

PAPER

Deducing neutron star equation of state parameters directly from telescope spectra with uncertainty-aware machine learning

To cite this article: Delaney Farrell *et al* JCAP02(2023)016

View the [article online](#) for updates and enhancements.

Deducing neutron star equation of state parameters directly from telescope spectra with uncertainty-aware machine learning

Delaney Farrell,^a Pierre Baldi,^b Jordan Ott,^b Aishik Ghosh,^{c,d}
Andrew W. Steiner,^{e,f} Atharva Kavitkar,^g Lee Lindblom,^h
Daniel Whiteson^c and Fridolin Weber^{a,h}

^aDepartment of Physics, San Diego State University,
San Diego, CA 92115, U.S.A.

^bDepartment of Computer Science, University of California Irvine,
Irvine, CA 92697, U.S.A.

^cDepartment of Physics and Astronomy, University of California Irvine,
Irvine, CA 92697, U.S.A.

^dPhysics Division, Lawrence Berkeley National Laboratory,
Berkeley, CA 94720, U.S.A.

^eDepartment of Physics and Astronomy, University of Tennessee,
Knoxville, TN 37996, U.S.A.

^fPhysics Division, Oak Ridge National Laboratory,
Oak Ridge, TN 37831, U.S.A.

^gDepartment of Computer Science,
TU Kaiserslautern, Germany

^hCenter for Astrophysics and Space Sciences, University of California at San Diego,
San Diego, CA 92093, U.S.A.

E-mail: dfarrell@sdsu.edu, pfbaldi@uci.edu, jott1@uci.edu,
aishikghoshoffice@gmail.com, awsteiner@utk.edu, a_kavitkar19@cs.uni-kl.de,
llindblom@ucsd.edu, daniel@uci.edu, fweber@mail.sdsu.edu

Received September 8, 2022

Accepted January 12, 2023

Published February 8, 2023

Abstract. Neutron stars provide a unique laboratory for studying matter at extreme pressures and densities. While there is no direct way to explore their interior structure, X-rays emitted from these stars can indirectly provide clues to the equation of state (EOS) of the superdense nuclear matter through the inference of the star's mass and radius. However,

inference of EOS directly from a star's X-ray spectra is extremely challenging and is complicated by systematic uncertainties. The current state of the art is to use simulation-based likelihoods in a piece-wise method which relies on certain theoretical assumptions and simplifications about the uncertainties. It first infers the star's mass and radius to reduce the dimensionality of the problem, and from those quantities infer the EOS. We demonstrate a series of enhancements to the state of the art, in terms of realistic uncertainty quantification and a path towards circumventing the need for theoretical assumptions to infer physical properties with machine learning. We also demonstrate novel inference of the EOS directly from the high-dimensional spectra of observed stars, avoiding the intermediate mass-radius step. Our network is conditioned on the sources of uncertainty of each star, allowing for natural and complete propagation of uncertainties to the EOS.

Keywords: X-ray telescopes, Machine learning, neutron stars, X-rays

ArXiv ePrint: [2209.02817](https://arxiv.org/abs/2209.02817)

Contents

1	Introduction	1
2	Background	3
2.1	Equation of state for dense matter	3
2.2	X-ray spectroscopy for neutron stars	5
3	Machine learning	6
4	Training samples	6
4.1	Equation of state	7
4.2	Modeling X-ray spectra	9
4.3	Nuisance parameters	10
5	Inference of EOS from mass and radius	11
5.1	Mass and radius inference by XSPEC	13
5.2	Inference of EOS from mass and radius	13
5.2.1	Neural network regression	14
5.2.2	Polynomial regression	15
5.2.3	Estimation of uncertainty	16
5.2.4	Performance	16
6	Inference of mass and radius from spectra	16
6.1	MR_Net method	18
6.2	MR_Net performance in mass, radius	19
6.3	Network uncertainty on mass and radius	20
6.4	Revisiting EOS regression	21
7	Inference of EOS from spectra	21
7.1	Architecture	24
7.2	Training	26
7.3	Results	27
8	Discussion	28
9	Conclusions	29

1 Introduction

Neutron stars are the densest stellar objects, providing a unique laboratory for studying matter in physical conditions that cannot be replicated on Earth and are only found in these neutron-packed remnants of massive stars ($\gtrsim 8 M_{\odot}$). Insights about the forms of matter which emerge under these extreme conditions can improve our understanding of two of the least well-understood fundamental forces, quantum chromodynamics and gravity.

The neutron-rich matter within the inner regions of a neutron star can reach supranuclear densities of 10^{15} g/cm³, potentially leading to transitions to stable non-nucleonic states of strange matter in the form of hyperons [1–3], deconfined quark matter made of up, down, and strange quarks [4, 5], color superconducting phases [6, 7], or Bose-Einstein condensates

made of negatively charged pions or K^- mesons [8–10]. The structure and behavior of matter at such extreme densities are one of the great mysteries in modern science, prompting decades of theoretical and experimental research into the interior composition of a neutron star.

The nature of matter within a neutron star is compactly summarized by its equation of state (EOS), the relationship between pressure P and energy density ϵ , which show starkly different behaviors under the various states of strange matter hypotheses described above. This relationship is determined by the microphysical interactions between various particles within the star, relative abundances of different particle species, as well as the star’s temperature. Understanding the EOS of supranuclear density matter has been of interest to the nuclear and astrophysics communities for decades, resulting in many proposed phenomenological models for the equation of state of neutron star matter. Variations in models arise from a lack of precise knowledge of nuclear interactions between particles at such extreme conditions, as well as the wide range of densities and isospin asymmetries that are thought to exist within neutron stars [11].

While the internal pressure and density cannot be directly observed, the EOS of a star in static gravitational equilibrium determines stellar properties such as its mass and radius, which in turn determine observables such as the stellar X-ray spectrum. Conversely, the stellar spectra can be used to infer the masses and radii, which in principle allow inference of the EOS [12–17], though inversion of this second step is numerically very difficult. Additional challenges are due to the small number of neutron star observations, $\mathcal{O}(10)$, and the significant uncertainty of individual measurements. It is therefore vital that as much information as possible is extracted from each star, and that the uncertainties be propagated accurately, to provide the most complete information possible about the EOS.

At the same time, there has been a dramatic burst of progress in artificial intelligence, specifically deep learning [22], a modern re-branding of neural networks. This progress has led to breakthroughs not only in traditional areas such as natural language processing and computer vision but also in the natural sciences, including particle physics, often increasing the statistical power of difficult-to-collect data [23] while allowing robust handling of uncertainties [24]. Where earlier neural networks were limited in size, computing progress especially in the form of Graphical Processing Units (GPUs), has enabled the deployment of larger and deeper networks that can handle more complex and higher-dimensional data [25, 26], allowing direct analysis of data without requiring dimensional reduction, or other preprocessing steps, that can often sacrifice useful information. The full power of these techniques has not yet been brought to bear on many astrophysical tasks.

In the context of the inference of neutron star EOS, recent work by Fujimoto et al. [18, 19] demonstrated the ability of deep networks to regress the EOS directly from a set of stellar mass-radius pairs, without the need to extract the functional relationship between mass and radius. Their analysis used a toy model to describe the uncertainties in mass and radius, assuming uncorrelated Gaussian errors randomly drawn from ad-hoc priors. Real measurements, of course, do not often obey these simplifying assumptions, and show complex correlations between mass and radius [27]. Related work [20] has demonstrated similar regression, again assuming Gaussian uncertainty on mass and radius values, but with clever efforts to reduce dependence on EOS parameterization. An alternative approach [21] uses both neural networks and support vector machines to regress the EOS from stellar radii and tidal deformations.

More realistic characterization of the uncertainties in the mass-radius plane can be extracted using the state-of-the-art tool XSPEC [28], which assumes a theoretical model for

the star and telescope response, allowing for explicit calculation of the likelihood of telescope spectra for various mass and radius values. The likelihood can be used in the standard way to extract best-estimates and uncertainty contours of any shape in the mass-radius plane. However, these complex mass-radius likelihoods cannot be trivially incorporated into the existing EOS inference schemes, motivating the simplifying assumptions of uncorrelation normal distributions which can be described by two width values. An additional concern is that XSPEC’s contours rely on the simplifying assumptions of the theoretical model.

What has received less attention in the literature are likelihood-free methods to infer the EOS directly from the telescope spectra, without the intermediate stepping stone of the mass-radius determination and the challenges of its representation. This would allow for the full propagation of realistic uncertainties and the relaxation of assumptions about the theoretical model.

In this paper, we present a technique of EOS inference which allows for the full propagation of the uncertainties in the X-ray spectra, without making simplifying assumptions about the shape of the contours in the mass-radius plane. We proceed in three steps, beginning from an approach similar to the state of the art but with realistic uncertainty propagation, and moving towards end-to-end inference. In the first step, our neural network model infers the neutron star EOS from a set of stellar masses and radii extracted from XSPEC, but rather than making simplifying assumptions or extracting uncertainty contours from XSPEC, we vary the assumed nuisance parameters (NPs) which are the source of the uncertainty to produce new best-estimate mass-radius points. The EOS inference can then be performed on many sets, each corresponding to varied NP values, producing a variation in the inferred which represents the propagated uncertainty. In the second step, we investigate a more flexible method of inferring the mass and radius that does not explicitly rely on XSPEC’s specific theoretical model. We introduce a network capable of directly analyzing high-dimensional neutron star spectra, performing inference of stellar mass and radii from telescope spectra, a demonstration of the impressive capacity of modern deep networks. Finally, we perform a first-of-its-kind inference of the EOS parameters directly from a multi-star *set* of stellar spectra, without requiring the intermediate step of collapsing the information into mass and radius; see figure 1. In both cases, we allow for full propagation of uncertainties by conditioning the networks on the stellar nuisance parameters. As this is — to our knowledge — the first attempt at full propagation of these uncertainties for this task, there are no direct benchmarks in prior work. Instead, we show comparisons between our three methods and visualize the impact on the EOS inference of variation of the nuisance parameters for a fixed x-ray spectrum.

The paper is organized as follows. In section 2, we provide background on the physics of the connection between the nuclear equation of state and the stellar observations. Section 3 describes the fundamentals of the machine learning concepts on which our studies rely. Details of the samples of simulated data are given in section 4. Section 5 demonstrates inference of the EOS parameters from mass and radius, while section 6 describes how mass and radius parameters can be inferred directly from stellar spectra, and section 7 shows end-to-end inference of EOS parameters from a set of neutron star spectra.

2 Background

2.1 Equation of state for dense matter

Neutron star interiors present a unique opportunity to study matter under conditions beyond the reach of terrestrial laboratories: matter that is extremely high in density, relatively cold in

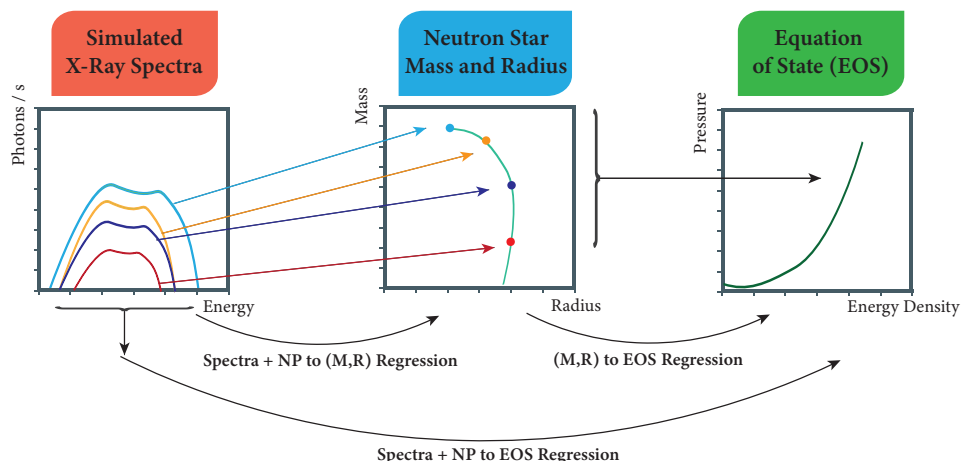


Figure 1. Overview of the regression task, which involves either inferring stellar summary quantities such as mass and radius, which can then be used to deduce the equation of state as in earlier work [18–21] or inference of EOS directly from stellar spectra, as is demonstrated in this study.

temperature, and isospin-asymmetric [29]. Perhaps the closest experimental constraints have come from ultrarelativistic heavy-ion collisions (as conducted at the RHIC [30], the LHC [31], and FAIR [32]), which probe the nature of hot, symmetric nuclear matter. Extremely neutron-rich matter is more recently probed in studies of neutron skin or giant monopole and dipole resonances, but these studies are limited in nature [33]. Unfortunately, these experiments currently lack the temperature and density constraints present within the core of a neutron star — meaning the dense matter encountered within a neutron star cannot yet be replicated by experiment.

While properties of superdense matter cannot be derived directly from experiments, certain general principles from general relativity and quantum chromodynamics (QCD) guide the theoretical investigation into neutron star matter. Neutron star structure is controlled by the long-range gravitational force [34, 35] which holds the star together and short-range strong interactions between nucleons and nuclei which provide the pressure that prevents the star from collapsing. At low nuclear densities (below nuclear saturation), effective field theories based on QCD provide a systematic basis for nuclear forces, which offers good constraints on two-nucleon interactions [35–37]. At higher densities, the QCD framework predicts that baryonic matter (where quarks are confined within hadrons) will experience a phase transition to quark-gluon plasma (QGP), where quarks and gluons are freed from hadronic boundaries [37]. Other stable states of non-nucleonic matter may also occur, like the formation of hyperons, color superconducting phases of quark matter, or Bose-Einstein condensates of different mesons [8, 10]. Theoretical uncertainties have resulted in a wide range of proposed phenomenological models for the EOS of neutron star matter, which can then be tested by experiment or observation.

The EOS of neutron star matter is intrinsically linked to macroscopic characteristics like gravitational mass M and radius R through the general relativistic stellar structure equation known as the Tolman-Oppenheimer-Volkoff (TOV) equation [38, 39]. This equation assumes

the object is spherically symmetric, non-rotating, and non-magnetic. The TOV equation is given by (assuming geometrized units where $G = c = 1$):

$$\frac{dP}{dr} = -\frac{(\epsilon + P)(m + 4\pi r^3 P)}{r^2 \left(1 - \frac{2m}{r}\right)} \quad (2.1)$$

where m is the gravitational mass enclosed within a sphere of radius r . The mass of the star can be solved for as:

$$\frac{dm}{dr} = 4\pi r^2 \epsilon \quad (2.2)$$

where the total gravitational mass M of a star with radius R is given by $M \equiv m(R) = 4\pi \int_0^R dr r^2 \epsilon$. Given an EOS, numerically solving the TOV equation for M and R is straightforward. These equations create a one-to-one map from the EOS to the $M - R$ relation [17]; the inverse form of this map can therefore provide constraints on the EOS from observable properties. To mathematically invert the TOV equation, at least two stars' mass and radius must be known exactly, a feat is not possible with current observational technology. Solving the inverse problem is therefore much more complicated, potentially even intractable without making significant numerical assumptions.

2.2 X-ray spectroscopy for neutron stars

Many reliable observations of neutron stars come from X-ray emission, either from electromagnetic radiation from pulsars or thermal emission in quiescent low-mass X-ray binaries (qLMXBs). qLMXBs are particularly desirable to place strong constraints on neutron star structure as they are likely to have low magnetic fields (10^{8-9} G), resulting in minimal effects on the radiation transport or temperature distribution on the star's surface [27, 40, 41]. Additionally, these binaries are identified in globular clusters where distances, ages, and reddening are well-constrained [42]. The distinctive soft thermal spectra from these sources come from a long-lived thermal glow resulting from heat stored in the deep crust of the neutron stars within the binary system during accretion, which is then re-radiated from the whole surface when accretion stops [43]. For the context of this work, the inference of EOS will come from simulated thermal spectra from qLMXBs.

Observation of neutron star emission, whether X-ray or gravitational wave, has long served as a way to constrain mass and radius for neutron stars (e.g. [44, 45]), but uncertainties arise in the inference of these properties for a variety of reasons. In the case of X-ray radiation from qLMXBs, constraints on mass and radius are determined by fitting the emitted spectrum with an appropriate atmosphere model (where the surface composition is known or can be determined by the X-ray spectrum) and combining the spectroscopic measurements with the distance of the source. Models for thermal X-ray radiation are based on a light-element atmosphere, as the lightest element that is present in the atmosphere floats to the top due to rapid gravitational settling on neutron star surfaces [27]. Atmospheric models used on X-ray spectra from qLMXBs gave the first broad constraints on neutron star radius, and more modern analyses of X-ray spectra have provided tighter constraints on both radius and EOS.

The high-resolution imaging and spectroscopy of NASA's *Chandra X-ray Observatory* have provided powerful insight into neutron star properties like cooling [46, 47], mass and radius [27], and binary mergers of exotic stars [48]. *Chandra's* telescope contains a system of four pairs of mirrors that focus incoming X-ray photons to the Advanced CCD Imaging Spectrometer (ACIS), which measures the energy of each incoming X-ray. The observed

spectrum, along with a corresponding instrument response, is then fit to a well-motivated parameterized model. Many such models for spectral fitting exist in XSPEC [28], an X-ray spectral fitting package distributed and maintained by the aegis of the GSFC High Energy Astrophysics Science Archival Research Center (HEASARC). These parameterized models differ for different types of X-ray sources, as well as assumptions about the source’s atmosphere, magnetic field, temperature (a full list of models can be found in the XSPEC manual [28]). XSPEC has been used numerous times in the past to analyze data from *Chandra* as well as other spectrometers like *NICER*, *Nustar*, and *XMM-Newton*, making it a valuable resource for inference of neutron star properties.

3 Machine learning

Machine learning methods, in particular deep learning, aim to extract useful knowledge from data automatically and are rapidly being applied across many data-rich fields of science [22]. In regression tasks such as EOS inference, one is interested in constructing a function f whose inputs are the observed data and whose outputs are an estimate of some parameter of interest. The optimal function f is not known initially, but an approximation can be learned from a set of input-output example pairs.

In order to approximate f , machine learning methods first begin with a suitable class C of parameterized functions (e.g. polynomials of a certain degree, neural networks of a certain architecture) and then seek to find the best approximation to f within the class C . This is typically done through a stochastic gradient descent procedure that seeks to iteratively minimize the approximation error on the training set.

The well-known technique of linear regression is the most elementary form of regression, and can be viewed as a form of shallow learning (no hidden layers). Deep learning generalizes linear regression by using multi-layer neural networks as the class C and thus enabling the construction of sophisticated and flexible non-linear approximations. With sufficient training data and computing power, deep learning methods can handle large-scale problems with high dimensional data and avoid heuristic simplifications that lose information. It is not uncommon to deal with problems with input sizes in the range of up to 10^9 examples, each with dimensions of 10^{3-4} , with neural networks that can have up to 10^{11} free parameters. Training sets can range in size from 10^1 to 10^{10} or more. Unlike shallow learning and linear regression, deep learning does not require that the number of parameters be equal to the number of training examples [49]. More recent, attention-based architectures, such as transformers [50, 51], allow networks to take advantage of structures and symmetries in the data, and are applied in sections below.

When the interpretation of data depends on external unmeasured or poorly-known parameters, such as neutron star temperature or distance, it can be useful to apply *parameterized networks* [25]. Such networks learn a task as a function of the external parameter, allowing for evaluation of a fixed dataset under varying assumptions about the parameter [24].

4 Training samples

Samples of simulated neutron stars, used to train networks and evaluate their performance, are described below.

Each simulated star is described by two high-level summary quantities, mass and radius, which are drawn from the mass-radius relation determined by the EOS, as well as three nuisance parameters that are independent of the EOS and can vary from star to star. These five

Saturation Property	Value	Units
n_0	0.153	fm^{-3}
E_0	-16.3	MeV
K_0	300.0	MeV
m_N^*/m_N	0.70	—
J	32.5	MeV
L_0	55.0	MeV
U_N	-65.5	MeV

Table 1. Parameters of the model used to select example equations of state for the generation of simulated data samples. Shown are properties of the symmetric nuclear matter at saturation density for the GM1L parametrization of neutron star interiors [52]; see text for details.

parameters are sufficient to determine the expected simulated Chandra telescope spectrum in the chosen NS theoretical model. In the case of EOS inference, sets of stars with consistent EOS are grouped to form training and testing sets. Details of each step of the generation are provided below.

4.1 Equation of state

The equation of state of the hadronic matter within the core is modeled with the relativistic non-linear mean field model GM1L [52]. The version used here only accounts for protons and neutrons but can be extended to include hyperons and Δ baryons [53]. The corresponding saturation properties of symmetric nuclear matter for the GM1L parametrization are shown in table 1 [53, 54]. These properties include the nuclear saturation density n_0 , energy per nucleon E_0 , nuclear compressibility K_0 , effective nucleon mass m_N^*/m_N , asymmetry energy J , asymmetry energy slope L_0 , and the value of the nucleon potential U_N . The value of L_0 listed in table 1 is in agreement with the value of the slope of the symmetry energy deduced from nuclear experiments and astrophysical observations [55].

The most commonly used constraints on K_0 come from experimental values of the giant monopole resonance, which lie in the range of 220 to 260 MeV [56, 57]. The analysis of [58], however, suggests a higher range of 250 to 315 MeV. The value of $K_0 = 300$ MeV considered in our paper falls into the latter category, but this will not dramatically impact the neutron-rich equation of state appropriate for neutron star interiors. The GM1L equation of state for the core is paired with two models for the crust. For the outer crust, which falls in the density range 10^4 – 10^{11} g/cm³, we use the Baym-Pethick-Sutherland (BPS) model [59]. For the inner crust, with densities in the range 10^{11} – 10^{14} g/cm³, we use the Baym-Bethe-Pethick (BBP) model [60].

To limit the number of parameters the networks must learn, the essential features of the high-density portion of the EOS needed to be represented efficiently by just a few values. This can be done accurately by constructing parametric representations based on spectral fits, formed as generalizations of the Fourier series used to represent periodic functions [15]. An EOS, defined as $P = P(\epsilon)$ or $\epsilon(P) = \epsilon$, can be represented as a linear combination of basis functions $\epsilon_k(\phi)$:

$$\epsilon(p) = \sum_k \epsilon_k \Phi_k(p) \quad (4.1)$$

where $\phi_k(p)$ can be any complete set of functions. The EOS is therefore determined by the spectral coefficients ϵ_k , making $\epsilon_k = \epsilon(p, \epsilon_k)$. There are two important conditions that a physical EOS must satisfy to ensure microscopic stability. The first is that the EOS must be non-negative, or $p(\epsilon) \geq 0$, and the second is that pressure must be monotonically increasing with density [15]. Because these conditions are not naturally respected by arbitrary basis functions in a spectral representation, representing an EOS with a straightforward spectral expansion will likely produce data that violates microscopic stability and is therefore erroneous. To ensure these two conditions are met, we instead turn to a faithful construction of spectral representations of the EOS; the process for constructing these is outlined in detail in [61] and [15].

The spectral representation of GM1L is formed from representing the EOS in terms of the relativistic enthalpy, h , where the EOS can be rewritten as a pair of equations $P = P(h)$ and $\epsilon = \epsilon(h)$. The enthalpy can be defined as

$$h(P) = \int_0^P \frac{dP'}{\epsilon(P')c^2 + P'} \quad (4.2)$$

where c is the speed of light [61]. Inverting eq. (4.2) obtains the equation $P = P(h)$, which can recover the EOS $\epsilon(P)$ as $\epsilon(h) = \epsilon[P(h)]$. The pair of equations $P = P(h)$ and $\epsilon = \epsilon(h)$ can be expressed from a reduction to quadrature:

$$P(h) = P_0 + (\epsilon_0 c^2 + P_0) \int_{h_0}^h \mu(h') dh', \quad (4.3)$$

$$\epsilon(h) = -P(h)c^{-2} + (\epsilon_0 + p_0 c^{-2})\mu(h). \quad (4.4)$$

The function $\mu(h)$ is defined as:

$$\mu(h) = \exp \left\{ \int_{h_0}^h [2 + \Gamma(h')] dh' \right\} \quad (4.5)$$

which is dependent on the sound speed or velocity function $\Gamma(h)$ and constants $P_0 = P(h_0)$ and $\epsilon_0 = \epsilon(h_0)$ [61]. Similar to eq. (4.1), the velocity function can be represented as a spectral expansion:

$$\Gamma(h, \nu_k) = \exp \left[\sum_k \nu_k \Phi_k(h) \right] \quad (4.6)$$

where $\Phi_k(h)$ is any complete set of basis functions on the domain $[h_0, h_{\max}]$.

We constructed enthalpy-based causal spectral fits for the GM1L EOS with up to 10 parameters. Figure 2 shows that just two spectral parameters produce parameterizations with a mean relative error of only 10%, and additional parameters can reduce the error to 5% or lower. A small number of parameters is preferred due to the increased complexity of learning multiple parameters, and the danger of Runge's phenomenon when applying our networks to current neutron star observations with accurate readings of mass and radius, which are still relatively few. Runge's phenomenon arises when attempting to fit equispaced data points with polynomials of high degree; increasing the order of the polynomial interpolation can result in issues with convergence or divergence rates for certain functions [62]. When applied to neutron star observation, attempting to fit a small number of data points (with varying accuracy in observation method) with a model having many parameters may result in fits that accurately fit the data, but are very poor representations of the actual physics. Based

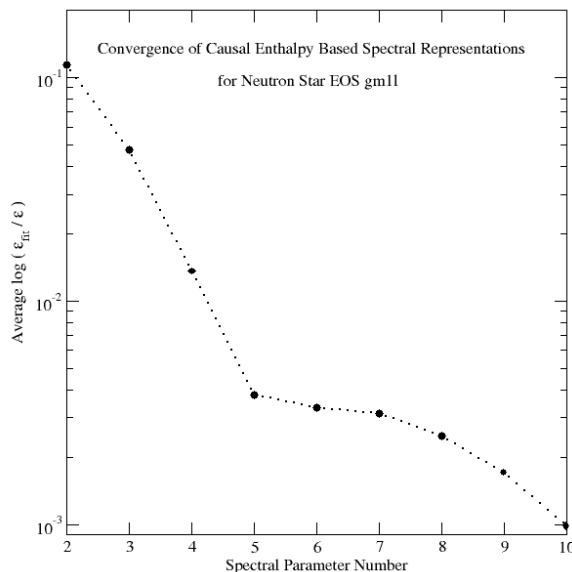


Figure 2. Relative error in spectral parameterization of the equation of state, shown as a function of the number of parameters used.

on the reasons listed above, we chose to use two spectral parameters to represent the EOS, hereafter referred to as λ_1 and λ_2 .

To create many samples needed for training and testing, spectral parameters were then constructed from the expression:

$$\lambda_{\text{generated}} = \lambda_{\text{true}} \cdot (1 + 2 \cdot \text{scale}(-0.5 + \text{ran2}))$$

where $\lambda_{\text{generated}}$ represents the newly constructed spectral parameter, λ_{true} is the best fit (true) spectral parameter of GM1L, and *scale* is a scaling parameter set to 0.05. *ran2* are uniformly distributed random numbers in the range 0 to 1 generated by the *ran2* function given in [63]. This process was repeated to create 10^4 different EOS variations. Each EOS variation was used to generate a coinciding M – R relation using equations (2.1), (2.2), examples of which are seen in figure 3, from which 100 (M, R) pairs are selected, each representing stellar parameters consistent with that EOS. Due to the random component of our EOS generation, some models have a mass peak below the current observed mass limit, $2.1 M_{\odot}$. All models have a minimum mass of at least $1 M_{\odot}$. The physicality of predicted results will be discussed in further detail 8.

4.2 Modeling X-ray spectra

The relation between stellar parameters (M, R) is determined by the EOS, and samples from the allowed curve are used as input to generate simulated X-ray Chandra spectra, such as the Chandra observation of the quiescent low-mass X-ray binary (qLMXB) X7 in the globular cluster 47 Tuc [27].

The XSPEC program [28], which can be used for spectral fitting, is also capable of generation of simulated spectra, via the *fakeit* command when a NS model and telescope response matrix are provided.

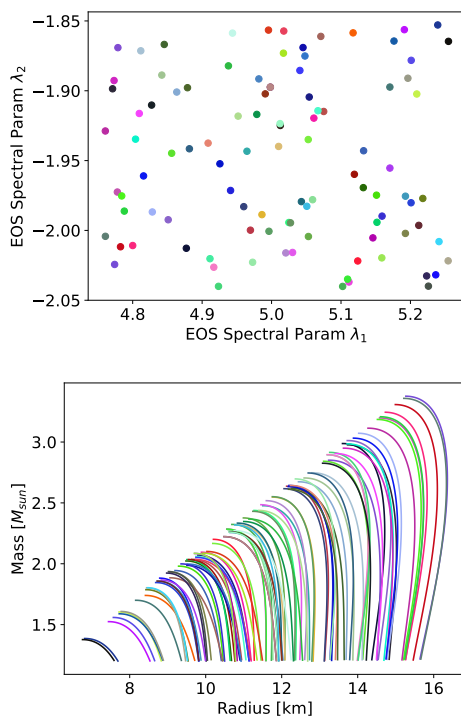


Figure 3. Examples of training data. Top, 100 samples in EOS spectral parameter space (λ_1, λ_2) randomly selected from the full set of 10,000 EOS spectral pairs. Bottom, neutron star mass-radius curves determined by the selected EOS parameters.

The NS theoretical model NSATMOS [13] selected includes a hydrogen atmosphere model with electron conduction and self-irradiation. The Chandra telescope response specified in ref. [13] was also used to describe the instrument response and telescope effective area.

4.3 Nuisance parameters

The NSATMOS model has five parameters to describe each star: gravitational mass M in units of M_\odot , radius R in units of km, and three additional parameters related to observation. For the context of M – R and subsequent EOS inference, only M and R are parameters of interest, whose values come from those generated by the GM1L EOS and so provide information relevant to the physical question. The remaining three nuisance parameters are the effective temperature of the surface, T_{eff} , the distance to the star, d , and the hydrogen column, N_{H} which parameterizes the reddening of the spectrum by the interstellar medium. These parameters influence the observed spectrum of a given neutron star. Lack of knowledge of these values is a leading source of uncertainty in the inference of mass and radius, and hence EOS.

Using table 1 in ref. [45] as a guide, we find that distances typically range between 2 and 10 kpc, and hydrogen columns lie between 0.2 and $5 \times 10^{21} \text{ cm}^{-2}$. While neutron stars with larger distances and larger hydrogen columns exist, they are sufficiently distant as to be difficult to obtain spectral information. From table 3 in ref. [64], effective temperatures at the surface typically lie between 50 and 200 eV, or from 6×10^5 and 2.4×10^6 K. Note that core temperatures are typically a few orders of magnitude larger. Again colder neutron stars most certainly exist but are more difficult to observe.

Nuis. Param.	True	Tight	Loose
Distance	exact	5%	20%
Hydrogen Column N_{H}	exact	30%	50%
$\log(T_{\text{eff}})$	exact	± 0.1	± 0.2

Table 2. Description of “true”, “tight”, and “loose” nuisance parameter (NP) scenarios. Shown are the width of each Gaussian distribution representing the prior knowledge of each NP. For distance and N_{H} , width is relative; for $\log(T_{\text{eff}})$, it is absolute. See text for details and references.

Examples of generated spectra for varying stellar parameters are shown in figure 4. The generated spectra are very sensitive to the effective surface temperature, with lesser sensitivity to other parameters. The dependence of the curves in figure 4 to the changing nuisance parameters is not surprising: roughly proportional to radius and distance squared, but higher power in temperature.

The networks detailed below provide estimates of either the neutron star mass and radius or the EOS parameters, conditioned on NP values. Uncertainty in regressed parameters of interest due to uncertainties in the NP can then be fully propagated via variation of the NPs used during regression. To demonstrate the impact of NP uncertainties, we define three example scenarios of uncertainties, dubbed “true”, “tight”, and “loose”, which describe the quality of prior information on the NP values for each star.

In the “true” scenario, the NPs are set to the true value used to generate the spectra, such that the NP prior is essentially a delta function. In the “tight” scenario, the uncertainty is described as a narrow Gaussian for each NP, with distance having a width of 5%, hydrogen column having a width of 30%, and $\log(T_{\text{eff}})$ having a width of 0.1. In the “loose” scenario, the uncertainties are described by a wider Gaussian, with distance having a width of 20%, hydrogen column having a width of 50%, and $\log(T_{\text{eff}})$ having a width of 0.2. These ranges are shown in table 2. The sensitivity to NP values is reflected in the performance of the networks below.

5 Inference of EOS from mass and radius

Previous applications of machine learning to the task of inferring the equation of state have begun from the stellar mass and radii [18–20], or equivalent parameters [21], though with simple ad-hoc descriptions of the uncertainty on stellar mass and radius values, often modeled as two-dimensional uncorrelated Gaussians rather than fully propagating the underlying uncertainties. In this section, we tackle the same problem, but where the stellar data are more realistic and the underlying uncertainties are fully propagated to the EOS estimation. Specifically, the best estimates of stellar mass and radius are derived using state-of-the-art tools that extract them from realistic stellar spectra, which include the impact of stellar nuisance parameters and limited observation time. In addition, this mass-radius estimation is conditioned on the nuisance parameters, such that variations in those nuisance parameters lead to variations in the mass and radius estimates. This connects directly to the neural network regression of EOS parameters from mass-radius values parameterized in the nuisance parameters, allowing for the direct propagation of the underlying uncertainties to give a measure of the resulting uncertainty on the regressed EOS parameters.

Below, we describe the extraction of realistic mass and radius values with XSPEC and their subsequent use in NN regression of the EOS parameters and the estimation of the

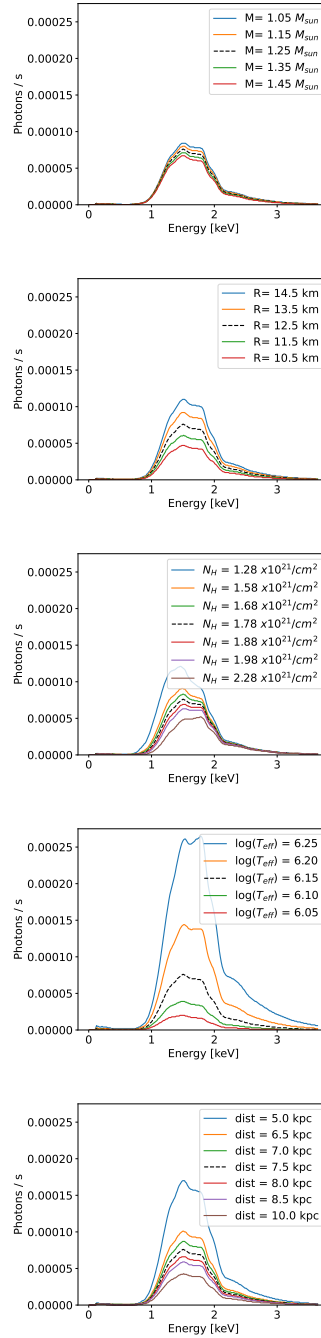


Figure 4. Examples of simulated stellar spectra expected for several values of stellar parameters. Each pane shows the expected rate of photons in Chandra per energy bin, for variations of the parameters of interest (mass M , radius R) as well as for variations of the nuisance parameters (N_H , $\log(T_{\text{eff}})$, distance). The dashed black line has the same parameters in each pane.

uncertainty. For comparison, we also provide a demonstration of the regression of EOS parameters using polynomial regression. In subsequent sections, we consider an alternative extraction of mass and radius using a NN, as well as end-to-end regression of EOS directly from stellar spectra.

5.1 Mass and radius inference by XSPEC

Sample stellar spectra are generated as described above, including Poisson noise corresponding to an observation time of 100 ks, and nuisance parameters variations as specified in table 2.

Given a sample observed X-ray spectrum, the XSPEC code scans the mass and radius parameter space, searching for values that best describe it. For each mass-radius pair, the expected spectrum is calculated using the chosen model and telescope response function, identical to those used to generate the sample spectra being fit. The fitted values are those which minimize a bin-wise χ^2 , and reported errors are those which generate a fixed increase in the χ^2 metric.

To propagate the uncertainty due to the lack of knowledge of the NP values, the fit on a given spectrum is performed several times with varying assumed values of the NPs drawn from the appropriate prior. The variation in the resulting fitted values then describes the impact of uncertainty on the NPs. For this reason, during each single XSPEC fit, the NP values are not allowed to vary, but are frozen. Figure 5 shows examples for individual stars, demonstrating the variation of the stellar parameter estimates with varying NP values. For illustrative comparison to ad-hoc models of uncertainty, the standard deviation in mass and radius are used to define the widths of a 2d error ellipse, though it is clear that this fails to capture the complex nature of the impacts of the underlying uncertainties; these simple error models are not used in our analysis.

XSPEC is also capable of *floating* the nuisance parameters, varying their values to improve an individual fit, and reporting an uncertainty envelope in the mass-radius plane. This can be helpful in the case where the mass and radius and their envelope are the final targets. However, to propagate the uncertainty downstream requires that we have the full posterior in the mass-radius plane or samples from it. An estimate and envelope do not provide that capacity, though they can allow for ad-hoc parameterizations of the prior as have been performed previously. We condition on the nuisance parameters to allow full propagation of the NP uncertainty through to EOS estimation, as we do below.

Performance of XSPEC regression of mass and radius is shown in figure 6, where the residuals increase as expected with wider priors on the nuisance parameters. In addition, note that in the case of the “loose” priors, there is a small fraction of cases where XSPEC fails to converge on an estimate, as the nuisance parameters are fixed to a value far from the value used to generate the spectrum.

One important note regarding XSPEC’s performance is the same theoretical model, NSATMOS, is used in both the data generation and regression. Because of this, the regression models discussed in the next section can, at best, match XSPEC’s performance for the evaluation dataset. Nonetheless, it is an important step in demonstrating the capacity of these methods to perform such inference without explicitly relying on a single theoretical model.

5.2 Inference of EOS from mass and radius

In this section, we demonstrate the inference of EOS parameters from the stellar mass and radius data, using neural network regression parameterized in the NPs to allow propagation of the uncertainty. In addition, we build polynomial regression models which serve as a benchmark, following ref. [18].

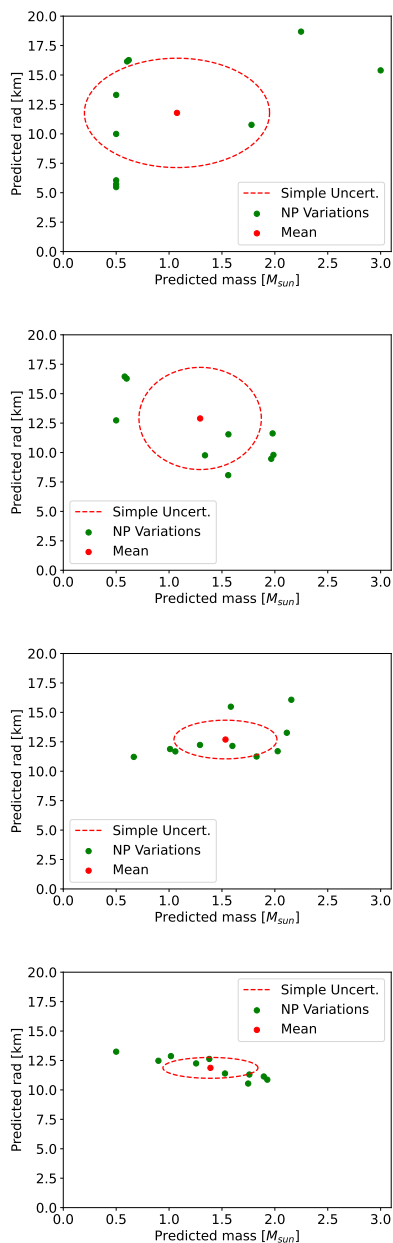


Figure 5. Estimation of the mass and radius of a neutron star from the underlying stellar spectra, by XSPEC. Each pane represents one star, and shown (green) are estimates for several independent values of the nuisance parameters drawn from the associated priors, and the mean value (red). Top two cases have loose priors, bottom two have tight. The dashed ellipse, whose widths are set to the standard deviation of the mass and radius estimates, is a demonstration of the inadequacy of a simple uncertainty model.

5.2.1 Neural network regression

Deep feed-forward neural networks are trained to provide the EOS parameters given a collection of ten stars, each represented by their mass and radius. Each network has two outputs, λ_1 and λ_2 .

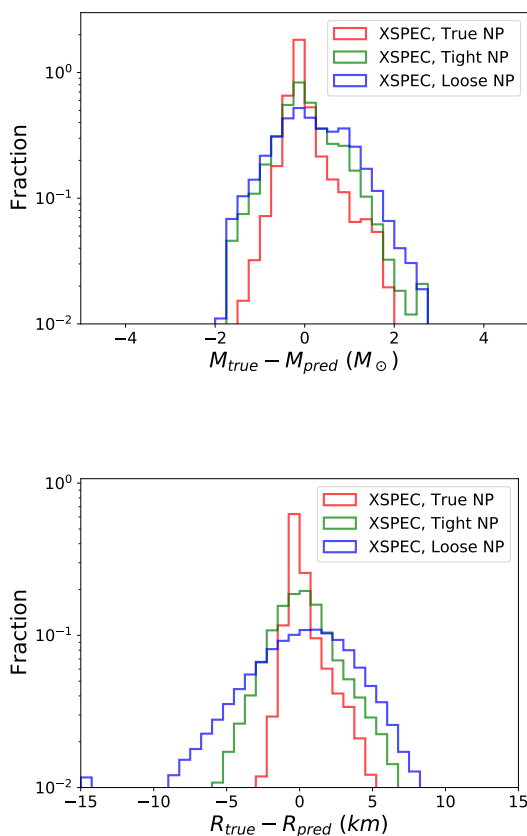


Figure 6. Performance of XSPEC inference of neutron star mass and radius, as measured by residuals between the fitted (“pred”) and true values under three treatments of the nuisance parameters (NPs). In the “true” case, the NPs are fixed to their true values; in the “tight” and “loose” cases, they are drawn from narrow or wide priors, respectively; see text for details. Cases in which XSPEC fails to converge are shown as large negative residuals.

We train three networks, one for each of the true, tight, and loose NP scenarios. All networks have identical architecture, 10 hidden layers with 32 nodes each followed by an output layer with 2 nodes. Rectified linear units are used as activation functions for the hidden layers while linear activations are used for the output layers. They were trained up to 1000 epochs with a mean squared error (MSE) loss and an Adam optimizer [65], and the performance is evaluated on independent validation data. The performance was not found to be highly sensitive to hyper-parameter tuning, so permutation symmetry preserving architectures were not explored. The networks were implemented using `Tensorflow 2.7.0` on a single NVIDIA RTX A5000 GPU.

5.2.2 Polynomial regression

Following the example of ref. [18], as a performance benchmark we also construct a polynomial regression model to regress EOS parameters λ_1 and λ_2 from stellar mass and radius information. The input for each model is a 20×1 vector containing the following information from each of ten stars: mass M , radius R .

Each 20×1 input to the polynomial regression network represents a set of stars, all of which are chosen from the same EOS. We construct two multivariate polynomial regression models of degree two, following the general form:

$$\lambda = \beta_0 + \sum_{i=1,2}^N \beta_i x^i$$

where each β are coefficients and $N = 20$ to represent the cluster size, or 10 mass-radius pairs. Polynomial features were created using the machine learning toolbox scikit-learn [66], and subsequently fit to a linear regression model. This model uses optimization in the form of ordinary least squares, which takes the form:

$$\min_{\beta} \|X\beta - y\|_2^2$$

where X is the $N = 20$ input vector, and y is the target EOS parameter, either λ_1 or λ_2 .

5.2.3 Estimation of uncertainty

The uncertainty in the underlying NS nuisance parameters has a significant impact on the estimation of EOS parameters. Because the mass and radius estimation is conditioned on the NPs, leading to variations in the mass and radius, (e.g. see figure 5), those variations can be propagated through the EOS estimation. The significant uncertainties in stellar nuisance parameters and the small number of stars observed to date make the treatment of those uncertainties vital. The validity of the final result is only as powerful as the validity of its uncertainties. In figure 7, examples of the variation of the EOS estimates are shown, where the underlying stellar spectra are fixed. Thus, this provides a measure of the uncertainty in the EOS due to the uncertainty in the NPs.

5.2.4 Performance

Performance of neural network regression of the EOS parameters λ_1 and λ_2 are compared to polynomial regression of the same quantities using identical datasets via comparison of the residuals, the difference between the true and regressed values. As seen in figures 8 and 9, while PR is able to achieve narrower residuals in the true case, the network regression is more robust in cases with larger uncertainties.

This result confirms what has been seen in earlier studies of NN regression from mass-radius pairs [18], but our study extends previous work by using realistic values of the mass and radius inferred from realistic simulated spectra, as well as by demonstrating uncertainty quantification, via full propagation of the underlying uncertainties due to nuisance parameters.

6 Inference of mass and radius from spectra

Previous applications of machine learning to neutron star datasets focus on analysis of mass-radius pairs, as demonstrated above, rather than direct analysis of the stellar spectra by neural networks.

A potential obstacle to direct analysis of spectra by neural networks is that the spectra are high-dimensional, often with $\mathcal{O}(10^3)$ bins of photon energy. However deep learning methods combined with GPUs have no trouble analyzing data with similarly high dimensionality [25], opening up new opportunities to tackle this important topic. While the mass

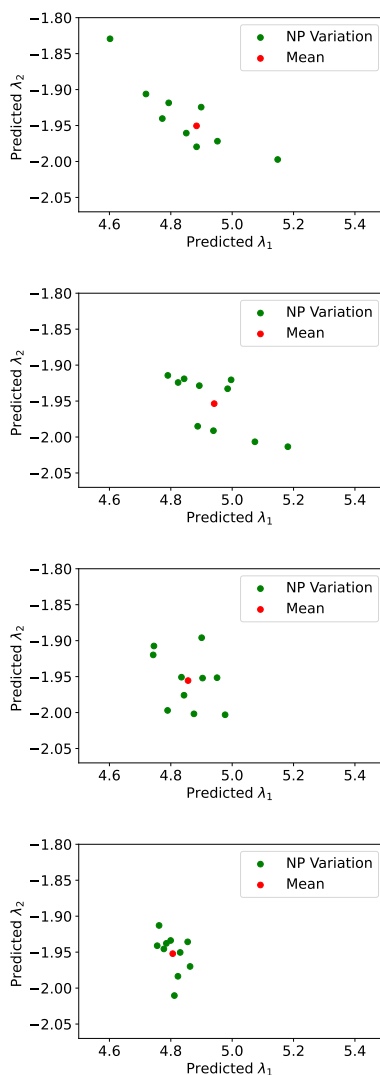


Figure 7. Neural network regression of the EOS parameters λ_1 and λ_2 of a set of 10 neutron stars from their masses and radii as estimated by XSPEC from each stars spectrum. Each pane represents an example dataset of 10 simulated stars, and shown (green) are EOS estimates for several independent values of the stellar nuisance parameters drawn from the associated priors, and the mean value (red). Top two cases have loose priors, bottom two have tight.

and radius are powerful summaries of the information in the lower-level spectra which is relevant to the equation of state, direct ML analysis of the spectra themselves may allow for the extraction of additional information or provide more robust propagation of uncertainties. As an initial step, we begin by estimating the stellar mass and radius from a single stellar spectrum before moving on to end-to-end inference of EOS parameters directly from a set of spectra in the next section.

In this section, we apply machine learning to the task of extracting the mass and radius from the stellar spectra, training a network we refer to as MR_Net. This serves as a demonstration of the capabilities of ML to grapple with high-dimensional datasets, allows us to harmonize the treatment of nuisance parameters end-to-end from spectra to EOS, and potentially extract more relevant information.

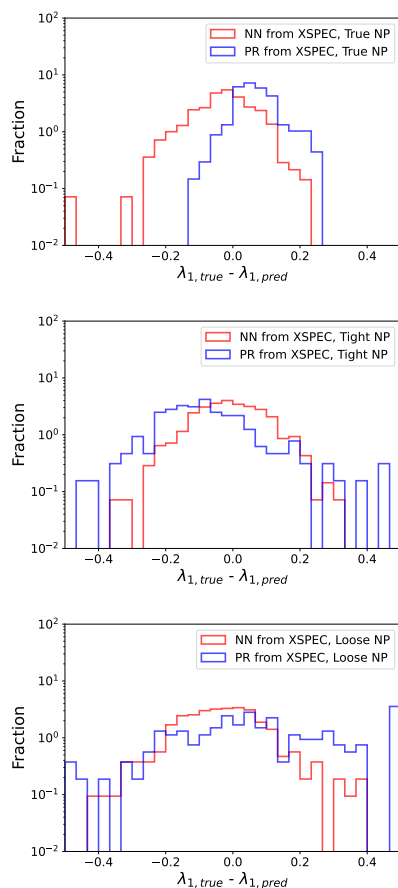


Figure 8. Comparison of the performance of NN regression and polynomial regression of EOS parameter λ_1 from mass radius pairs inferred by XSPEC from stellar spectra. Shown is the residual, the difference between the predicted and true values for each of three treatments of the stellar nuisance parameters. In the “true” case, the NPs are fixed to their true values; in the “tight” and “loose” cases, they are drawn from narrow or wide priors, respectively; see text for details.

6.1 MR_Net method

We build a network whose inputs correspond to the bins of the stellar X-ray spectrum, and whose outputs are the estimates of the star’s mass and radius. In addition, the mass and radius regressor is parameterized on the stellar nuisance parameters (distance, N_{H} , $\log(T_{\text{eff}})$), which allows the results to be conditioned on the nuisance parameters.

This architecture is composed of two input branches, one to process the star’s spectra and another to process the corresponding nuisance parameters. Each branch contains a series of layers that process its inputs in isolation. Following these initial layers, the output from the branches is combined, forming a single vector containing all the information. This vector is then passed to a final series of layers to predict the star’s mass and radius. Each segment of the network, both the branches and the main trunk of the network, contain four layers, giving the network eight layers in total. All fully connected layers contain 275 nodes and utilize a dropout probability of 0.25. The network employs skip connections between alternate layers. This stabilizes the training process and adds robustness to the network overall. The network is trained with an MSE loss and an Adam optimizer with an initial learning rate of 0.00017 which is slowly decayed over the course of training.

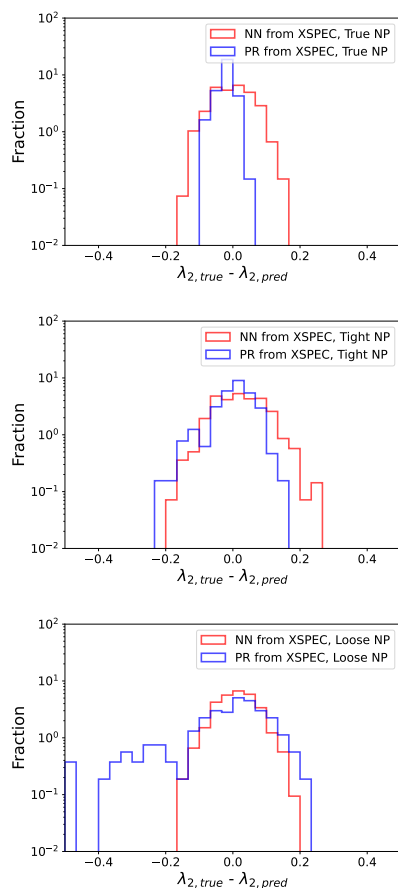


Figure 9. Comparison of the performance of NN regression and polynomial regression of EOS parameter λ_2 from mass radius pairs inferred by XSPEC from stellar spectra. Shown is the residual, the difference between the predicted and true values for each of three treatments of the stellar nuisance parameters. In the “true” case, the NPs are fixed to their true values; in the “tight” and “loose” cases, they are drawn from narrow or wide priors, respectively; see text for details.

Mass-radius regression is formulated as a supervised learning problem, where the network learns to minimize the error between the true mass and radius and its predictions. The network weights are updated by stochastic gradient descent using backpropagation. The Huber loss function is used and the Adam optimizer computes gradients and schedules the backward passes.

6.2 MR_Net performance in mass, radius

We begin with the best-case scenario in which the nuisance parameters are known with zero uncertainty, referred to as “true NP” above. Figure 10 shows the performance of MR_Net given neutron star spectra with statistical noise corresponding to 100,000 seconds (100 ks) of observation time, as well as for spectra without statistical noise. This demonstrates the contribution of statistical noise to the residual and demonstrates the network’s ability to digest the spectral information and understand the dependence on mass and radius.

What is clear is that MR_Net is capable of extracting the mass and radius values of the star directly from the spectrum. We emphasize that the network is trained on examples, but

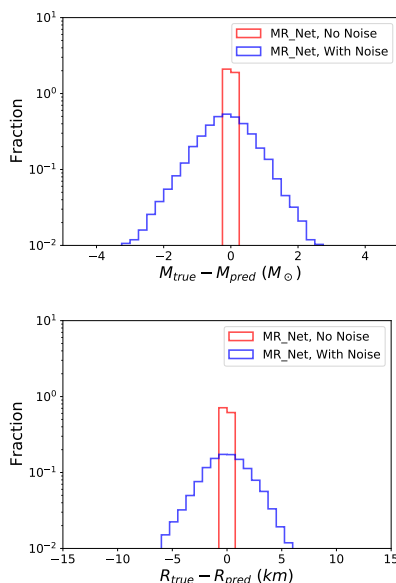


Figure 10. Performance of the MR_Net regression of a neutron star mass (top) and radius (bottom) from its stellar X-ray spectrum. Shown is the residual, the difference between the true and predicted values, for spectra with statistical noise (blue) corresponding to an observation time of 100 ks, and for spectra without statistical noise (red), which demonstrates the capacity of the network. Nuisance parameters are fixed to their true values.

does not benefit from the knowledge of the theoretical model used to generate these stars, while XSPEC requires precise specification of the theoretical model. This lack of requirement of a theoretical model opens new possibilities, such as training MR_Net to interpolate smoothly between theoretical models by providing a mixed or parameterized training set.

The MR_Net is conditioned on the nuisance parameters, allowing for propagation of the NP uncertainties through to the regression target as was done with XSPEC estimates. To assess the impact of NP uncertainty, we draw from priors on the NPs under the “tight” and “loose” scenarios defined earlier. The residuals widen, as expected. Figures 11 and 12 show the mass and radius residuals, respectively, under each NP scenario for MR_Net and XSPEC. Table 3 shows the mean and width of each residual distribution, as well as the combined width. As an additional comparison, figure 13 shows the ratio of predicted values from MR_Net to true values subtracted from 1.

MR_Net is capable of analyzing the stellar spectrum directly and extracting stellar parameters in a robust manner that allows for propagation of NP uncertainties.

6.3 Network uncertainty on mass and radius

Conditioning the extraction of M and R on the nuisance parameters allows for the natural propagation of the corresponding uncertainties. As was done for XSPEC mass and radius estimates, we propagate the NP uncertainty through to mass and radius uncertainty by sampling from the stellar NP priors several times for a given stellar spectrum, performing the mass and radius regression multiple times. Figure 14 demonstrates this for several example stars. Note that the variation of NP values does not produce variation in M and R which can be accurately summarized by 2D uncorrelated Gaussians, as has been assumed in previous studies [18] with ad-hoc datasets.

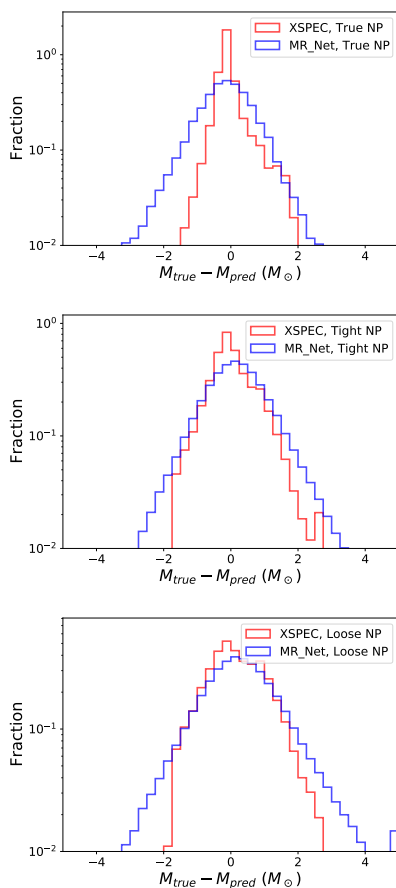


Figure 11. Performance of the MR_Net regression of a neutron star mass from its stellar X-ray spectrum, compared to regression using XSPEC. Shown is the residual, the difference between the true and predicted values, for three scenarios of nuisance parameter uncertainties. In the “true” case, the NPs are fixed to their true values; in the “tight” and “loose” cases, they are drawn from narrow or wide priors, respectively; see text for details.

6.4 Revisiting EOS regression

The question driving the analysis of neutron star spectra is not a desire to measure their masses and radii, but to use those to determine the equation of state parameters. In this section, we push the results of MR_Net through our NN regression of EOS parameters to analyze the performance of spectra $\rightarrow (M, R) \rightarrow$ EOS regression. In later sections, we remove the intermediate step and perform direct spectra \rightarrow EOS regression.

Performance of the EOS regression using stellar mass and radius information from MR_Net is shown in figures 18 and 19, and compared to regression from mass and radius information by XSPEC. Examples of uncertainty propagation through to EOS estimates are shown in figure 15.

7 Inference of EOS from spectra

In the previous section, we connected the two ML models, MR_Net (spectra $\rightarrow M, R$) and our EOS regression NN ($M, R \rightarrow$ EOS). This required the collapse of the full information

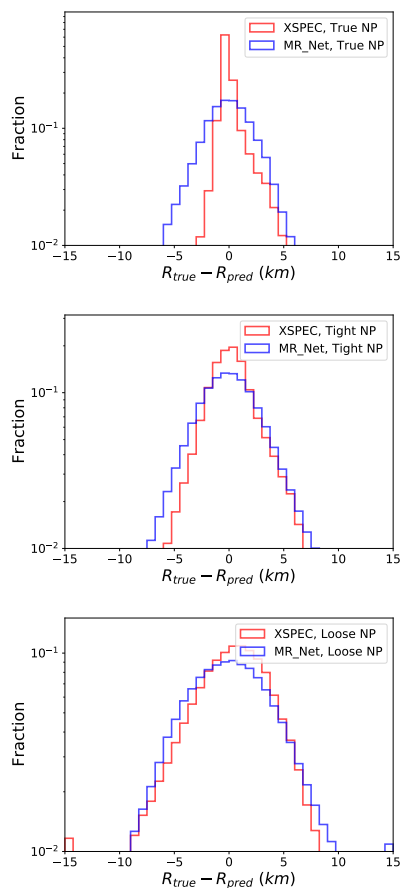


Figure 12. Performance of the MR_Net regression of a neutron star radius from its stellar X-ray spectrum, compared to regression using XSPEC. Shown is the residual, the difference between the true and predicted values, for three scenarios of nuisance parameter uncertainties. In the “true” case, the NPs are fixed to their true values; in the “tight” and “loose” cases, they are drawn from narrow or wide priors, respectively; see text for details.

and dependence on nuisance parameters into these two physical quantities. At the same time, MR_Net demonstrated that it is possible to regress physics quantities directly from high-dimensional stellar spectra. In this section, rather than connecting two networks via mass and radius, we use a single network to perform end-to-end regression of EOS parameters from a set of stellar spectra, avoiding the information collapse, keeping the full information, and allowing for robust propagation of the stellar nuisance parameters into uncertainty quantification for the EOS parameters of interest.

In addition to a demonstration of the power of networks to directly analyze low-level data, this allows us to probe the question of whether the mass and radius are sufficient statistics, and whether they contain *all* of the information relevant to the problem. There are many examples in the literature in which such well-motivated high-level heuristics fail to capture the complete information contained in lower-level data. In this case, while in principle the mass and radius are all that are required to infer the EOS in the context of a fixed theoretical stellar model, such information is never without uncertainty. Full propagation of the dependence on nuisance parameter uncertainty may allow for more accurate and robust estimates.

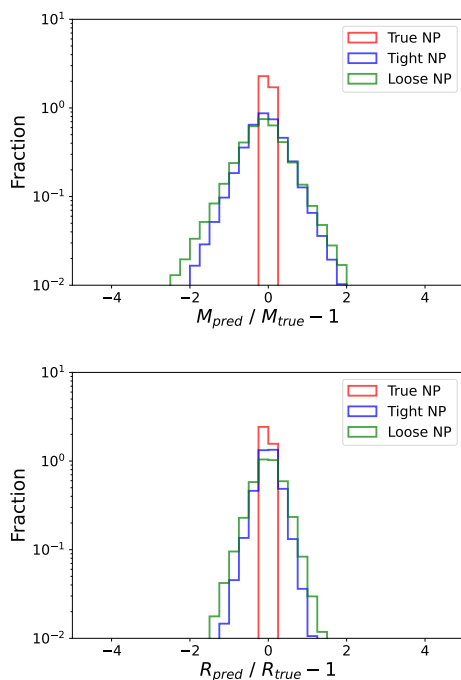


Figure 13. Performance of the MR_Net regression of a neutron star mass (top) and radius (bottom) from its stellar X-ray spectrum. Shown is the ratio of predicted (“pred”) to true values minus one for three scenarios of nuisance parameter uncertainties. In the “true” case, the NPs are fixed to their true values; in the “tight” and “loose” cases, they are drawn from narrow or wide priors, respectively; see text for details.

	Nuis.	Mass		Radius		Combined
Method	Params	μ	σ	μ	σ	σ
XSPEC	True	-0.01	0.50	0.23	1.44	1.51
MR_Net	True	-0.14	0.93	-0.07	2.80	2.99
XSPEC	Tight	-0.06	0.73	0.24	2.61	2.69
MR_Net	Tight	0.17	1.06	0.06	3.52	2.76
XSPEC	Loose	0.18	0.86	-0.06	4.32	4.40
MR_Net	Loose	0.28	1.29	0.14	4.93	5.10

Table 3. Performance of the regression of neutron star mass and radius for XSPEC as well as our neural network regression, MR_Net, which lacks any knowledge of the theoretical model. Shown are the mean (μ) and standard deviation (σ) of the residuals under three scenarios of nuisance parameter uncertainties. In the “true” case, the NPs are fixed to their true values; in the “tight” and “loose” cases, they are drawn from narrow or wide priors, respectively; see text for details. The combined column is a quadrature sum of the standard deviations of radius and mass.

Furthermore, there are properties of neutron stars that can be deciphered from spectra beyond simply mass and radius. Quantities like temperature inhomogeneities [67] may impact a star’s equation of state but are not captured by the mass and radius.

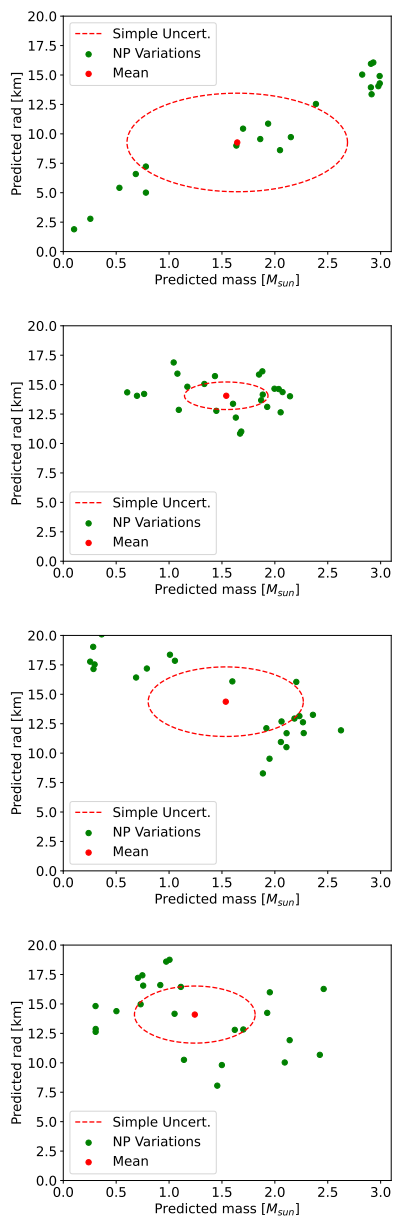


Figure 14. Estimation of the mass and radius of a neutron star from the underlying stellar spectra, by MR_Net. Each pane represents one star, and shown (green) are estimates for several independent values of the nuisance parameters drawn from the associated priors, and the mean value (red). Top two cases have loose priors, bottom two have tight. The dashed ellipse, whose widths are set to the standard deviation of the mass and radius estimates, is a demonstration of the inadequacy of a simple uncertainty model.

7.1 Architecture

Many neural network architectures operate on sequences of vectors, rather than set of vectors. For instance, in natural language processing, the input may be a sentence where each word is converted to a vector and the ordering of the vectors matters. However, in the case of neutron stars and other problems, we need neural networks that operate on sets of vectors, such as the independent spectra observations for multiple stars.

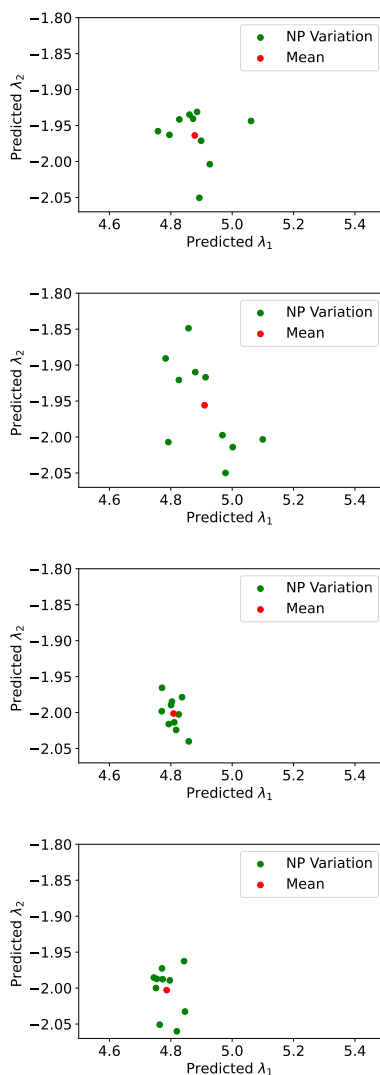


Figure 15. Neural network regression of the EOS parameters λ_1 and λ_2 of a set of 10 neutron stars from their masses and radii as estimated by MR_Net from each stars spectrum. Each pane represents an example dataset of 10 simulated stars, and shown (green) are EOS estimates for several independent values of the stellar nuisance parameters drawn from the associated priors, and the mean value (red). Top two cases have loose priors, bottom two have tight.

One architectures with invariant properties with respect to permutations of the input vectors is the transformer architecture [50, 51]. Perhaps surprisingly, transformers were originally developed for problems in natural language processing, thus requiring the addition of positional information bits to the vector encoding each word in order to recover the sequential dimension. More recently they have been used in other areas, including physics [68, 69] in order to leverage their permutation invariance properties. Transformer architectures typically consists of stacks of encoder modules followed by decoder modules. The structure of each encoder module and each decoder module is similar, thus we describe only a typical encoder module. A transformer encoder module accepts an unordered set of inputs and produces a set of outputs. The transformer employs a mechanism called *self-attention*, which allows it to

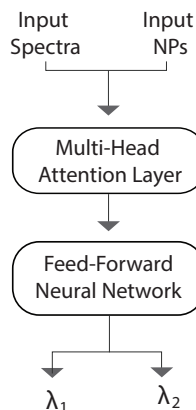


Figure 16. Schematic diagram of the transformer neural network used to determine EOS coefficients λ_1 and λ_2 from an input of spectra and NPs.

compare each element in the set against every other. This mechanism allows the network to attend to important features in the set while computing an output prediction. Briefly, self-attention (equation (7.1)) operates on an input matrix X , with N rows and D columns. Three matrices are produced from the projection of X with differing, trainable, weight matrices: $Q = XW_Q$, $K = XW_K$, $V = XW_V$ (termed Query, Key, and Value respectively).

$$S = D(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_q}} \right) V. \quad (7.1)$$

Thus in short each output corresponds to a different convex combination of the Value vectors, where each convex combination depends on the degree of similarity between the corresponding Query and Key vectors. The similarity is computed by taking dot products between corresponding Query and Key vectors, and then applying a softmax to yield a convex combination (see [50, 51] for additional details).

In the neutron star application, the transformer architecture takes as input spectra and corresponding nuisance parameters for each star in the set. In the results shown below, the network is given a set of 10 stars, though the structure of the network allows it to accept larger or smaller datasets with minimal modification. The final output of the network is the two EOS parameters. This is shown schematically in figure 16.

The network is composed of six consecutive transformer blocks. Each block processes the input through multi-head attention (with eight heads), followed by dropout ($p = 0.15$), normalization, and fully connected layers. Following these blocks the output is processed by one final fully connected layer in order to regress the EOS coefficients λ_1 and λ_2 . All fully connected layers, with the exception of the final one, use the ReLU activation function. The Adam optimizer was used to provide gradient updates with an initial learning rate of 0.000075, which was slowly decayed over the course of training.

7.2 Training

The parameters of the network architecture and the learning algorithm, the hyperparameters, were optimized with Sherpa [70], a Python library for hyperparameter tuning. The algorithm

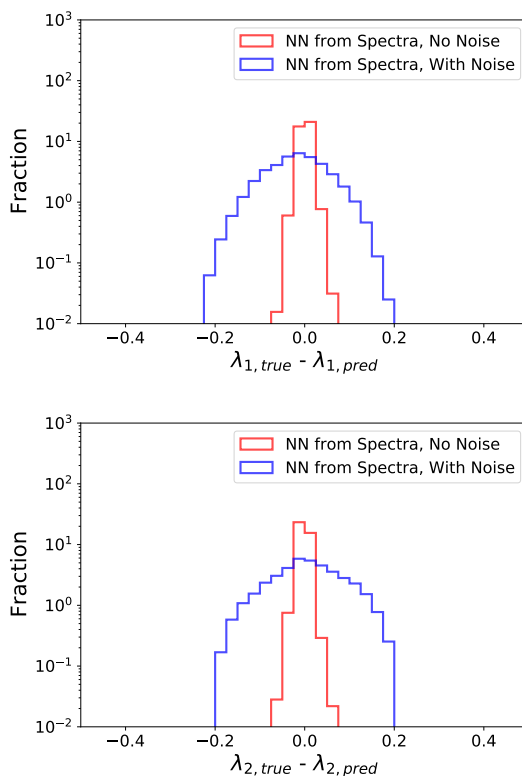


Figure 17. Performance of the neural network regression of the neutron star EOS parameters λ_1 (top) and λ_2 (bottom) directly from a set of stellar X-ray spectra, without intermediate prediction of the mass and radius. Shown is the residual, the difference between the true and predicted values, for spectra with statistical noise (blue) corresponding to an observation time of 100 k seconds per star, and for spectra without statistical noise (red), which demonstrates the capacity of the network. Nuisance parameters are fixed to their true values.

used is based on a random search and has the advantage of making no assumptions about the structure of the hyperparameter search problem and thus is well suited for exploring a variety of parameter settings. An initial exploratory search was conducted on a subset of the data to find appropriate hyperparameters.

Following this exploratory phase, the network is trained for 1,000 epochs on the full dataset. The Adam optimizer [65] is used for gradient descent, with early stopping monitoring of the validation loss to prevent overfitting.

7.3 Results

Figure 17 shows the performance of spectra \rightarrow EOS regression in the best-case scenario, where the nuisance parameters are perfectly known. Shown is the performance with statistical noise corresponding to 100 ks of observation time per star, as well as for spectra without statistical noise. While there is significant width to the residuals, this is dominated by the statistical uncertainty, not the network’s ability to digest the spectral information and understand the dependence on mass and radius. This clearly demonstrates the network’s capacity is sufficient for the regression task.

Method	Params.	Nuis.	λ_1		λ_2		Combined
			μ	σ	μ	σ	σ
NN(Spectra)	True		-0.02	0.066	0.01	0.075	0.099
NN(M, R via MR_Net)	True		-0.03	0.089	-0.02	0.068	0.112
NN(M, R via XSPEC)	True		-0.03	0.065	0.01	0.055	0.085
NN(Spectra)	Tight		0.02	0.085	-0.02	0.077	0.115
NN(M, R via MR_Net)	Tight		0.00	0.104	0.02	0.072	0.126
NN(M, R via XSPEC)	Tight		-0.03	0.081	0.01	0.056	0.098
NN(Spectra)	Loose		-0.03	0.131	-0.01	0.078	0.152
NN(M, R via MR_Net)	Loose		-0.01	0.135	-0.02	0.078	0.156
NN(M, R via XSPEC)	Loose		-0.03	0.123	0.01	0.058	0.136

Table 4. Performance of the regression of neutron star EOS parameters λ_1 and λ_2 using direct regression from spectra, as compared to NN regression from mass and radius (M, R) information extracted via MR_Net or XSPEC. Shown are the mean (μ) and standard deviation (σ) of the residuals under three scenarios of nuisance parameter uncertainties; distributions are given in figures 18 and 19. In the “true” case, the NPs are fixed to their true values; in the “tight” and “loose” cases, they are drawn from narrow or wide priors, respectively; see text for details. The combined column is a quadrature sum of the standard deviations of λ_1 and λ_2 .

We next analyze the performance of the direct regression in cases where the nuisance parameters are not perfectly determined. Figures 18 and 19 shows the residuals in the EOS parameters for the end-to-end regression, as compared to regression from mass and radius information provided by MR_Net or XSPEC from the stellar spectra. Table 4 summarizes the performance for each method.

As the full network is again conditioned on the NPs, we can propagate this uncertainty directly through our regression. Figure 20 demonstrates how variations of the NPs, drawn from the appropriate priors, provide a measure of the uncertainty on the final result.

8 Discussion

The performance of the three methods are measured in simulated stellar samples generated with the same theoretical stellar model that is assumed by XSPEC, which makes it a valuable upper limit for the two fully neural network based methods, which must infer the relationships. The three approaches perform comparably, and the end-to-end method slightly but consistently outperforms the two-step method using MR_Net. Once trained, the end-to-end network can handle any prior on the nuisance parameters, whereas the networks that rely on XSPEC fits or MR_Net predictions need to first be trained on data with the desired prior.

When the data are simulated and drawn from a known theoretical model, one cannot achieve more statistical power than directly calculating the likelihood. But even powerful theoretical models for spectral fitting still rely on a variety of assumptions about the spectrum’s source. The flexibility of these fully neural network based approaches is an important advantage, opening the door to interpolating between theoretical models [24], or even learning directly from observational data [71, 72]. An inference approach with this flexibility and the capacity for robust propagation of uncertainties is vital.

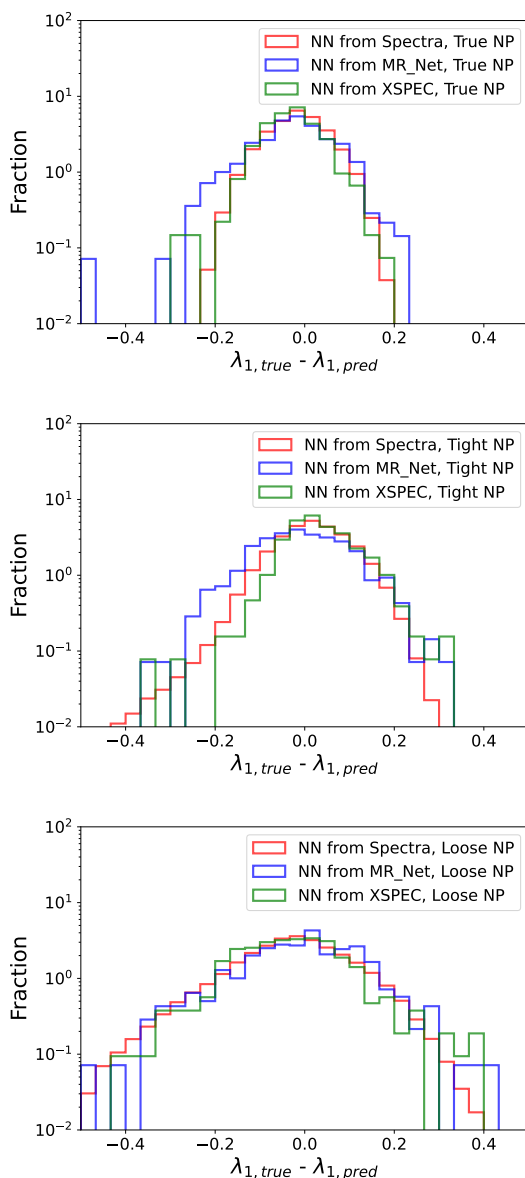


Figure 18. Performance of the regression of neutron star EOS parameter λ_1 using direct regression from spectra, as compared to regression from mass and radius information extracted via MR_Net or XSPEC. Shown are the residual distributions, the difference between the true and predicted values, under three scenarios of nuisance parameter uncertainties. See table 4 for quantitative analysis. In the “true” case, the NPs are fixed to their true values; in the “tight” and “loose” cases, they are drawn from narrow or wide priors, respectively (see text for details).

As a further visualization, figure 21 shows several example curves in the mass-radius plane fitted to the same stellar spectra with varying nuisance parameters.

9 Conclusions

We have demonstrated the network regression of EOS parameters from realistic neutron star mass and radius estimates drawn from simulated stellar spectra. Our approach of condition-

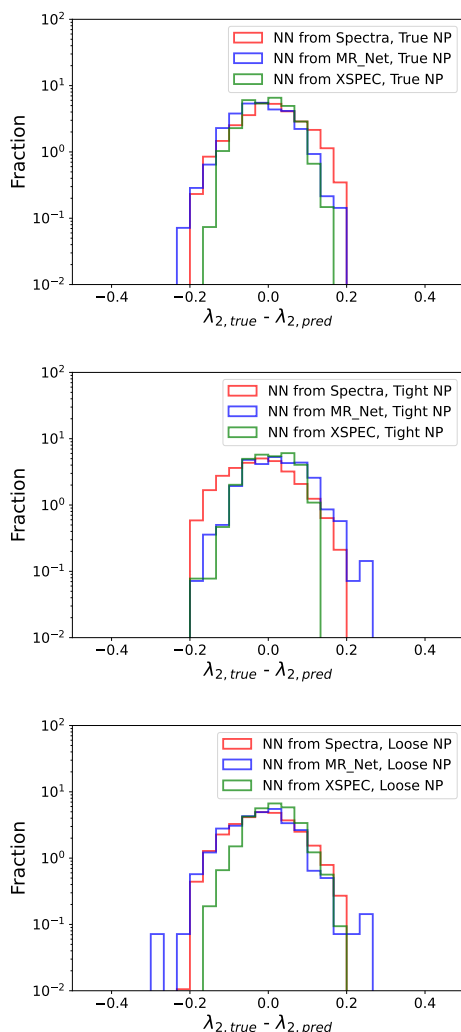


Figure 19. Performance of the regression of neutron star EOS parameter λ_2 using direct regression from spectra, as compared to regression from mass and radius information extracted via MR_Net or XSPEC. Shown are the residual distributions, the difference between the true and predicted values, under three scenarios of nuisance parameter uncertainties. See table 4 for quantitative analysis. In the “true” case, the NPs are fixed to their true values; in the “tight” and “loose” cases, they are drawn from narrow or wide priors, respectively (see text for details).

ing each step on nuisance parameters allows us to fold in the NP uncertainty via multiple sampling from priors and permits full propagation of the uncertainty through to the final regression targets. The full propagation is important because variation in NPs does not produce variations in the mass and radius of neutron stars that can be accurately summarised as two dimensional uncorrelated Gaussians (see figure 14), as has been assumed in previous studies. In addition, we have shown that networks can analyze high-dimensional telescope data directly, including sets of multiple stars, and achieve comparable performance to methods that assume perfect knowledge of the theoretical model used to generate the simulated samples. In realistic cases where the nuisance parameter uncertainties are significant, the proposed end-to-end network regression achieves comparable precision in EOS regression to the network using XSPEC fits or MR_Net predictions.

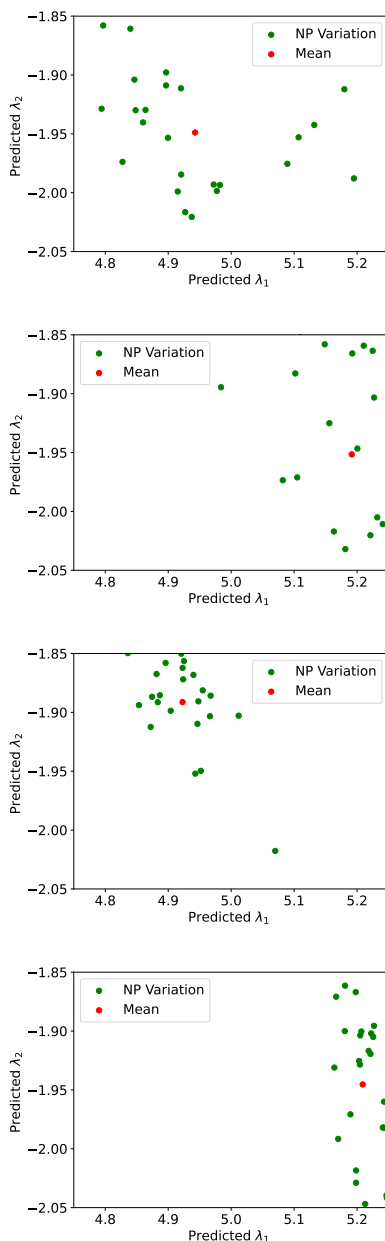


Figure 20. Neural network regression of the EOS parameters λ_1 and λ_2 of a set of 10 neutron stars directly from the set of stellar spectra. Each pane represents an example dataset of 10 simulated stars, and shown (green) are EOS estimates for several independent values of the stellar nuisance parameters drawn from the associated priors, and the mean value (red). Top two cases have loose priors, bottom two have tight.

These results suggest many future directions. Our networks are parameterized in the nuisance parameters, allowing for the propagation of prior uncertainties which are implicitly derived from auxiliary data. But the stellar spectra may also offer information that constrains the NP uncertainty. Profiling over the nuisance parameters could reduce this uncertainty, though it may be computationally very expensive without neural likelihood estimation techniques [73].

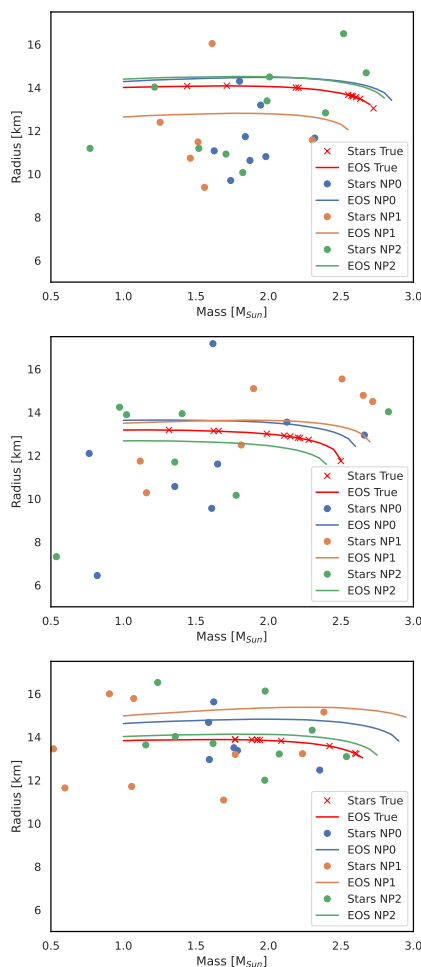


Figure 21. Demonstration of the impact of nuisance parameters on the regression of EOS parameters for three sets of observed stellar spectra. In red are the true mass and radius of the stars, drawn from the mass-radius curve determined by the true EOS parameters. In blue dots are the values of the mass and radius deduced by MR_Net for each star given a set of stellar NPs drawn from the priors; the blue line shows the mass-radius curve corresponding to the EOS parameters deduced directly from the stellar spectra and NPs by the proposed end-to-end regression. The results of MR_Net are not used in the EOS regression, and only appear to aid the visualization. Brown and green are similar to blue, but for independent draws of the NPs from the same stellar priors.

Alternatively, rather than employing regression to directly produce estimates of the EOS parameters, one might train a generative model to operate as a surrogate of the likelihood [74], allowing for fast evaluation of the likelihood as a function of the EOS parameters and potentially direct profiling.

To reduce the impact on one particular set of theoretical assumptions, such networks may be trained on a collection of theoretical models, and in the future when more telescope data becomes available, even trained directly from observed spectra.

Other future directions for the networks described in this paper would be to test more exotic neutron star equations of state, including those with phase transitions. Including additional parameters will play a key role in conducting similar research using alternative models in XSPEC that rely on different nuisance parameters, like a Helium atmospheric model.

Even more interesting would be extending this type of EOS inference to other compact objects like white dwarfs.

Acknowledgments

DW and AG are supported by The Department of Energy Office of Science. LL was supported by NSF Grant No. 2012857 to the University of California at San Diego. AWS was supported by NSF AST 19-09490, PHY 21-16686, and the Department of Energy Office of Nuclear Physics. DF and FW are supported by the National Science Foundation (U.S.A.) under Grant No. PHY-2012152. The authors are grateful to C.O. Heinke and W.C.G. Ho for their assistance in understanding and navigating the XSPEC software.

References

- [1] L. Tolos and L. Fabbietti, *Strangeness in Nuclei and Neutron Stars*, *Prog. Part. Nucl. Phys.* **112** (2020) 103770 [[arXiv:2002.09223](#)] [[INSPIRE](#)].
- [2] J.J. Li, A. Sedrakian and F. Weber, *Competition between delta isobars and hyperons and properties of compact stars*, *Phys. Lett. B* **783** (2018) 234 [[arXiv:1803.03661](#)] [[INSPIRE](#)].
- [3] W.M. Spinella and F. Weber, *Hyperonic Neutron Star Matter in Light of GW170817*, *Astron. Nachr.* **340** (2019) 145 [[arXiv:1812.03600](#)] [[INSPIRE](#)].
- [4] K. Fukushima and C. Sasaki, *The phase diagram of nuclear and quark matter at high baryon density*, *Prog. Part. Nucl. Phys.* **72** (2013) 99 [[arXiv:1301.6377](#)] [[INSPIRE](#)].
- [5] M. Orsaria, H. Rodrigues, F. Weber and G.A. Contrera, *Quark deconfinement in high-mass neutron stars*, *Phys. Rev. C* **89** (2014) 015806 [[arXiv:1308.1657](#)] [[INSPIRE](#)].
- [6] M.G. Alford, A. Schmitt, K. Rajagopal and T. Schäfer, *Color superconductivity in dense quark matter*, *Rev. Mod. Phys.* **80** (2008) 1455 [[arXiv:0709.4635](#)] [[INSPIRE](#)].
- [7] J.L. Zdunik and P. Haensel, *Maximum mass of neutron stars and strange neutron-star cores*, *Astron. Astrophys.* **551** (2013) A61 [[arXiv:1211.1231](#)] [[INSPIRE](#)].
- [8] G. Baym, *Pion condensation in nuclear and neutron star matter*, *Phys. Rev. Lett.* **30** (1973) 1340 [[INSPIRE](#)].
- [9] D.B. Kaplan and A.E. Nelson, *Strange Goings on in Dense Nucleonic Matter*, *Phys. Lett. B* **175** (1986) 57 [[INSPIRE](#)].
- [10] P.J. Ellis, R. Knorren and M. Prakash, *Kaon condensation in neutron star matter with hyperons*, *Phys. Lett. B* **349** (1995) 11 [[nucl-th/9502033](#)] [[INSPIRE](#)].
- [11] J.L. Friedman and J.R. Ipser, *On the maximum mass of a uniformly rotating neutron star*, *Astrophys. J.* **314** (1987) 594.
- [12] R.E. Rutledge, L. Bildsten, E.F. Brown, G.G. Pavlov and V.E. Zavlin, *The Thermal X-Ray Spectra of Centaurus X-4, Aquila X-1, and 4U 1608-522 in Quiescence*, *Astrophys. J.* **514** (1999) 945 [[astro-ph/9810288](#)] [[INSPIRE](#)].
- [13] C.O. Heinke, G.B. Rybicki, R. Narayan and J.E. Grindlay, *A Hydrogen atmosphere spectral model applied to the neutron star X7 in the globular cluster 47 Tucanae*, *Astrophys. J.* **644** (2006) 1090 [[astro-ph/0506563](#)] [[INSPIRE](#)].
- [14] J.M. Lattimer and M. Prakash, *Neutron star structure and the equation of state*, *Astrophys. J.* **550** (2001) 426 [[astro-ph/0002232](#)] [[INSPIRE](#)].
- [15] L. Lindblom, *Spectral Representations of Neutron-Star Equations of State*, *Phys. Rev. D* **82** (2010) 103011 [[arXiv:1009.0738](#)] [[INSPIRE](#)].

- [16] A.W. Steiner, J.M. Lattimer and E.F. Brown, *The Equation of State from Observed Masses and Radii of Neutron Stars*, *Astrophys. J.* **722** (2010) 33 [[arXiv:1005.0811](#)] [[INSPIRE](#)].
- [17] L. Lindblom and N.M. Indik, *Spectral Approach to the Relativistic Inverse Stellar Structure Problem. Part II*, *Phys. Rev. D* **89** (2014) 064003 [[arXiv:1310.0803](#)] [*Erratum ibid.* **93** (2016) 129903] [[INSPIRE](#)].
- [18] Y. Fujimoto, K. Fukushima and K. Murase, *Mapping neutron star data to the equation of state using the deep neural network*, *Phys. Rev. D* **101** (2020) 054016 [[arXiv:1903.03400](#)] [[INSPIRE](#)].
- [19] Y. Fujimoto, K. Fukushima and K. Murase, *Extensive Studies of the Neutron Star Equation of State from the Deep Learning Inference with the Observational Data Augmentation*, *JHEP* **03** (2021) 273 [[arXiv:2101.08156](#)] [[INSPIRE](#)].
- [20] F. Morawski and M. Bejger, *Neural network reconstruction of the dense matter equation of state derived from the parameters of neutron stars*, *Astron. Astrophys.* **642** (2020) A78 [[arXiv:2006.07194](#)] [[INSPIRE](#)].
- [21] M. Ferreira and C. Providência, *Unveiling the nuclear matter EoS from neutron star properties: a supervised machine learning approach*, *JCAP* **07** (2021) 011 [[arXiv:1910.05554](#)] [[INSPIRE](#)].
- [22] P. Baldi, *Deep Learning in Science*, Cambridge University Press, Cambridge, U.K. (2021).
- [23] P. Baldi, P. Sadowski and D. Whiteson, *Searching for Exotic Particles in High-Energy Physics with Deep Learning*, *Nature Commun.* **5** (2014) 4308 [[arXiv:1402.4735](#)] [[INSPIRE](#)].
- [24] A. Ghosh, B. Nachman and D. Whiteson, *Uncertainty-aware machine learning for high energy physics*, *Phys. Rev. D* **104** (2021) 056026 [[arXiv:2105.08742](#)] [[INSPIRE](#)].
- [25] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski and D. Whiteson, *Parameterized neural networks for high-energy physics*, *Eur. Phys. J. C* **76** (2016) 235 [[arXiv:1601.07913](#)] [[INSPIRE](#)].
- [26] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban and D. Whiteson, *Jet Flavor Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev. D* **94** (2016) 112002 [[arXiv:1607.08633](#)] [[INSPIRE](#)].
- [27] S. Bogdanov, C.O. Heinke, F. Özel and T. Güver, *Neutron Star Mass-Radius Constraints of the Quiescent Low-mass X-ray Binaries X7 and X5 in the Globular Cluster 47 Tuc*, *Astrophys. J.* **831** (2016) 184 [[arXiv:1603.01630](#)] [[INSPIRE](#)].
- [28] K.A. Arnaud, *XSPEC: The First Ten Years*, in *Astronomical Data Analysis Software and Systems V*, *Astronomical Society of the Pacific Conference Series* **101**, G.H. Jacoby and J. Barnes eds., Astronomical Society of the Pacific (1996).
- [29] J.M. Lattimer, *Neutron Star Mass and Radius Measurements*, *Universe* **5** (2019) 159 [[INSPIRE](#)].
- [30] B. Müller and J.L. Nagle, *Results from the relativistic heavy ion collider*, *Ann. Rev. Nucl. Part. Sci.* **56** (2006) 93 [[nucl-th/0602029](#)] [[INSPIRE](#)].
- [31] ATLAS collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, 2008 *JINST* **3** S08003 [[INSPIRE](#)].
- [32] P. Spiller and G. Franchetti, *The FAIR accelerator project at GSI*, *Nucl. Instrum. Meth. A* **561** (2006) 305 [[INSPIRE](#)].
- [33] X. Roca-Maza, M. Centelles, X. Viñas and M. Warda, *Neutron skin of ^{208}Pb , nuclear symmetry energy, and the parity radius experiment*, *Phys. Rev. Lett.* **106** (2011) 252501 [[arXiv:1103.1762](#)] [[INSPIRE](#)].
- [34] N.K. Glendenning, *Compact stars: Nuclear physics, particle physics, and general relativity*, in *Astronomy and Astrophysics Library*, Springer, New York, NY, U.S.A. (1997) [[INSPIRE](#)].
- [35] A.L. Watts et al., *Colloquium: Measuring the neutron star equation of state using x-ray timing*, *Rev. Mod. Phys.* **88** (2016) 021001 [[arXiv:1602.01081](#)] [[INSPIRE](#)].

- [36] R. Machleidt and D.R. Entem, *Chiral effective field theory and nuclear forces*, *Phys. Rept.* **503** (2011) 1 [[arXiv:1105.2919](#)] [[INSPIRE](#)].
- [37] E.S. Fraga, A. Kurkela and A. Vuorinen, *Neutron star structure from QCD*, *Eur. Phys. J. A* **52** (2016) 49 [[arXiv:1508.05019](#)] [[INSPIRE](#)].
- [38] J.R. Oppenheimer and G.M. Volkoff, *On massive neutron cores*, *Phys. Rev.* **55** (1939) 374 [[INSPIRE](#)].
- [39] R.C. Tolman, *Static solutions of Einstein's field equations for spheres of fluid*, *Phys. Rev.* **55** (1939) 364 [[INSPIRE](#)].
- [40] S. Campana, M. Colpi, S. Mereghetti, L. Stella and M. Tavani, *The neutron stars of soft x-ray transients*, *Astron. Astrophys. Rev.* **8** (1998) 279 [[astro-ph/9805079](#)] [[INSPIRE](#)].
- [41] A.Y. Potekhin, *Atmospheres and radiating surfaces of neutron stars*, *Phys. Usp.* **57** (2014) 735 [[arXiv:1403.0074](#)] [[INSPIRE](#)].
- [42] C.O. Heinke et al., *Analysis of the quiescent low-mass x-ray binary population in galactic globular clusters*, *Astrophys. J.* **598** (2003) 501 [[astro-ph/0305445](#)] [[INSPIRE](#)].
- [43] E.F. Brown, L. Bildsten and R.E. Rutledge, *Crustal heating and quiescent emission from transiently accreting neutron stars*, *Astrophys. J. Lett.* **504** (1998) L95 [[astro-ph/9807179](#)] [[INSPIRE](#)].
- [44] K. Hebeler, J.M. Lattimer, C.J. Pethick and A. Schwenk, *Equation of state and neutron star properties constrained by nuclear physics and observation*, *Astrophys. J.* **773** (2013) 11 [[arXiv:1303.4662](#)] [[INSPIRE](#)].
- [45] A.W. Steiner et al., *Constraining the Mass and Radius of Neutron Stars in Globular Clusters*, *Mon. Not. Roy. Astron. Soc.* **476** (2018) 421 [[arXiv:1709.05013](#)] [[INSPIRE](#)].
- [46] P. Slane, D.J. Helfand and S.S. Murray, *New constraints on neutron star cooling from chandra observations of 3C 58*, *Astrophys. J. Lett.* **571** (2002) L45 [[astro-ph/0204151](#)] [[INSPIRE](#)].
- [47] R. Wijnands, N. Degenaar and D. Page, *Cooling of Accretion-Heated Neutron Stars*, *J. Astrophys. Astron.* **38** (2017) 49 [[arXiv:1709.07034](#)] [[INSPIRE](#)].
- [48] H. Sun et al., *A unified binary neutron star merger magnetar model for the Chandra X-ray transients CDF-S XT1 and XT2*, *Astrophys. J.* **886** (2019) 129 [[arXiv:1908.01107](#)] [[INSPIRE](#)].
- [49] B. Neyshabur, S. Bhojanapalli, D. McAllester and N. Srebro, *Exploring generalization in deep learning*, in *Advances in Neural Information Processing Systems* **30**, Curran Associates, Inc. (2017).
- [50] A. Vaswani et al., *Attention is all you need*, in *Advances in Neural Information Processing Systems* **30**, I. Guyon et al. eds., Curran Associates, Inc. (2017), pp. 5998–6008 and online pdf version at <http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf>.
- [51] P. Baldi and R. Vershynin, *The Quarks of Attention*, [arXiv:2202.08371](#).
- [52] S. Typel, G. Ropke, T. Klähn, D. Blaschke and H.H. Wolter, *Composition and thermodynamics of nuclear matter with light clusters*, *Phys. Rev. C* **81** (2010) 015803 [[arXiv:0908.2344](#)] [[INSPIRE](#)].
- [53] W.M. Spinella and F. Weber, *Dense Baryonic Matter in the Cores of Neutron Stars*, World Scientific (2020), pp. 85–152 [[DOI:10.1142/9789813277342_0004](#)] [[INSPIRE](#)].
- [54] G. Malfatti, M.G. Orsaria, G.A. Contrera, F. Weber and I.F. Ranea-Sandoval, *Hot quark matter and (proto-) neutron stars*, *Phys. Rev. C* **100** (2019) 015803 [[arXiv:1907.06597](#)] [[INSPIRE](#)].
- [55] M. Oertel, M. Hempel, T. Klähn and S. Typel, *Equations of state for supernovae and compact stars*, *Rev. Mod. Phys.* **89** (2017) 015007 [[arXiv:1610.03361](#)] [[INSPIRE](#)].

- [56] S. Shlomo, V.M. Kolomietz and G. Colò, *Deducing the nuclear-matter incompressibility coefficient from data on isoscalar compression modes*, *Eur. Phys. J. A* **30** (2006) 23.
- [57] U. Garg and G. Colò, *The compression-mode giant resonances and nuclear incompressibility*, *Prog. Part. Nucl. Phys.* **101** (2018) 55 [[arXiv:1801.03672](#)] [[INSPIRE](#)].
- [58] J.R. Stone, N.J. Stone and S.A. Moszkowski, *Incompressibility in finite nuclei and nuclear matter*, *Phys. Rev. C* **89** (2014) 044316 [[arXiv:1404.0744](#)] [[INSPIRE](#)].
- [59] G. Baym, C.J. Pethick and P. Sutherland, *The Ground state of matter at high densities: Equation of state and stellar models*, *Astrophys. J.* **170** (1971) 299 [[INSPIRE](#)].
- [60] G. Baym, H.A. Bethe and C.J. Pethick, *Neutron star matter*, *Nucl. Phys. A* **175** (1971) 225 [[INSPIRE](#)].
- [61] L. Lindblom, *Causal Representations of Neutron-Star Equations of State*, *Phys. Rev. D* **97** (2018) 123019 [[arXiv:1804.04072](#)] [[INSPIRE](#)].
- [62] B. Fornberg and J. Zuev, *The runge phenomenon and spatially variable shape parameters in RBF interpolation*, *Comput. Math. Appl.* **54** (2007) 379.
- [63] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes. The Art of Scientific Computing*, 3rd edition, Cambridge University Press (2007).
- [64] J.M. Lattimer and A.W. Steiner, *Neutron Star Masses and Radii from Quiescent Low-Mass X-ray Binaries*, *Astrophys. J.* **784** (2014) 123 [[arXiv:1305.3242](#)] [[INSPIRE](#)].
- [65] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, [arXiv:1412.6980](#) [[INSPIRE](#)].
- [66] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *J. Mach. Learn. Res.* **12** (2011) 2825 [[arXiv:1201.0490](#)] [[INSPIRE](#)].
- [67] K. Elshamouty, C. Heinke, S. Morsink, S. Bogdanov and A. Stevens, *The impact of surface temperature inhomogeneities on quiescent neutron star radius measurements*, *Astrophys. J.* **826** (2016) 162 [[arXiv:1605.09400](#)] [[INSPIRE](#)].
- [68] M.J. Fenton, A. Shmakov, T.-W. Ho, S.-C. Hsu, D. Whiteson and P. Baldi, *Permutationless many-jet event reconstruction with symmetry preserving attention networks*, *Phys. Rev. D* **105** (2022) 112008 [[arXiv:2010.09206](#)] [[INSPIRE](#)].
- [69] A. Shmakov, M.J. Fenton, T.-W. Ho, S.-C. Hsu, D. Whiteson and P. Baldi, *SPANet: Generalized permutationless set assignment for particle physics using symmetry preserving attention*, *SciPost Phys.* **12** (2022) 178 [[arXiv:2106.03898](#)] [[INSPIRE](#)].
- [70] L. Hertel, J. Collado, P. Sadowski, J. Ott and P. Baldi, *Sherpa: Robust hyperparameter optimization for machine learning*, *SoftwareX* **12** (2020) 100591 [[arXiv:2005.04048](#)].
- [71] J.N. Howard, S. Mandt, D. Whiteson and Y. Yang, *Learning to simulate high energy particle collisions from unlabeled data*, *Sci. Rep.* **12** (2022) 7567 [[arXiv:2101.08944](#)] [[INSPIRE](#)].
- [72] A. Ghosh, X. Ju, B. Nachman and A. Siodmok, *Towards a deep learning model for hadronization*, *Phys. Rev. D* **106** (2022) 096020 [[arXiv:2203.12660](#)] [[INSPIRE](#)].
- [73] L. Heinrich, *Learning Optimal Test Statistics in the Presence of Nuisance Parameters*, [arXiv:2203.13079](#) [[INSPIRE](#)].
- [74] K. Cranmer, J. Brehmer and G. Louppe, *The frontier of simulation-based inference*, *Proc. Nat. Acad. Sci.* **117** (2020) 30055 [[arXiv:1911.01429](#)] [[INSPIRE](#)].