asymmetric posterior extending to large radii is expected when the waveform constrains compactness only weakly on the low-compactness side.

The geometric parameters adjust in a manner consistent with this picture. In the hydrogen-fit-to-helium runs the non-radius parameters move only modestly. The observer angle is recovered well ($\theta_{obs} = 1.530$ and 1.524 at $20\times$ and $60\times$ v.s. the injected 1.561), both spots migrate to the equator ($\theta_{c,1/2} = 1.579/1.77$ at $20\times$, $1.508/1.483$ at $60\times$, near $\pi/2$ and near the line of sight), and the spot sizes shrink ($\Delta\theta_{1\&2} \approx 0.071$ vs. the injected 0.087 and 0.066). These adjustments broaden the effective angular beaming, and the incorrect hydrogen model recovers acceptable waveforms without forcing the radius into a comparably biased region. The radius remains within 2σ of the incorrect hydrogen-model

posterior for all cases, even though the Bayesian evidence consistently favors the correct helium model.

We stress that this interpretation of the asymmetry is suggested by the results presented in §4, not derived as a general concept. Whether the asymmetry persists for other pulsar geometries, count levels, compactness configurations, and background conditions remains an open question.

5.3. *Why χ^2 Should Be Supplemented by the Bayesian Evidence*

A central finding of this work is that a model can pass both phase-channel and bolometric χ^2 checks while returning a misleading radius posterior. This is a concern because, in a real analysis, the actual atmospheric composition is unknown, meaning that given the statistically acceptable fit from χ^2 alone, there is no way to tell whether the acceptable fit corresponds to the correct composition or to an incorrect one that has absorbed the mismatch through shifting other parameters. This is especially apparent in the $He, m \rightarrow H, d$ direction, where the χ^2 values for the helium model are acceptable for every case we examined despite the radius being biased beyond 2σ in most spot-dominated cases.

This result is particularly important for analyses that aim to provide accurate equation-of-state constraints. If χ^2 acceptability were treated as sufficient validation of the atmospheric composition, one could unknowingly adopt a model that produces a significantly biased radius. Holt et al. (2025) showed that in joint NICER/XMM-Newton analyses, even a factor-of-five underestimate of the XMM-Newton background shifts the radius posterior by only $\sim 1\sigma$, and the Bayesian evidence identified the preferred background model. The present results show that atmospheric composition mismatch can produce a larger bias while similarly evading detection by χ^2 diagnostics.

5.4. *Why the Bayesian Evidence Is Valuable*

The Bayesian evidence retains sensitivity to atmospheric composition mismatch even when both atmosphere models produce acceptable χ^2 values. In both mismatch directions and for all spot-dominated waveforms, the evidence favors the correct atmosphere, with $\Delta \ln \mathcal{Z}$ values ranging from $\sim +4$ to $\sim +31.7$.

This sensitivity arises because the evidence integrates over the full posterior volume, rather than evaluating the fit quality only at the best-fit point. A model that achieves an acceptable best-fit χ^2 but requires a fine-tuned region of parameter space to do so will be penalized relative to a model that fits comparably well over a broader parameter range. The evidence therefore provides information about model adequacy and should be

included in the diagnostic framework for pulse-profile analyses.

5.5. *Why the Asymmetry May Not Be Universal*

The asymmetry we have identified between the two mismatch directions should not be treated as a general result. Our analysis has checked a specific configuration: a J0740+6620-like two-spot geometry, J0740+6620-like total counts ($\sim 5.5 \times 10^5$), a range of spot-to-background ratios from the realistic baseline to $100\times$ the baseline, and the particular fully ionized hydrogen and helium atmosphere models used in this work. The degree and direction of the asymmetry may change for other pulsar geometries, different count levels, different background conditions, or if the beaming patterns of the two atmosphere models become more or less distinguishable in another regime.

The mechanism we have described above does suggest one respect in which our results may generalize. If the asymmetry arises because the light bending responds steeply to increases in compactness and weakly to decreases, then it is controlled by the compactness of the star rather than by the atmospheric properties alone. We therefore expect the two mismatch directions to become more symmetric for lower-mass, and thus less compact, neutron stars, and the asymmetry to be most pronounced for the most massive sources, of which PSR J0740+6620 is a representative example.

In particular, at the actual count level and spot-to-background ratio estimated for PSR J0740+6620 (the $1\times$ multiplier), fitting hydrogen-generated data with a helium model does not produce a bias in the inferred radius, and the two atmosphere models are effectively indistinguishable. This baseline result is consistent with the PSR J0740+6620 analysis of [Salmi et al. \(2023\)](#), who found no significant radius change among the atmosphere models they examined. Our higher-multiplier results identify a controlled regime of larger pulsed fraction in which the atmospheric assumption becomes more consequential. The bias becomes pronounced only at higher spot-to-background ratios, where the pulsed signal contributes a larger fraction of the total counts.

5.6. *Broader Implications for NICER-like Analyses*

Previous work has shown that PPM radius estimates are robust against incorrect assumptions about hot spot shapes and temperature distributions ([Lo et al. 2013](#); [Miller & Lamb 2015, 2016](#)) and against substantial errors in the assumed unmodulated background ([Holt et al. 2025](#)), provided the fit is statistically acceptable. Unlike the systematic errors studied previously, atmospheric composition mismatch can hide inside an acceptable fit while still biasing the inferred radius. The

Bayesian evidence provides a more reliable diagnostic than χ^2 for identifying this type of model inadequacy.

For pulsars with substantially higher signal-to-noise ratios than PSR J0740+6620, such as PSR J0437–4715, the pulsed signal constitutes a much larger number of total counts and fraction of those counts. Our spot-to-background scaling experiment shows that as this fraction increases, the bias introduced by the incorrect atmosphere grows and the Bayesian evidence increasingly favors the correct model. This suggests that atmospheric composition mismatch is more consequential, but also detectable, for brighter sources, and that analyses of high-signal-to-noise data should routinely compare atmosphere models using the Bayesian evidence.

In the $1\times$ spot-to-background case, our analysis shows no significant bias in either mismatch direction, because the pulsed signal is too weak relative to the background to distinguish the beaming patterns. The bias becomes apparent only at higher spot-to-background ratios. Increasing the exposure time at the same spot-to-background ratio would reduce the statistical uncertainties without changing the fractional contribution of the pulsed signal, so the beaming mismatch would remain difficult to detect. A significant bias at the $1\times$ spot-to-background ratio would require either substantially more data than is currently feasible or a qualitative change in the analysis, such as the inclusion of imaging data that constrain the background independently.

6. CONCLUSIONS

In this work we have investigated whether hydrogen and helium atmospheric composition mismatch can bias the inferred radius of a PSR J0740+6620-like pulsar while remaining hidden inside a statistically acceptable fit. We generated synthetic NICER-like pulse-profile data assuming fully ionized hydrogen and helium atmospheres and fit each synthetic dataset with both atmosphere models, varying the spot-to-background ratio while holding the total expected counts fixed at $\sim 5.5 \times 10^5$.

Our central finding is that χ^2 acceptability is not a sufficient safeguard against atmospheric composition bias. In both mismatch directions, posterior sampling allows the incorrect atmosphere model to recover apparently good phase-channel and bolometric fits, yet the resulting radius posteriors can still be significantly displaced from the injected value. This is qualitatively different from the systematics studied in prior work ([Lo et al. 2013](#); [Miller & Lamb 2016](#); [Holt et al. 2025](#)), where statistically acceptable fits were reliable indicators of unbiased radii in the checked configurations.

Fitting hydrogen-generated data with a helium atmosphere model can increase the inferred radius to more than 2σ above its true value for strongly spot-dominated waveforms. Fitting helium-generated data with a hydrogen atmosphere model produces smaller shifts that remain within 2σ for all checked waveforms. The broader beaming produced by a helium atmosphere requires more aggressive parameter compensation than the narrower beaming produced by a hydrogen atmosphere, but should not be treated as a universal result, as it may depend on the assumed geometry, count level, and atmosphere models.

At the actual count level and spot-to-background ratio of PSR J0740+6620, neither mismatch direction produces a detectable bias. However, for sources with higher pulsed fractions, such as PSR J0437–4715, the present results suggest that atmospheric composition mismatch could be appreciable.

The Bayesian evidence consistently distinguishes the correct atmospheric model from the incorrect one for all spot-dominated waveforms. This makes the Bayesian evidence an essential complement to goodness-of-fit statistics in pulse-profile analyses.

ACKNOWLEDGMENTS

This work was supported in part by NASA ADAP grant 80NSSC21K0649, and was supported by NASA

through the NICER mission and the Astrophysics Explorers Program. The authors acknowledge the University of Maryland supercomputing resources (<http://hpcc.umd.edu>) that were made available for conducting the research reported in this paper. This work also utilized computational resources provided by the NASA Center for Climate Simulation (NCCS) Discover supercomputing facility. We are grateful to the NICER team for their ongoing efforts in instrument operation, data calibration, and analysis support. This work was performed in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452. AJD was supported by NASA through the NASA Hubble Fellowship grant No. HST-HF2-51553.001, awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555.

Software: pocoMC (Karamanis et al. 2022b), emcee (Foreman-Mackey et al. 2013), Python and NumPy (Oliphant 2007), Matplotlib (Hunter 2007), Cython (Behnel et al. 2011), schwimmbad (Price-Whelan & Foreman-Mackey 2017), HEASoft Nasa High Energy Astrophysics Science Archive Research Center (Heasarc) (2014)

APPENDIX

A. SAMPLER CONFIGURATION AND CONVERGENCE OF THE RADIUS POSTERIOR

We performed all posterior sampling with pocoMC (Karamanis et al. 2022a,b), which advances an ensemble of particles through a sequence of tempered distributions bridging the prior and the posterior, using a normalizing flow to precondition the target at each step. The two parameters that principally control sampling fidelity are the effective sample size N_{eff} and the number of active particles N_{act} . N_{eff} sets the target effective sample size of the importance-weighted particle ensemble maintained at each iteration. A larger N_{eff} therefore resolves the intermediate distributions more finely and reduces the variance of both the posterior and the evidence estimate, at proportionally greater cost. N_{act} sets how many of those particles are advanced by MCMC at each iteration, and hence the fraction of the ensemble refreshed per step. Particles are advanced in the space of the normalizing flow using the t-preconditioned Crank–Nicolson sampler, with the flow retrained at each itera-

tion. N_{total} is the number of samples in the final posterior, and N_{evidence} is the number of importance samples drawn from the preconditioned proposal for the final evidence estimate. We set both to 65,536 for every run.

Our production runs used $N_{\text{eff}} = 32,256$ and $N_{\text{act}} = 4,032$. The convergence check described in §A.1 doubles N_{eff} to 64,512 while holding N_{act} and all other settings fixed.

As an independent check on the pocoMC results, we continued each run with emcee (Foreman-Mackey et al. 2013). Unlike the nested-sampling scheme, whose accuracy depends on the fidelity of a learned or assumed proposal distribution and on the adequacy of a live-point set, emcee is a Markov chain Monte Carlo code constructed to satisfy detailed balance with respect to the target density. Its stationary distribution is therefore the posterior by construction, and given sufficient run time it is guaranteed to produce a representative sample of the posterior. Its convergence in twelve dimensions is substantially slower than that of a preconditioned sampler, which is why we use it as a validation tool rather

than as our primary sampler. However, for the same reason, a stable **emcee** continuation of a **pocoMC** run is strong evidence that the **pocoMC** posterior is not an artifact of the preconditioner or of the particle ensemble.

A.1. Convergence results

The radius R_e is the parameter of interest. We therefore verify directly that its posterior is converged. We do this in two ways. First, we repeat the **pocoMC** inference at another effective-sample-size setting, $N_{\text{eff}} = 64,512$. Doubling the original $N_{\text{eff}} = 32,256$ doubles the number of importance samples retained at each **pocoMC** iteration, so a sampling deficiency would appear as a disagreement between the two settings. We continue each **pocoMC** run with **emcee** and examine the “running credible intervals” of R_e . At each **emcee** step we compute the median and the central 1σ , 2σ , and 3σ intervals of R_e from all samples accumulated up to that step. If the chain has converged, these bands are flat. We perform both checks for all four atmosphere configurations (H data/H fit, H data/He fit, He data/He fit, He data/H fit) at the most demanding $180\times$ multiplier, where the pulsed signal is strongest.

Table 6 compares the radius posterior at the two N_{eff} settings. For all four configurations, the radius median shifts by at most 0.064 km when N_{eff} is doubled, and the $\pm 1\sigma$ and $\pm 2\sigma$ interval widths change by no more than a few percent, with no systematic direction. The quantile of the injected radius within the posterior is unchanged between the two settings in all four cases. Figure 8 shows the joint posteriors at $N_{\text{eff}} = 32,256$ and the $N_{\text{eff}} = 64,512$ with contours that overlie the correct and incorrect models. The strong radius bias in the mismatched fits, the asymmetric tail toward large radii in the H data/He fit case, and the compressed radius posterior in the He data/H fit case are all reproduced at both settings. The **emcee** running credible intervals, shown in Figures 9 and 10, remain stable over the full chain for every configuration. We conclude that the radius posteriors are converged at our adopted sampler settings.

Table 6. Convergence of the R_e posterior in km after doubling N_{eff} .

Configuration	N_{eff}	Median	$\pm 1\sigma$ width	$\pm 2\sigma$ width	Δ median
H data, H fit	32,256	13.532	1.010	2.293	+0.000
	64,512	13.532	1.036	2.352	
H data, He fit	32,256	18.080	3.461	7.225	+0.064
	64,512	18.144	3.464	7.287	
He data, He fit	32,256	13.666	1.733	3.737	+0.006
	64,512	13.672	1.752	3.841	
He data, H fit	32,256	13.112	0.712	1.454	−0.003
	64,512	13.109	0.721	1.480	

NOTE—The radius median and the $\pm 1\sigma$ and $\pm 2\sigma$ interval widths at each N_{eff} setting are shown in each column. Medians and widths are in km. Δ median is $\text{med}_{64} - \text{med}_{32}$. In every configuration, the median shift is far smaller than the $\pm 1\sigma$ width, indicating that the radius posterior has converged.

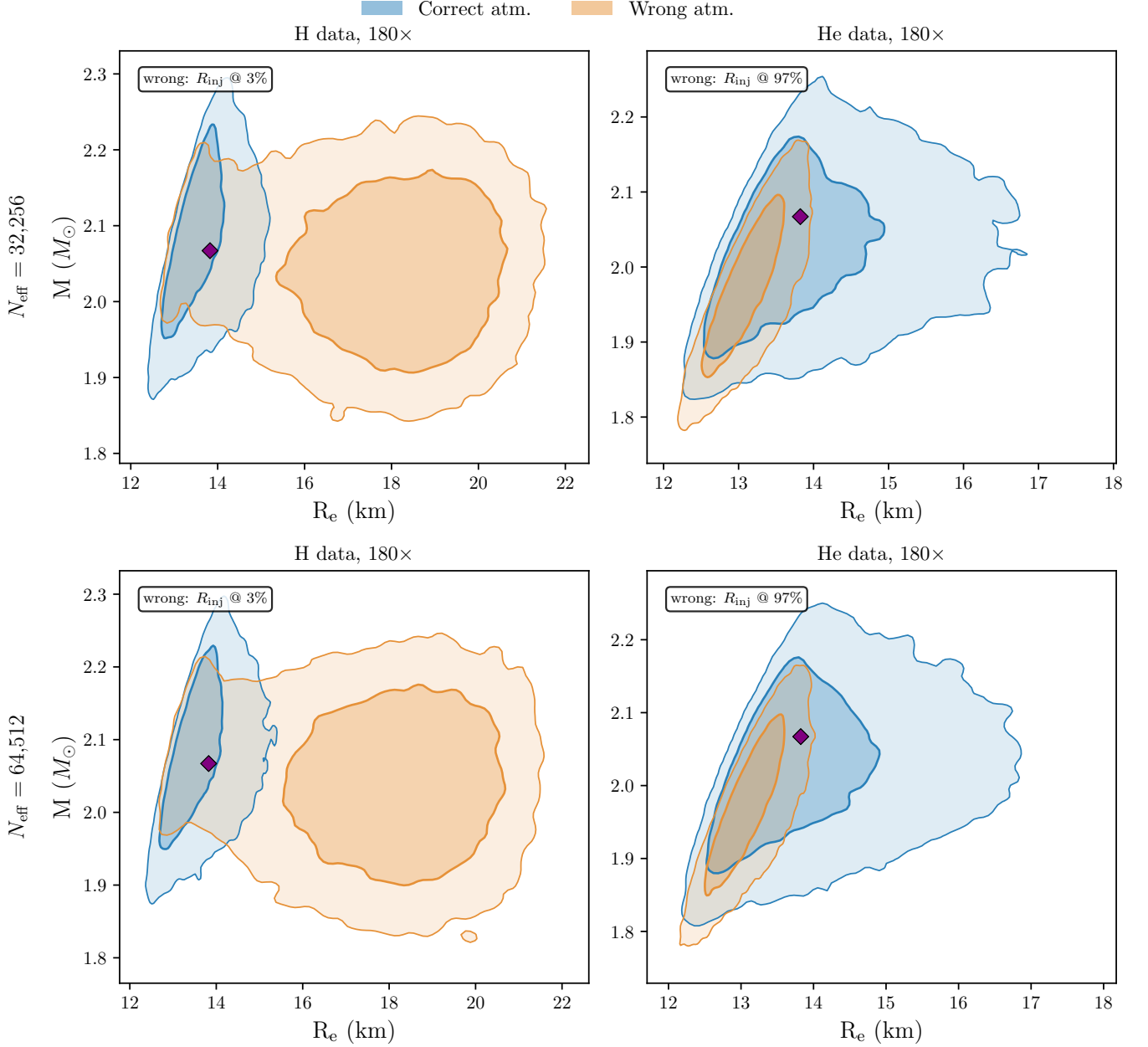


Figure 8. Two-dimensional R_e - M posteriors for $N_{\text{eff}} = 32,256$ (top row) and $N_{\text{eff}} = 64,512$ (bottom row) at the $180\times$ multiplier. *Left:* hydrogen-generated data, with the correct H fit and the mismatched He fit overlaid. *Right:* helium-generated data, with the correct He fit and the mismatched H fit overlaid. In each panel, the correct-atmosphere fit is shown in blue and the mismatched fit in orange; inner and outer contours are the 68% and 95% credible regions, and the purple diamond marks the injected value $(R_{\text{inj}}, M_{\text{inj}}) = (13.82 \text{ km}, 2.07 M_\odot)$. The radius bias of the mismatched fits, the large-radius tail in the H data/He fit case, and the compressed radius posterior in the He data/H fit case are reproduced identically at both N_{eff} settings.

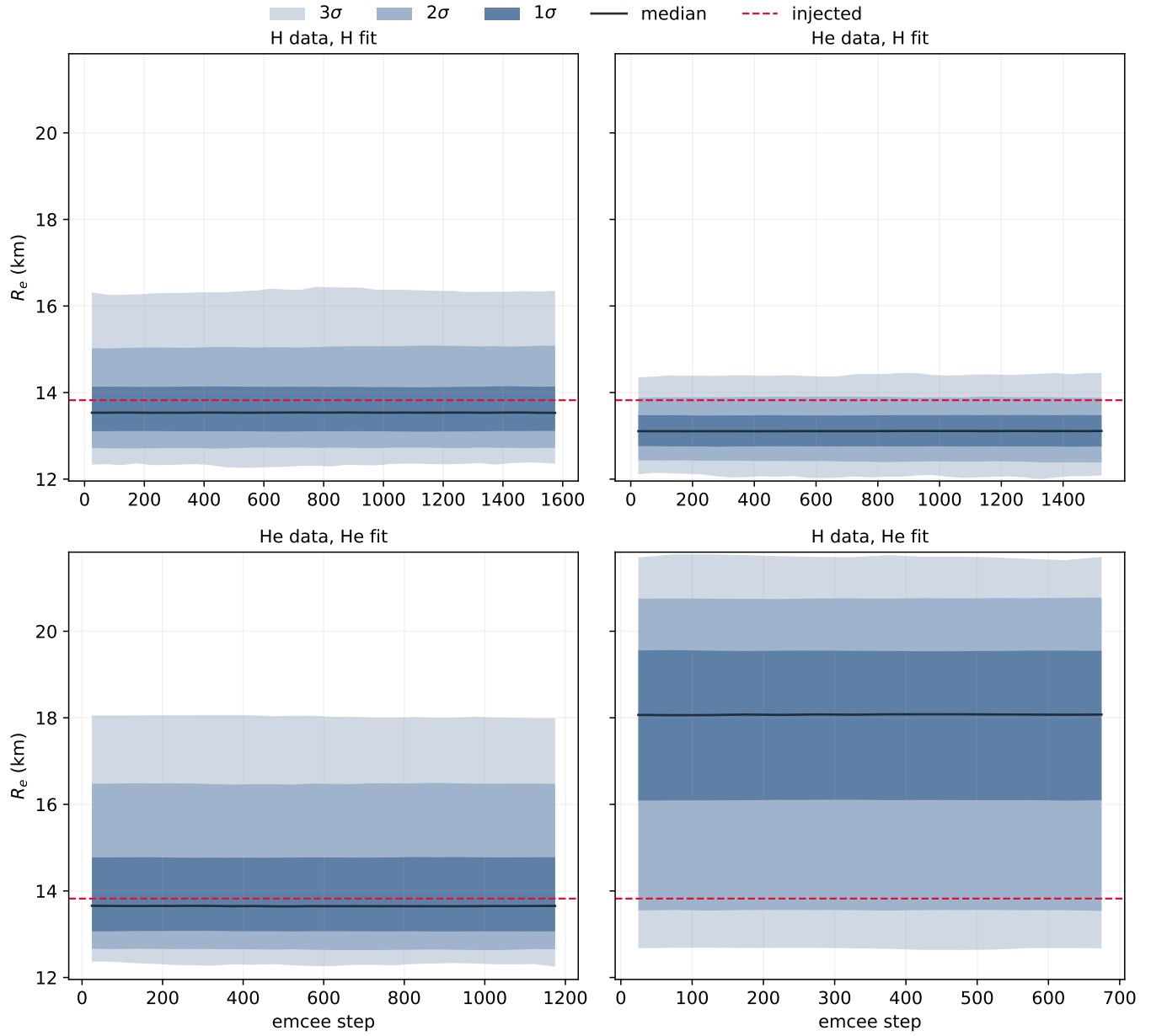


Figure 9. Running credible intervals of the radius R_e from `pocoMC` $N_{\text{eff}} = 32,256$ setting, using `emcee`, for the $180\times$ multiplier for the four atmosphere configurations. The shaded bands show the running 1σ , 2σ , and 3σ credible intervals, the solid line the running median, and the dashed line the injected radius. The radius posterior is stable over the full chain in every configuration, confirming convergence of the `pocoMC` results.

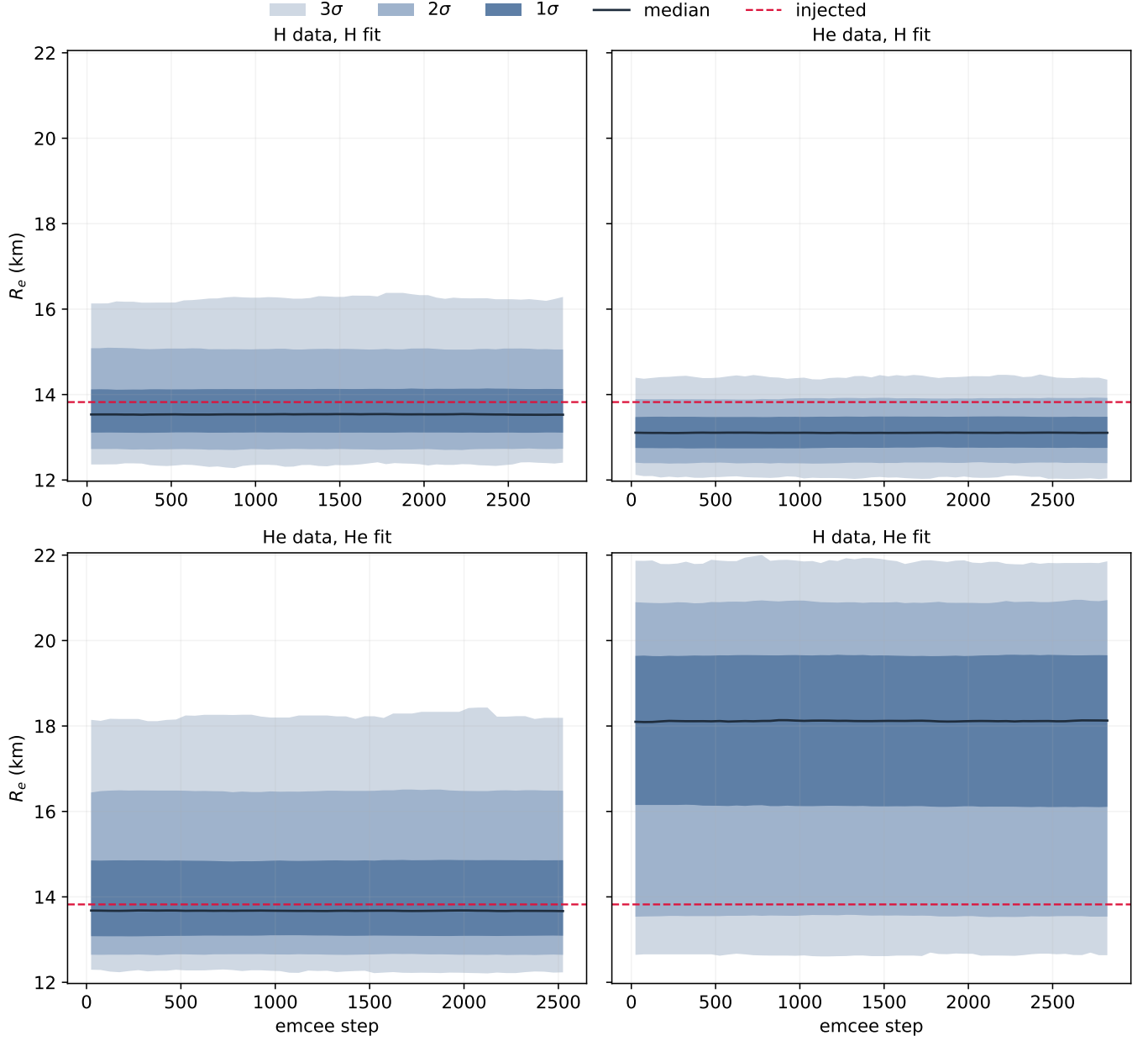


Figure 10. Running credible intervals of the radius R_e from `pocoMC` $N_{\text{eff}} = 64,512$ setting, using `emcee`, for the $180\times$ multiplier for the four atmosphere configurations. The shaded bands show the running 1σ , 2σ , and 3σ credible intervals, the solid line the running median, and the dashed line the injected radius. The radius posterior is stable over the full chain in every configuration, confirming convergence of the `pocoMC` results.

REFERENCES

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, *PhRvL*, 119, 161101, doi: [10.1103/PhysRevLett.119.161101](https://doi.org/10.1103/PhysRevLett.119.161101)
- . 2018, *PhRvL*, 121, 161101, doi: [10.1103/PhysRevLett.121.161101](https://doi.org/10.1103/PhysRevLett.121.161101)
- Alcock, C., & Illarionov, A. 1980, *ApJ*, 235, 534, doi: [10.1086/157656](https://doi.org/10.1086/157656)
- Behnel, S., Bradshaw, R., Citro, C., et al. 2011, *Computing in Science and Engineering*, 13, 31, doi: [10.1109/MCSE.2010.118](https://doi.org/10.1109/MCSE.2010.118)
- Bogdanov, S., Heinke, C. O., Özel, F., & Güver, T. 2016, *ApJ*, 831, 184, doi: [10.3847/0004-637X/831/2/184](https://doi.org/10.3847/0004-637X/831/2/184)
- Bogdanov, S., Lamb, F. K., Mahmoodifar, S., et al. 2019, *ApJL*, 887, L26, doi: [10.3847/2041-8213/ab5968](https://doi.org/10.3847/2041-8213/ab5968)
- Bogdanov, S., Dittmann, A. J., Ho, W. C. G., et al. 2021, *ApJL*, 914, L15, doi: [10.3847/2041-8213/abfb79](https://doi.org/10.3847/2041-8213/abfb79)
- Catuneanu, A., Heinke, C. O., Sivakoff, G. R., Ho, W. C. G., & Servillat, M. 2013, *ApJ*, 764, 145, doi: [10.1088/0004-637X/764/2/145](https://doi.org/10.1088/0004-637X/764/2/145)
- Choudhury, D., Salmi, T., Vinciguerra, S., et al. 2024, *ApJL*, 971, L20, doi: [10.3847/2041-8213/ad5a6f](https://doi.org/10.3847/2041-8213/ad5a6f)
- De, S., Finstad, D., Lattimer, J. M., et al. 2018, *PhRvL*, 121, 091102, doi: [10.1103/PhysRevLett.121.091102](https://doi.org/10.1103/PhysRevLett.121.091102)
- Demorest, P. B., Pennucci, T., Ransom, S. M., Roberts, M. S. E., & Hessels, J. W. T. 2010, *Nature*, 467, 1081, doi: [10.1038/nature09466](https://doi.org/10.1038/nature09466)
- Dittmann, A. J., Miller, M. C., Lamb, F. K., et al. 2024, *ApJ*, 974, 295, doi: [10.3847/1538-4357/ad5f1e](https://doi.org/10.3847/1538-4357/ad5f1e)
- Fonseca, E., Cromartie, H. T., Pennucci, T. T., et al. 2021, *ApJL*, 915, L12, doi: [10.3847/2041-8213/ac03b8](https://doi.org/10.3847/2041-8213/ac03b8)
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306, doi: [10.1086/670067](https://doi.org/10.1086/670067)
- Gendreau, K. C., Arzoumanian, Z., & Okajima, T. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8443, *Space Telescopes and Instrumentation 2012: Ultraviolet to Gamma Ray*, ed. T. Takahashi, S. S. Murray, & J.-W. A. den Herder, 844313, doi: [10.1117/12.926396](https://doi.org/10.1117/12.926396)
- González-Caniulef, D., Guillot, S., Stammel, P., et al. 2026, *arXiv e-prints*, arXiv:2607.03721, doi: [10.48550/arXiv.2607.03721](https://doi.org/10.48550/arXiv.2607.03721)
- Guillot, S., Rutledge, R. E., & Brown, E. F. 2011, *ApJ*, 732, 88, doi: [10.1088/0004-637X/732/2/88](https://doi.org/10.1088/0004-637X/732/2/88)
- Guillot, S., Servillat, M., Webb, N. A., & Rutledge, R. E. 2013, *ApJ*, 772, 7, doi: [10.1088/0004-637X/772/1/7](https://doi.org/10.1088/0004-637X/772/1/7)
- Heinke, C. O., Cohn, H. N., Lugger, P. M., et al. 2014, *MNRAS*, 444, 443, doi: [10.1093/mnras/stu1449](https://doi.org/10.1093/mnras/stu1449)
- Ho, W. C. G., & Lai, D. 2001, *MNRAS*, 327, 1081, doi: [10.1046/j.1365-8711.2001.04801.x](https://doi.org/10.1046/j.1365-8711.2001.04801.x)
- Holt, I. M., Miller, M. C., Dittmann, A. J., & Lamb, F. K. 2025, *arXiv e-prints*, arXiv:2511.16759, doi: [10.48550/arXiv.2511.16759](https://doi.org/10.48550/arXiv.2511.16759)
- Hoogkamer, M., Kini, Y., Salmi, T., Watts, A. L., & Buchner, J. 2025, *PhRvD*, 112, 023008, doi: [10.1103/cp8c-2nbk](https://doi.org/10.1103/cp8c-2nbk)
- Hunter, J. D. 2007, *Computing in Science and Engineering*, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Karamanis, M., Beutler, F., Peacock, J. A., Nabergoj, D., & Seljak, U. 2022a, *MNRAS*, 516, 1644, doi: [10.1093/mnras/stac2272](https://doi.org/10.1093/mnras/stac2272)
- Karamanis, M., Nabergoj, D., Beutler, F., Peacock, J., & Seljak, U. 2022b, *The Journal of Open Source Software*, 7, 4634, doi: [10.21105/joss.04634](https://doi.org/10.21105/joss.04634)
- Kini, Y., Mauviard, L., Salmi, T., et al. 2026, *ApJ*, 1005, 201, doi: [10.3847/1538-4357/ae733e](https://doi.org/10.3847/1538-4357/ae733e)
- Klochkov, D., Suleimanov, V., Pühlhofer, G., et al. 2015, *A&A*, 573, A53, doi: [10.1051/0004-6361/201424683](https://doi.org/10.1051/0004-6361/201424683)
- Lattimer, J. M., & Steiner, A. W. 2014, *ApJ*, 784, 123, doi: [10.1088/0004-637X/784/2/123](https://doi.org/10.1088/0004-637X/784/2/123)
- Lo, K. H., Miller, M. C., Bhattacharyya, S., & Lamb, F. K. 2013, *ApJ*, 776, 19, doi: [10.1088/0004-637X/776/1/19](https://doi.org/10.1088/0004-637X/776/1/19)
- Mauviard, L., Guillot, S., Salmi, T., et al. 2025, *ApJ*, 995, 60, doi: [10.3847/1538-4357/ae145d](https://doi.org/10.3847/1538-4357/ae145d)
- Miller, M. C., & Lamb, F. K. 2015, *ApJ*, 808, 31, doi: [10.1088/0004-637X/808/1/31](https://doi.org/10.1088/0004-637X/808/1/31)
- . 2016, *European Physical Journal A*, 52, 63, doi: [10.1140/epja/i2016-16063-8](https://doi.org/10.1140/epja/i2016-16063-8)
- Miller, M. C., Lamb, F. K., Dittmann, A. J., et al. 2019, *ApJL*, 887, L24, doi: [10.3847/2041-8213/ab50c5](https://doi.org/10.3847/2041-8213/ab50c5)
- . 2021, *ApJL*, 918, L28, doi: [10.3847/2041-8213/ac089b](https://doi.org/10.3847/2041-8213/ac089b)
- Miller, M. C., Dittmann, A. J., Holt, I. M., et al. 2026, *ApJL*, 1000, L48, doi: [10.3847/2041-8213/ae5057](https://doi.org/10.3847/2041-8213/ae5057)
- Nasa High Energy Astrophysics Science Archive Research Center (Heasarc). 2014, *HEASoft: Unified Release of FTOOLS and XANADU*, *Astrophysics Source Code Library*, record ascl:1408.004. <http://ascl.net/1408.004>
- Oliphant, T. E. 2007, *Computing in Science and Engineering*, 9, 10, doi: [10.1109/MCSE.2007.58](https://doi.org/10.1109/MCSE.2007.58)
- Pechenick, K. R., Ftaclas, C., & Cohen, J. M. 1983, *ApJ*, 274, 846, doi: [10.1086/161498](https://doi.org/10.1086/161498)
- Poutanen, J., & Gierliński, M. 2003, *MNRAS*, 343, 1301, doi: [10.1046/j.1365-8711.2003.06773.x](https://doi.org/10.1046/j.1365-8711.2003.06773.x)
- Price-Whelan, A. M., & Foreman-Mackey, D. 2017, *The Journal of Open Source Software*, 2, 357, doi: [10.21105/joss.00357](https://doi.org/10.21105/joss.00357)
- Qi, L., Zheng, S., Zhang, J., et al. 2025, *ApJ*, 981, 99, doi: [10.3847/1538-4357/adb42f](https://doi.org/10.3847/1538-4357/adb42f)

- Riley, T. E., Watts, A. L., Bogdanov, S., et al. 2019, *ApJL*, 887, L21, doi: [10.3847/2041-8213/ab481c](https://doi.org/10.3847/2041-8213/ab481c)
- Riley, T. E., Watts, A. L., Ray, P. S., et al. 2021, *ApJL*, 918, L27, doi: [10.3847/2041-8213/ac0a81](https://doi.org/10.3847/2041-8213/ac0a81)
- Romani, R. W. 1987, *ApJ*, 313, 718, doi: [10.1086/165010](https://doi.org/10.1086/165010)
- Romani, R. W., Kandel, D., Filippenko, A. V., Brink, T. G., & Zheng, W. 2022, *ApJL*, 934, L17, doi: [10.3847/2041-8213/ac8007](https://doi.org/10.3847/2041-8213/ac8007)
- Rybicki, G. B., & Lightman, A. P. 1979, *Radiative processes in astrophysics* (Wiley)
- Saffer, A., Fonseca, E., Ransom, S., et al. 2024, *arXiv e-prints*, arXiv:2412.02850, doi: [10.48550/arXiv.2412.02850](https://doi.org/10.48550/arXiv.2412.02850)
- Salmi, T., Vinciguerra, S., Choudhury, D., et al. 2022, *ApJ*, 941, 150, doi: [10.3847/1538-4357/ac983d](https://doi.org/10.3847/1538-4357/ac983d)
- . 2023, *ApJ*, 956, 138, doi: [10.3847/1538-4357/acf49d](https://doi.org/10.3847/1538-4357/acf49d)
- Salmi, T., Choudhury, D., Kini, Y., et al. 2024a, *ApJ*, 974, 294, doi: [10.3847/1538-4357/ad5f1f](https://doi.org/10.3847/1538-4357/ad5f1f)
- Salmi, T., Deneva, J. S., Ray, P. S., et al. 2024b, *ApJ*, 976, 58, doi: [10.3847/1538-4357/ad81d2](https://doi.org/10.3847/1538-4357/ad81d2)
- Servillat, M., Heinke, C. O., Ho, W. C. G., et al. 2012, *MNRAS*, 423, 1556, doi: [10.1111/j.1365-2966.2012.20976.x](https://doi.org/10.1111/j.1365-2966.2012.20976.x)
- Vinciguerra, S., Salmi, T., Watts, A. L., et al. 2024, *ApJ*, 961, 62, doi: [10.3847/1538-4357/acfb83](https://doi.org/10.3847/1538-4357/acfb83)
- Wolff, M. T., Guillot, S., Bogdanov, S., et al. 2021, *ApJL*, 918, L26, doi: [10.3847/2041-8213/ac158e](https://doi.org/10.3847/2041-8213/ac158e)