Neural Network-based Inference of the NS Dense-Matter Equation of State

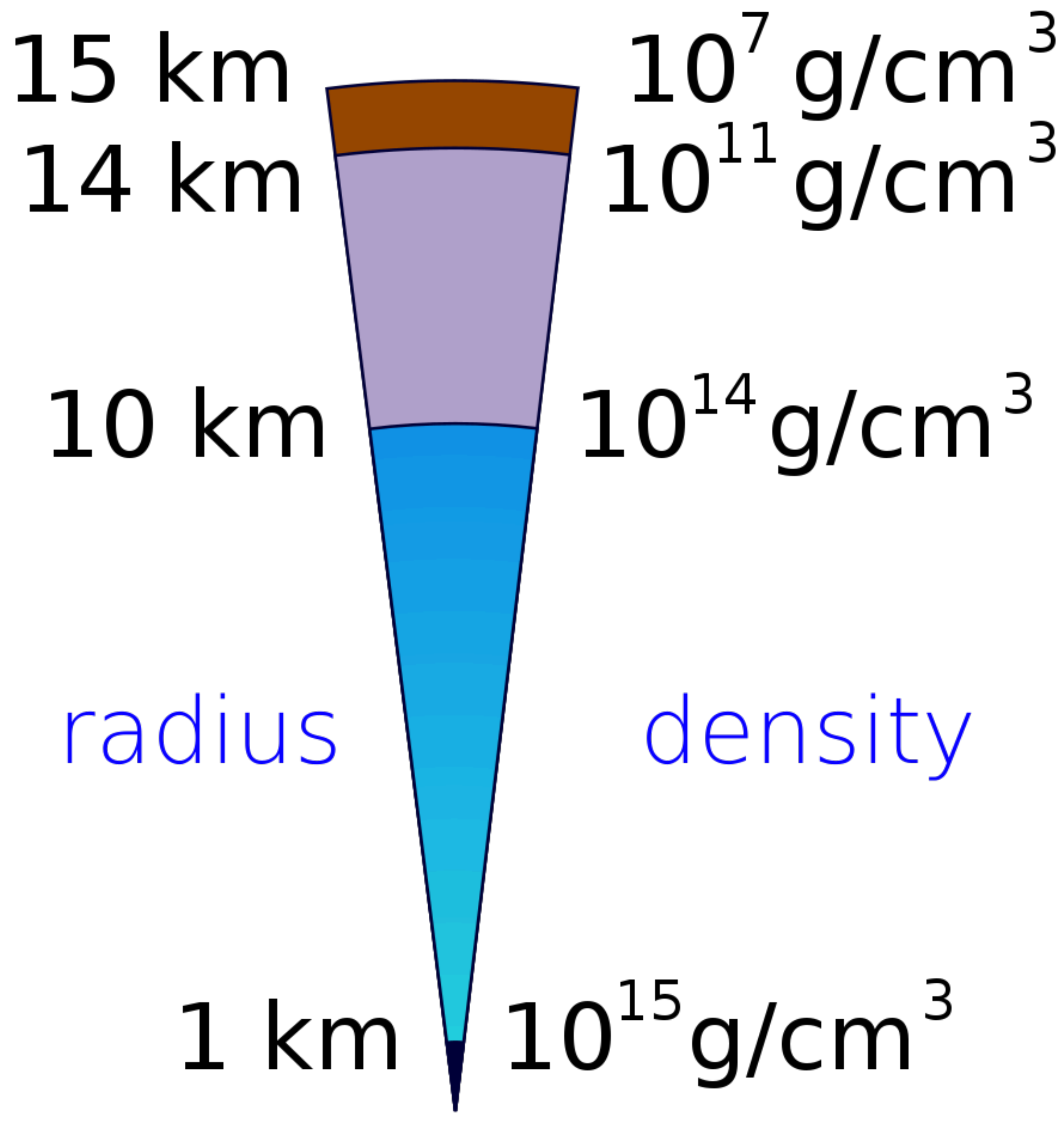
Ameya Thete[1]**,** Kinjal Banerjee[1] & Tuhin Malik[2]

[1]**BITS, Pilani - K.K. Birla Goa Campus**
[2]**CFisUC, University of Coimbra**

03 November 2022

Neutron Stars

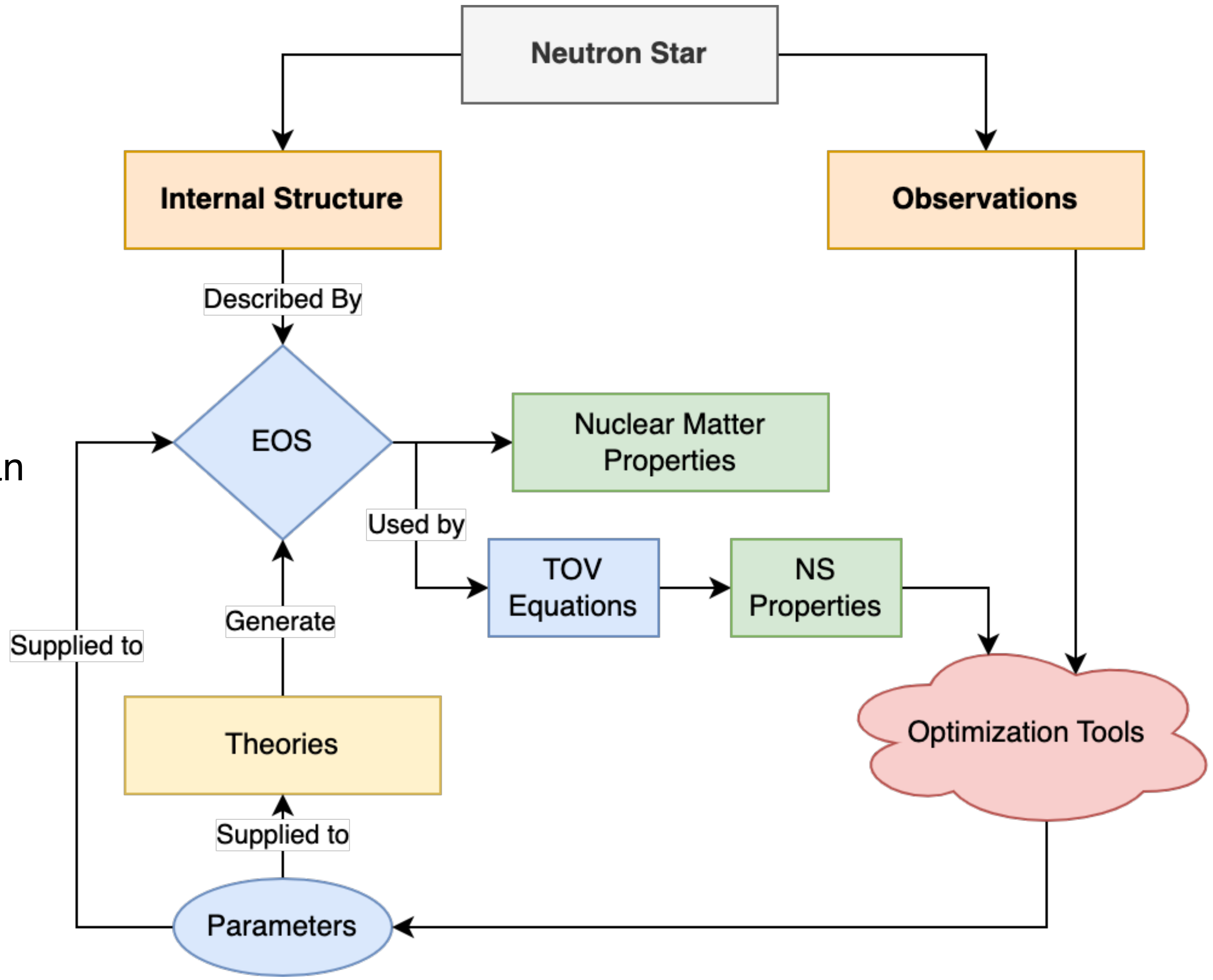
- Neutron Stars (NSs) are massive supernova remnants, formed by the gravitational collapse of stars heavier than $1.4M_{\odot}$
- First postulated by Baade and Zwicky in 1933, discovered by Hewish and Bell in 1967 via radio emissions from a pulsar.
- NSs are a degenerate gas of neutrons.
    - Pauli's exclusion principle forbids multiple neutrons from occupying the same state; only two neutrons can go into each energy level.
    - Hence, in a degenerate gas all low energy levels are filled. Neutrons have energy and are in motion => exert pressure, even at 0K
    - Hence, NSs are supported by degeneracy — this also leads to heavy compaction.
- Matter in NSs exists at extreme conditions — far beyond the kind accessible in a laboratory.



A cross-section of a neutron star, showing the layers and the matter density as a function of distance from the centre

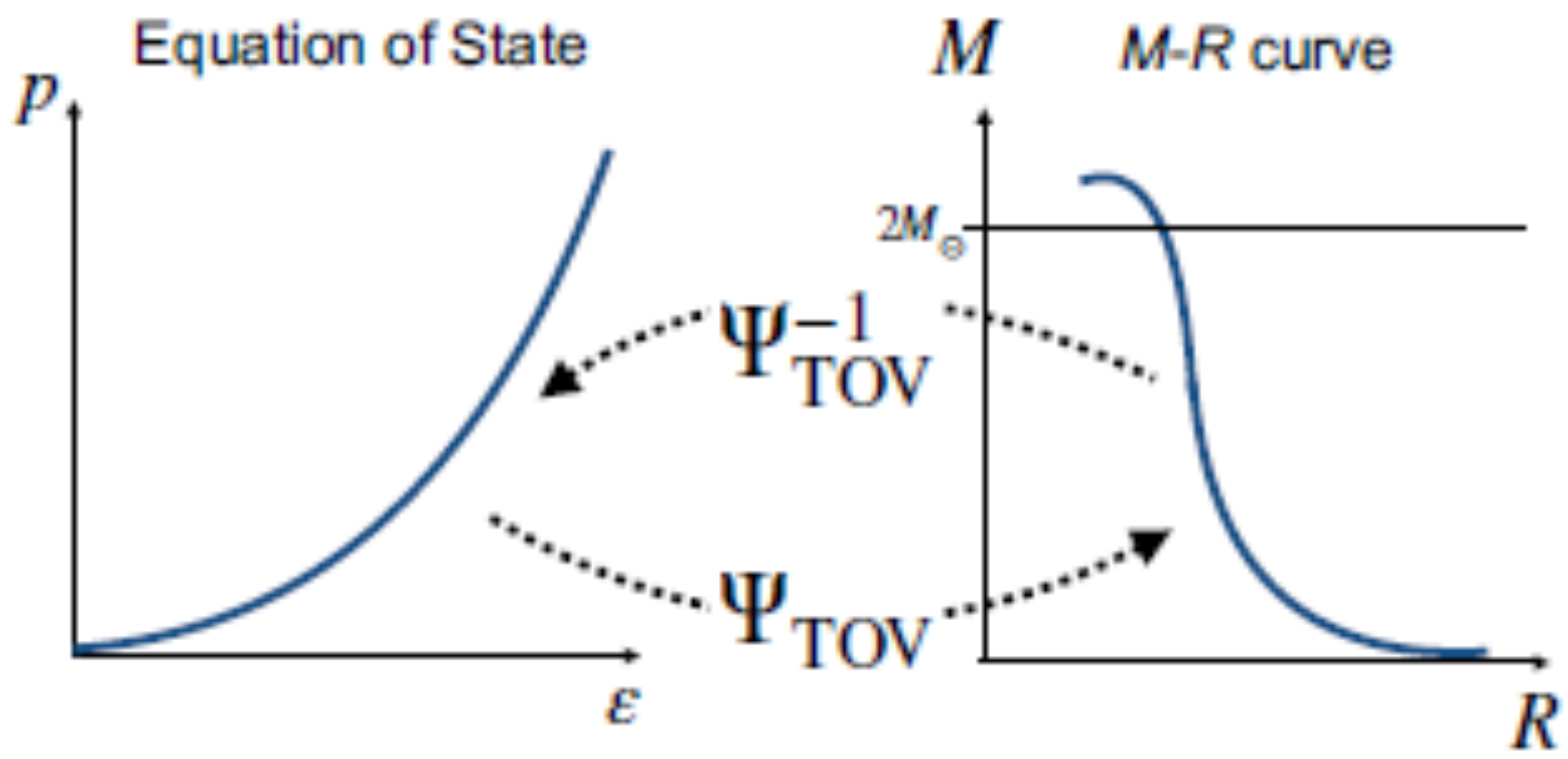
Neutron Stars

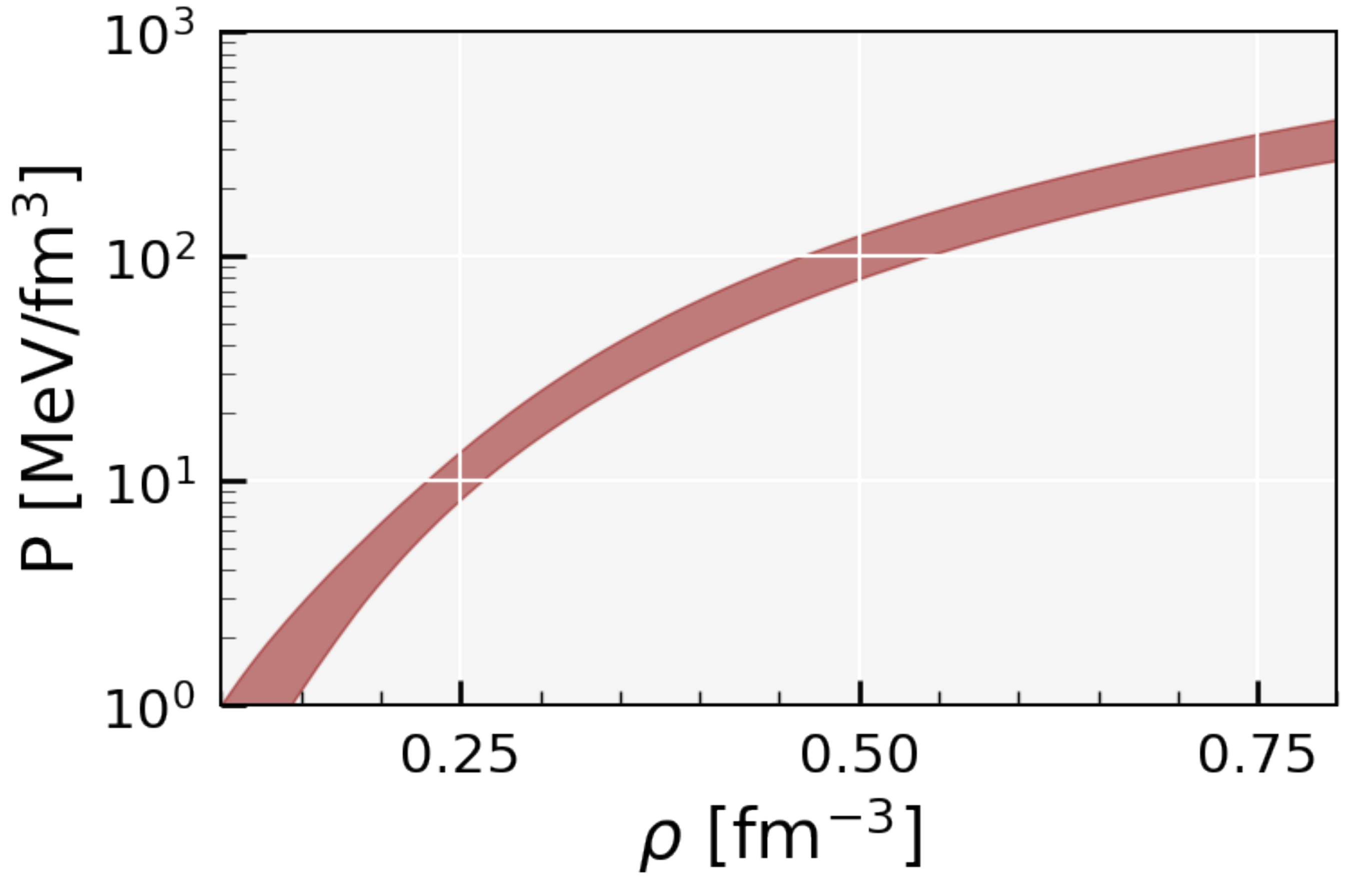
- Studies of NSs involve two different yet highly interlinked approaches: theoretical calculations of the internal structure and experimental observations of their properties, like the mass or radius.
- Since conditions on NSs cannot be replicated terrestrially, all experimental verification comes from observations made by multimessenger collaborations like NICER.
- Any neutron star is theoretically described by its equation of state (EoS): an equation relating the pressure as a function of energy.
- Given the unknown behaviour of matter at these densities, there are a multitude of theories that generate differing EoSs.
- Simultaneously, measurements of NS observables are made which help constrain the EoS.



Neutron Stars

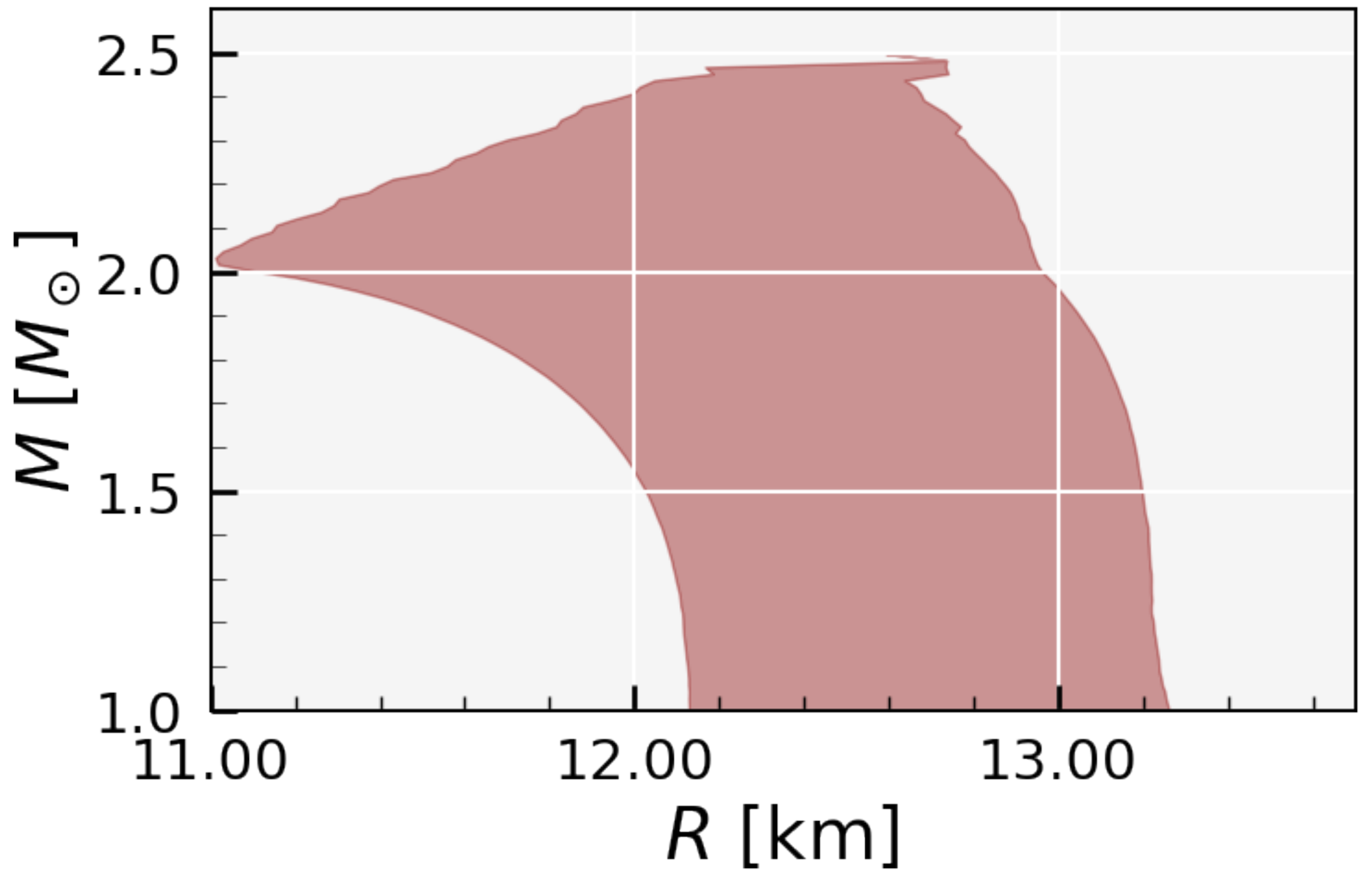
- The Tolman-Oppenheimer-Volkoff (TOV) equations can be used to generate the properties for a NS described by a given EoS
- Fortunately, GR guarantees a one-to-one correspondence between NS observables (like M-R curves or tidal deformability) and the EoS.
- Given the general tediousness of this task, is there a way to perform this conversion via an (efficient) functional map?
- Why would we want to construct such a map? Useful in statistical calculations!





EoS

Invertible!

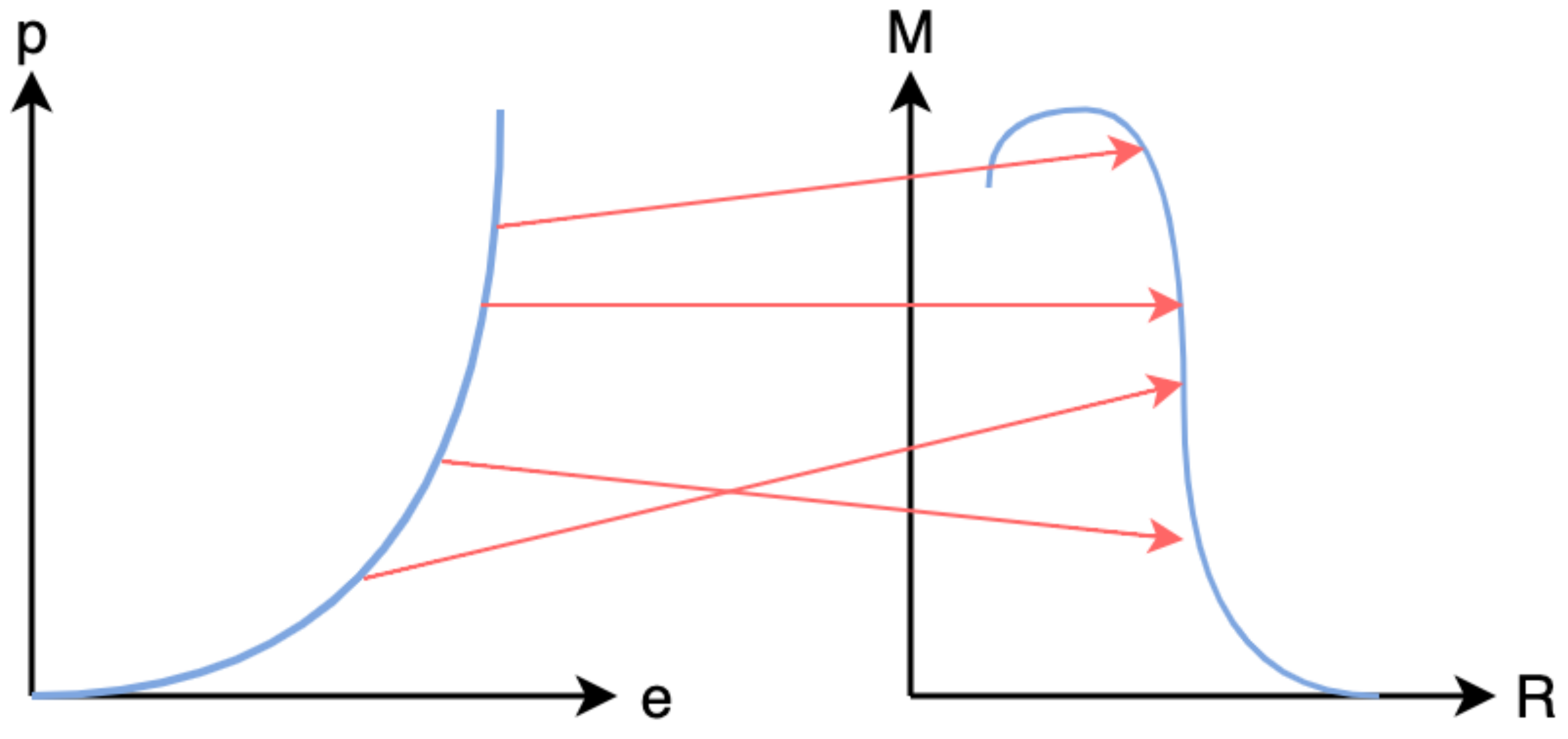


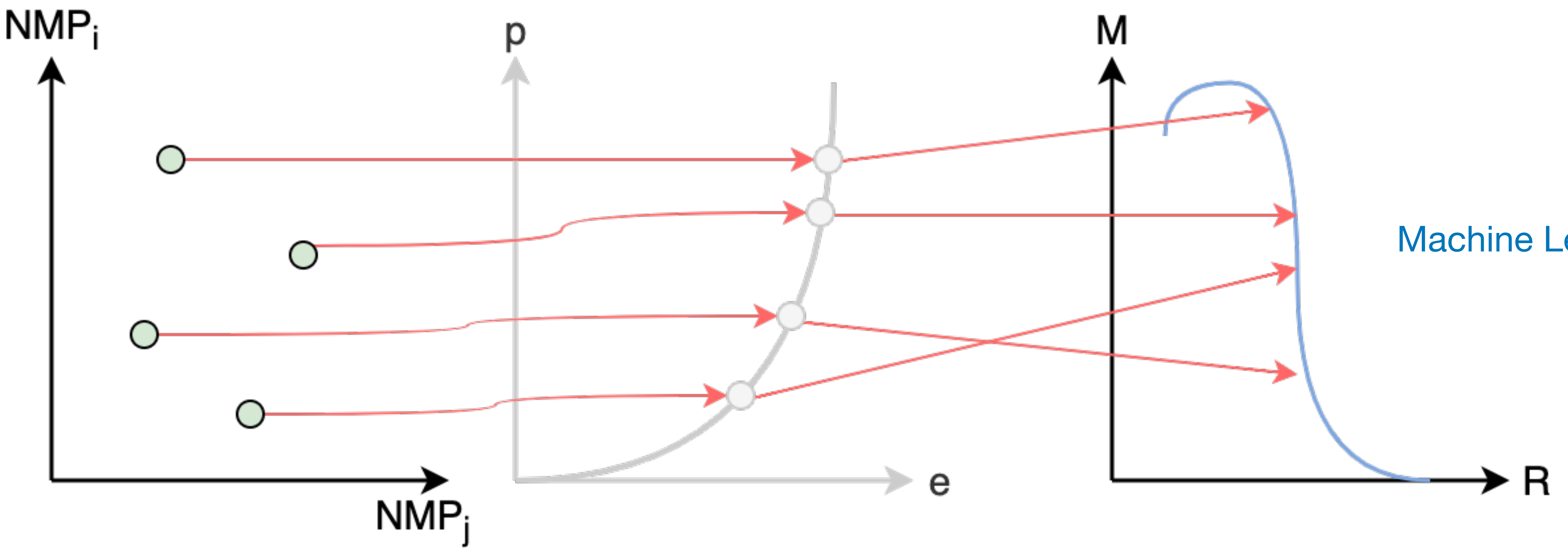
M-R Curve

Neutron Stars

- The answer is yes! We can construct a one-to-one parametric function between points on the EoS and the M-R curve and use real-world data to select the optimal map.

- But let's take this a step further. The EoS itself can be uniquely parameterised by NMPs (more on them in a bit), so we can ***completely bypass the EoS*** and construct this map from the space of these parameters to the space of NS observables.

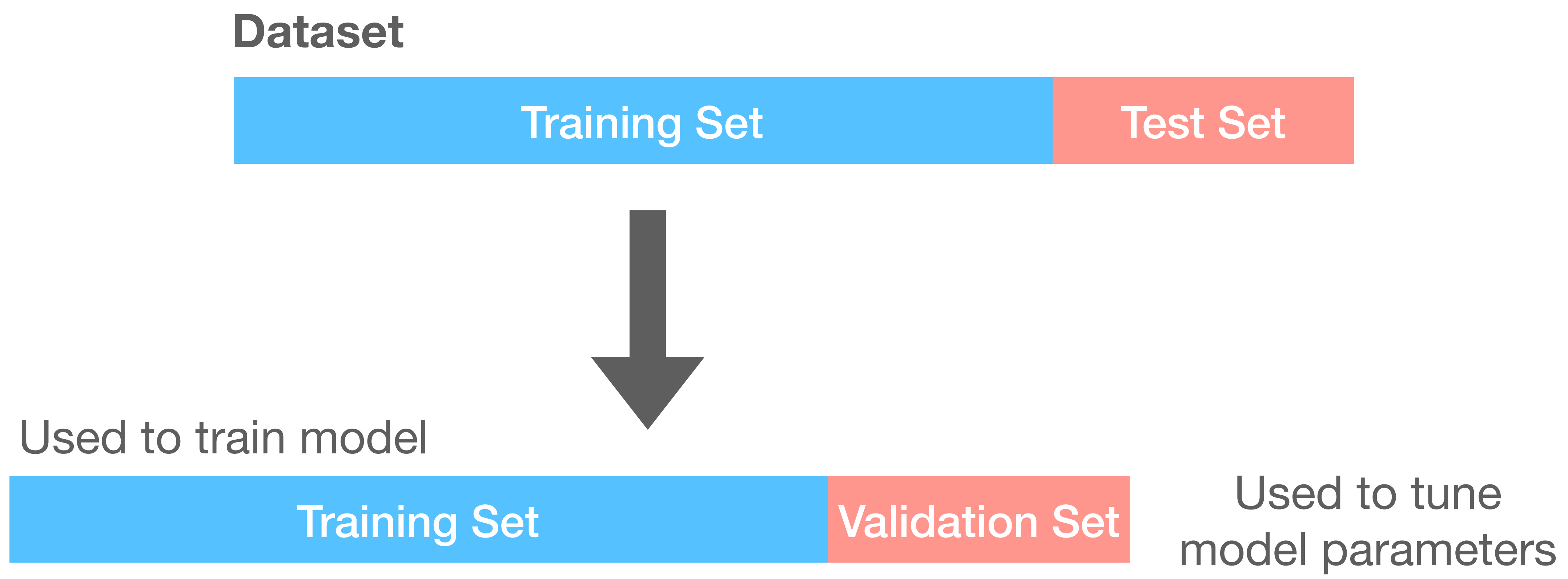


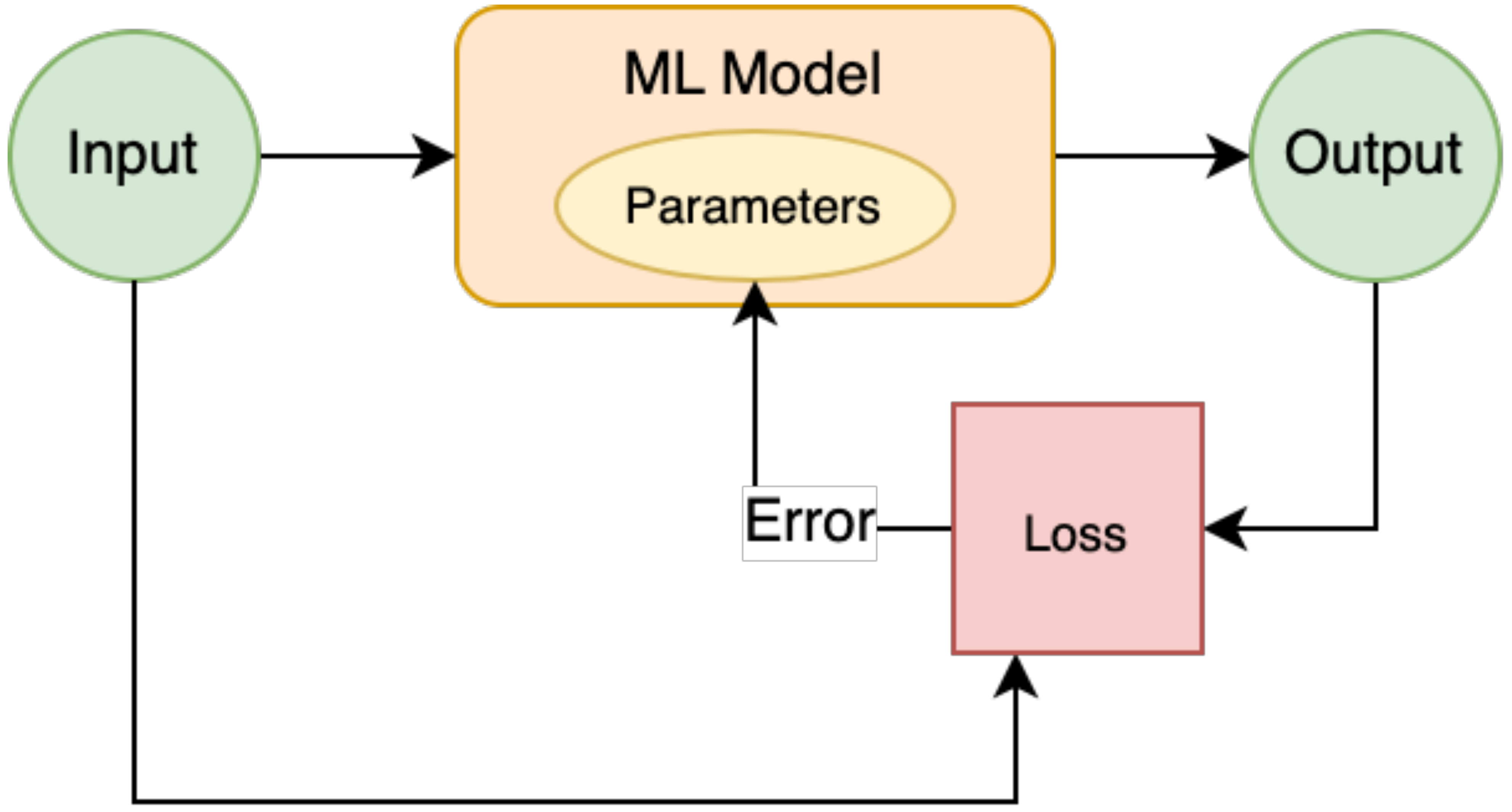


Machine Learning algorithms are useful to create such arbitrary functional maps!

Machine Learning

- Machine learning is a technique that allows us to build arbitrary functions to solve a task.
- The aim is to look at a lot of real world data and “learn” the underlying pattern by minimising the cost associated with the choice.



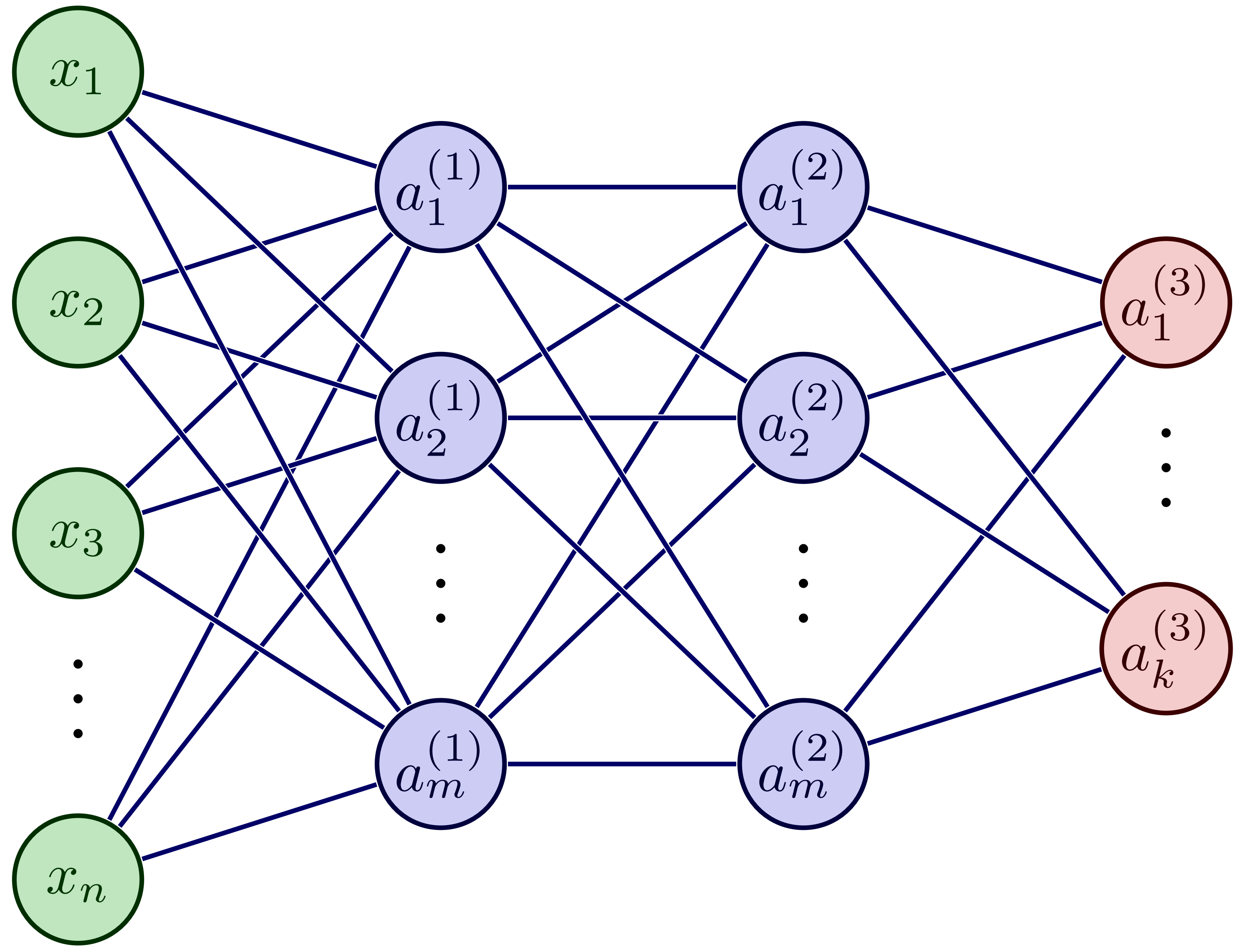


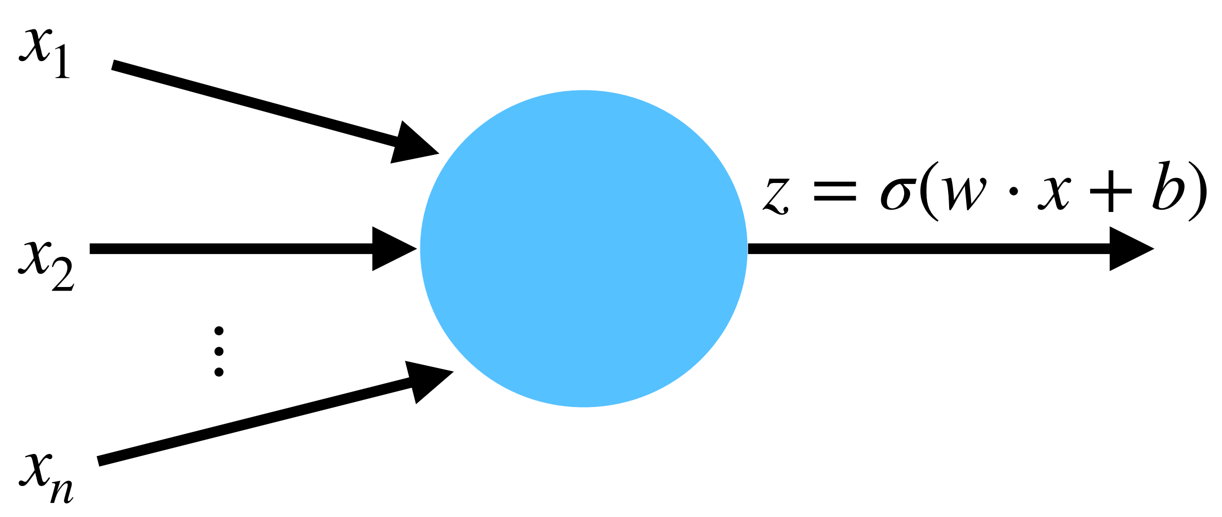
ML has some drawbacks:

- Overfitting and Underfitting
- Fails when not enough data is available

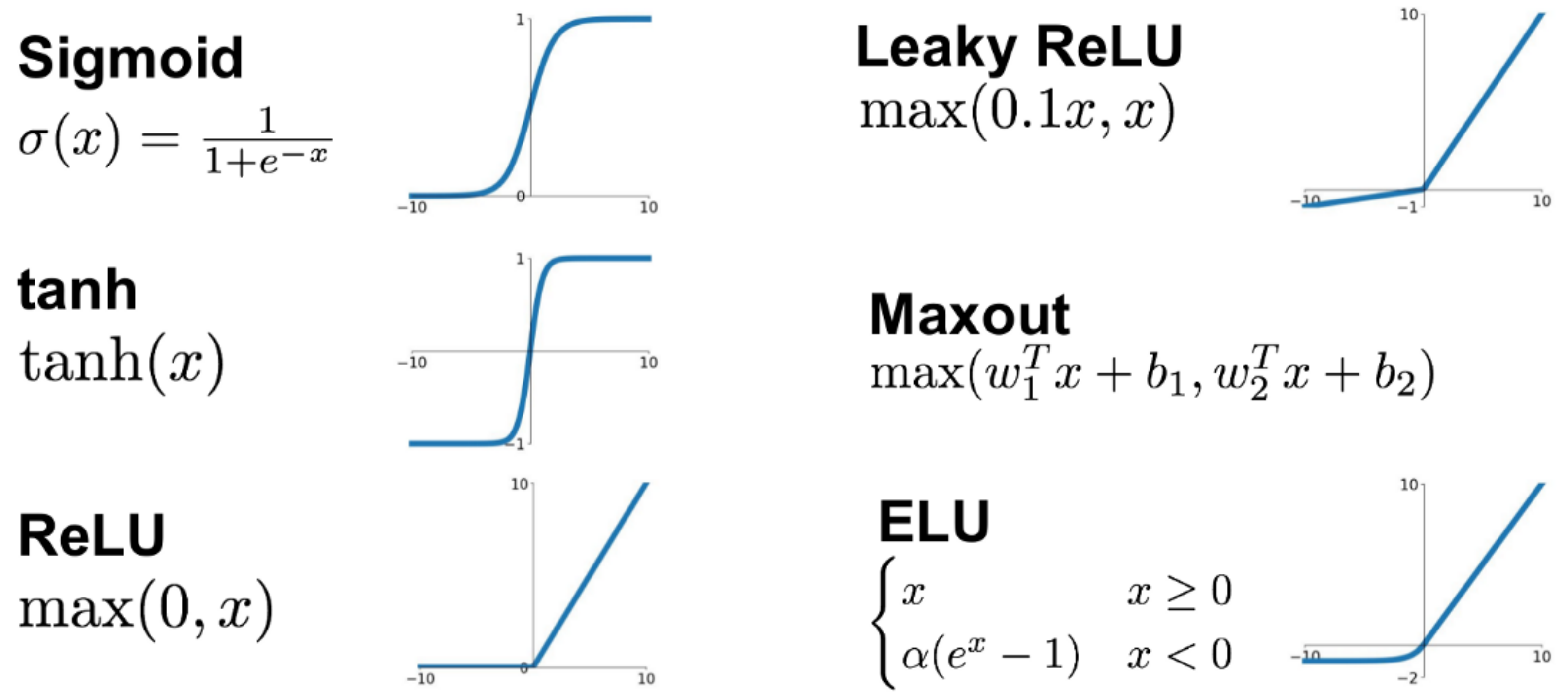
Artificial Neural Networks

- After trying out various ML algorithms, we found ANNs to be best suited for our task.
- Have the ability to learn intricate relationships from raw data.
- ANNs consist of units called neurons arranged in a stacked fashion.
  - A neuron performs a linear transformation of the input followed by the application of a non-linear function called the activation function.
  - For an input x, a neuron performs the operation $z = \sigma(w \cdot x + b)$





Activation Functions



Nuclear Matter Parameters

- Given an equation of state $p(\rho) = \rho^2 \frac{d}{d\rho}(e(\rho))$, the energy per particle at a given density ρ and asymmetry δ can be decomposed as

$$e(\rho, \delta) = e(\rho, 0) + S(\rho)\delta^2, \qquad \rho = \rho_p + \rho_n \qquad \delta = \frac{\rho_n - \rho_p}{\rho}$$

- The energy density of symmetric nuclear matter and the symmetry energy can be Taylor expanded around the saturation density, ρ_0. This gives us

$$e(\rho, 0) = e_0 + \frac{K_0}{2}\left(\frac{\rho - \rho_0}{3\rho_0}\right)^2 + \frac{Q_0}{6}\left(\frac{\rho - \rho_0}{3\rho_0}\right)^3 + \mathcal{O}(4)$$

$$S(\rho) = J_{\mathrm{sym},0} + L_{\mathrm{sym},0}\left(\frac{\rho - \rho_0}{3\rho_0}\right) + \frac{K_{\mathrm{sym},0}}{2}\left(\frac{\rho - \rho_0}{3\rho_0}\right)^2 + \mathcal{O}(3)$$

Any EoS can be well-reproduced by truncating to the third or fourth order![a]

- The collection of six Taylor coefficients along with the saturation density — called Nuclear Matter Parameters (NMPs) — can describe any EoS. Thus, the EoS is accorded a representation in the seven-dimensional parameter space of these NMPs. Symbolically,

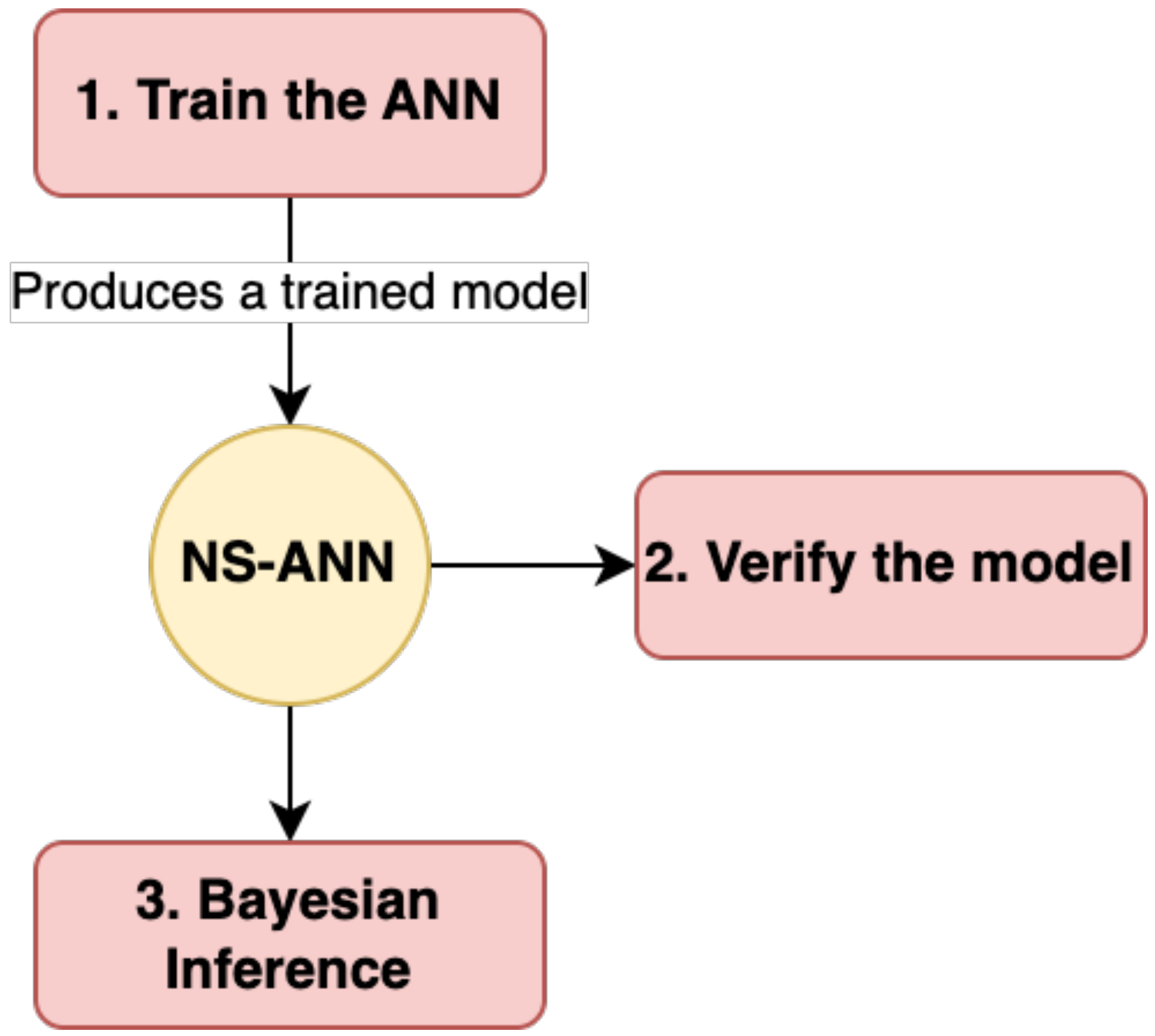
$$\begin{aligned}\mathrm{EoS}_j &= \{e_0, \rho_0, K_0, Q_0, J_{\mathrm{sym},0}, L_{\mathrm{sym},0}, \text{ and } K_{\mathrm{sym},0}\}_j \\ &\approx \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})\end{aligned}$$

[a] Phys. Rev. C 97 (2018) 025806

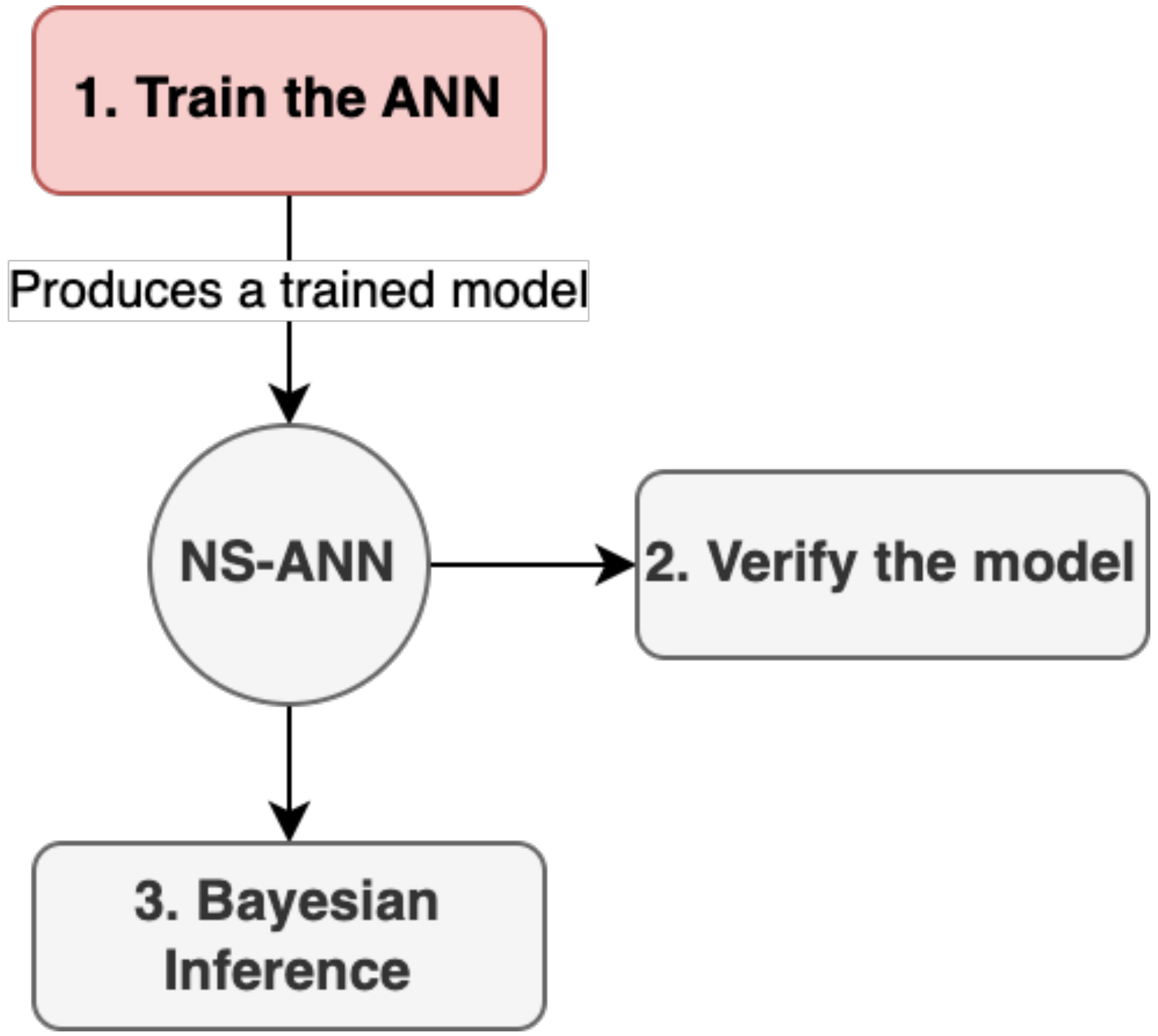
This Work

The work in our paper consists of three major sub-tasks, or “prongs”:

1. Constructing a map from the space of NMPs to the space of NS properties
2. Verifying the physical viability of the constructed map.
3. Applying the learnt map in a Bayesian setting.



Step 1: Train the ANN

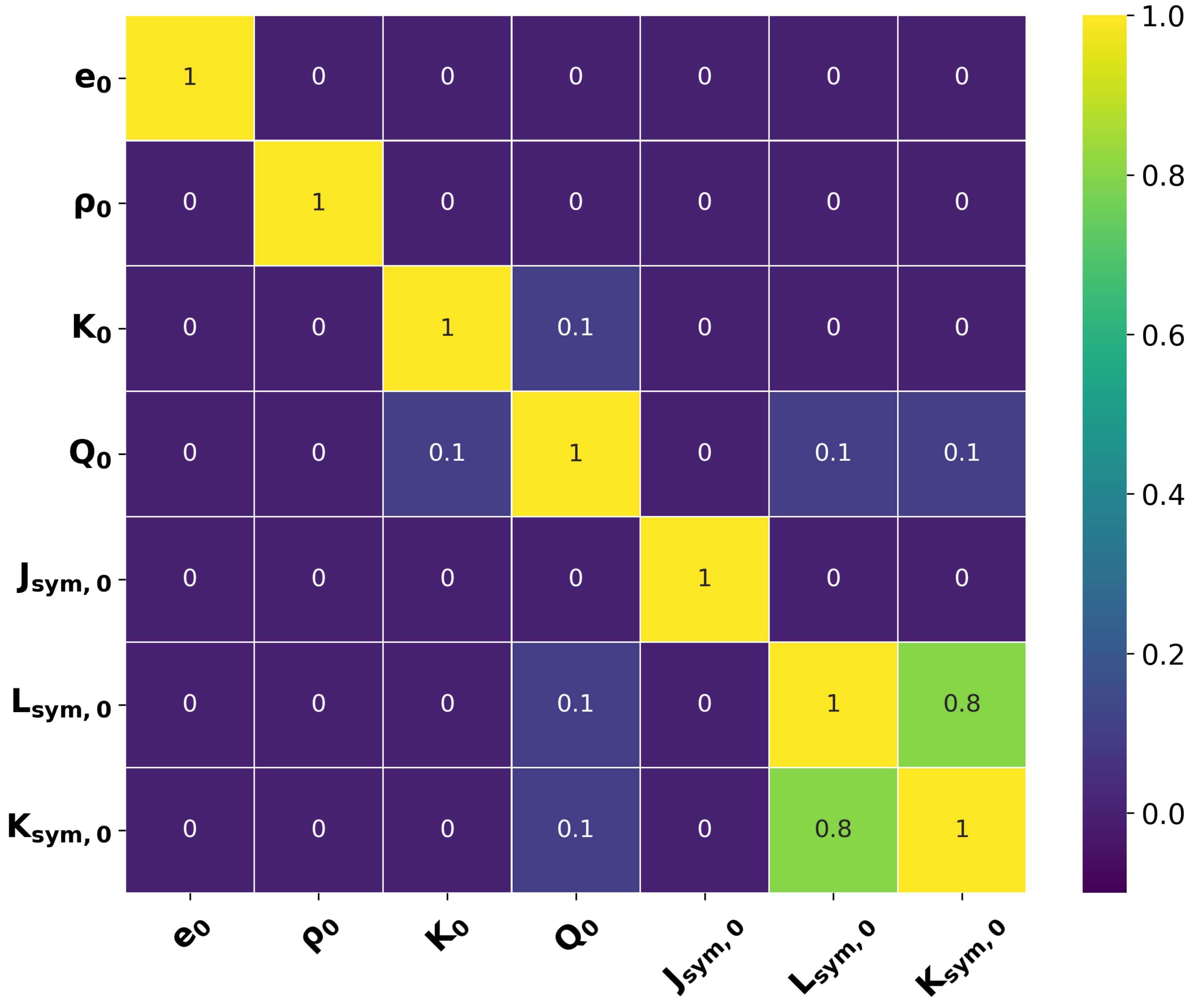


NS-ANN

- The ANN — called NS-ANN — is trained to generate neutron star observables directly from an input of NMPs.
    - Specifically, the inputs to the ANN are the seven NMPs described earlier, and the outputs are six neutron star properties: the maximum NS mass, $M_{\max}$; the maximum NS radius, $R_{\max}$; the radius for $1.4M_{\odot}$ NS, $R_{1.4}$; and the tidal deformability Λ_M for NS mass $M \in [1.0, 1.4, 1.8]M_{\odot}$.
- The motivation behind this mapping is to completely bypass any form of equation-solving. This model mimics a realistic nuclear physics model which satisfies finite nuclei properties.
- The model is a feedforward neural network with two hidden layers with 15 units in each layer.
    - ReLU activation function on both the hidden layers, no activation applied to the output
    - Parameters initialised using the Glorot uniform distribution
- ANN optimises a RMSE loss between the truth and predictions.

Dataset Generation

- We use a synthetically generated dataset to train the NS-ANN model.
- First, we construct a multivariate normal distribution of all possible EoSs, i.e., $\mathcal{N}(\mu, \Sigma)$ in the seven-dimensional NMP space.
    - We assume an *a priori* intercorrelation between $L_{sym,0}$ and $K_{sym,0}$ of 0.8 (c.f. Case-II data, Phys. Rev. C **102**, 052801(R))
- Samples are repeatedly drawn from this MVGD. The Skyrme EoS for each sample is constructed, and a check is performed to ensure that the resulting EoS
    - Predicts a $\Lambda_{1.4M_\odot} \leq 800$
    - Predicts a $M_{\max} \geq 2M_\odot$
    - Satisfies the causality condition $c_s \leq c$
- Samples that do not produces EoSs satisfying these constraints are discarded.
- For each EoS, the TOV and deformability equations are solved to obtain the NS observables.



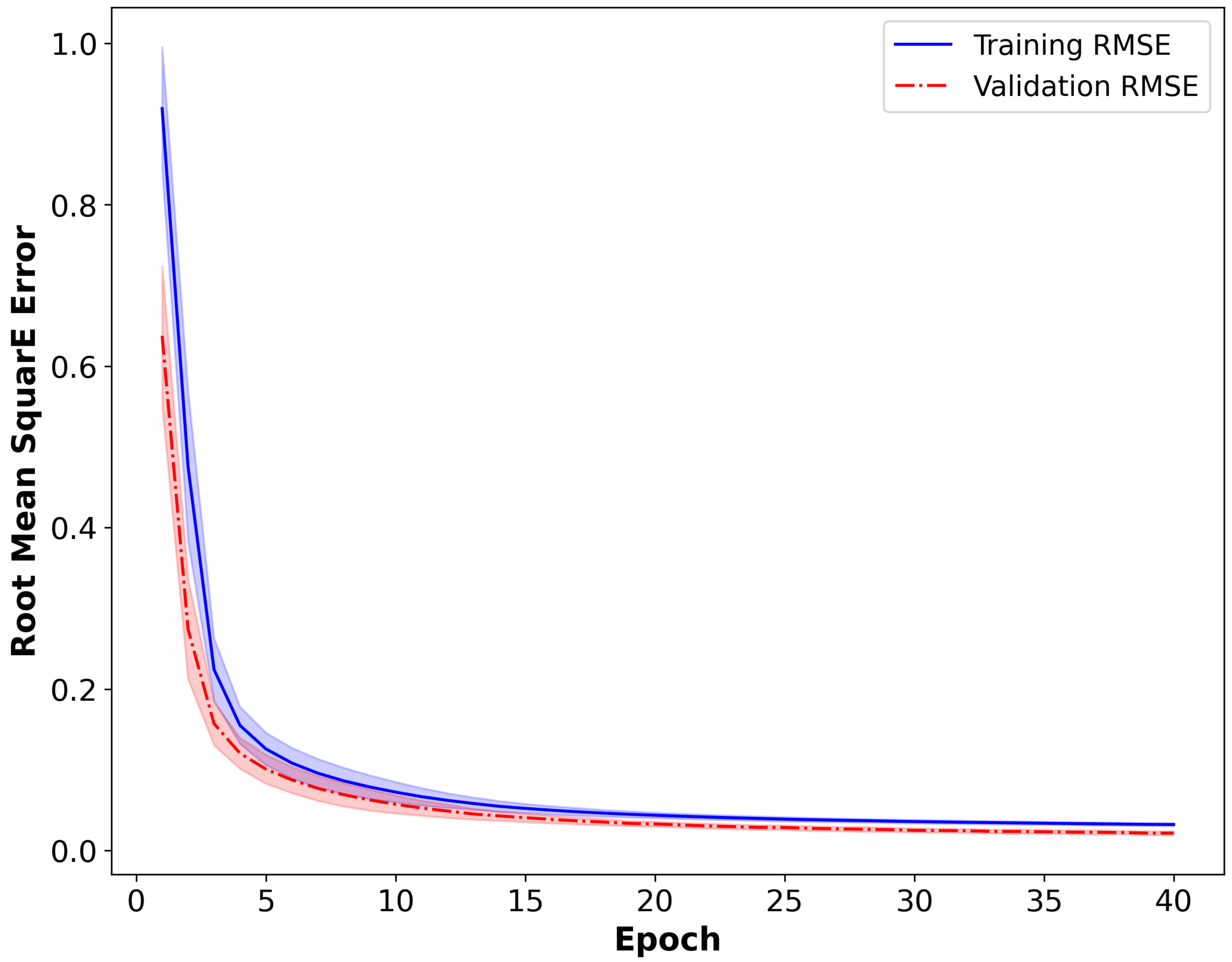
Correlations among the various NMPs for the sampled data set. Negligible correlations are observed for other pairs as the figure plots sample statistics

Training

- The ANN is trained for 40 epochs using the Adam optimiser on a 2-core Intel Xeon CPU.
- Training data batched into batches of size 16.

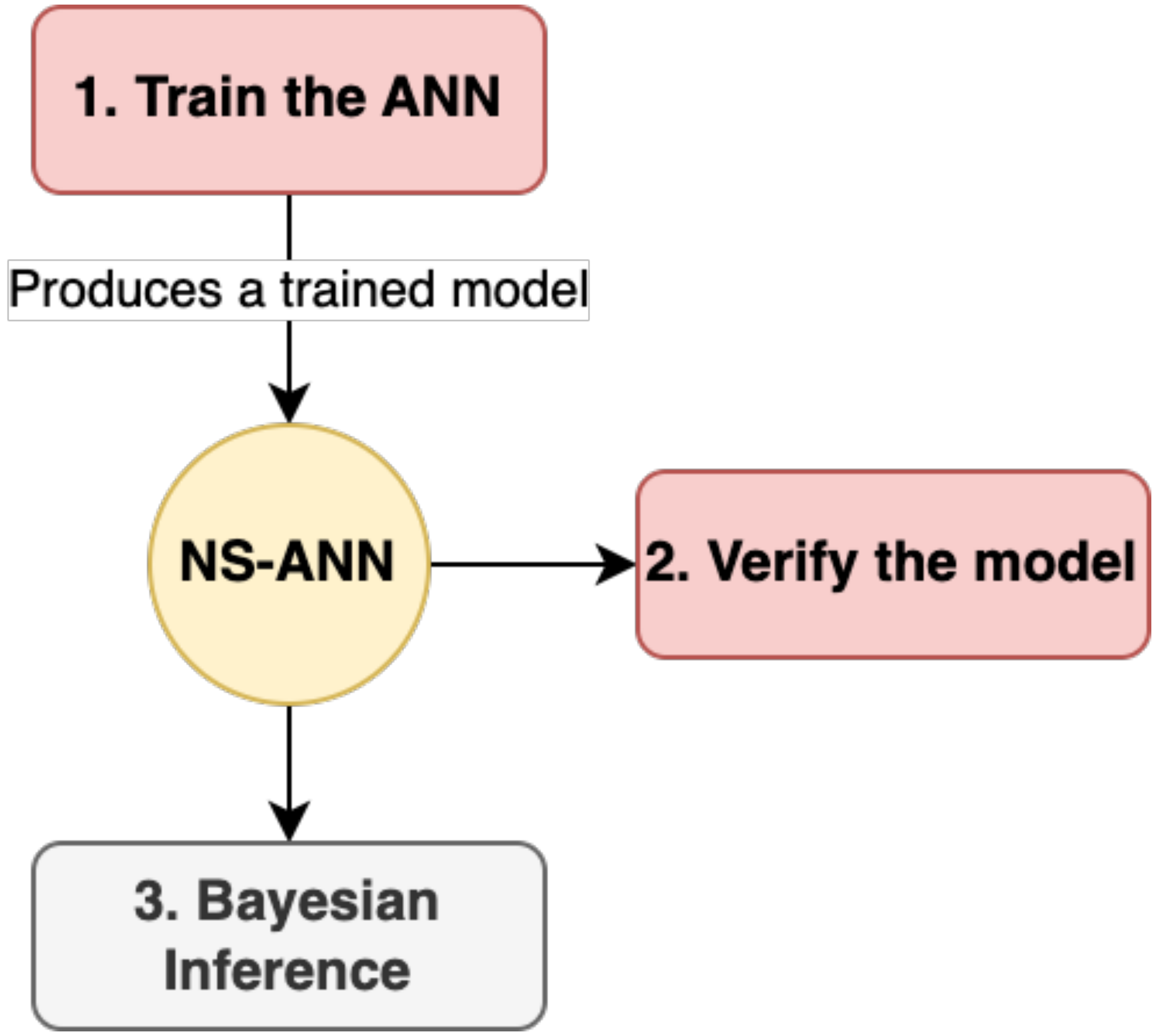
$\hat{y}$	**RMSE**	**(RMSE/$\bar{y}$) $\times$100**
$M_{\max}$	0.024	1.1
$R_{\max}$	0.088	0.8
$R_{1.4}$	0.194	1.5
$\Lambda_{1.0}$	123.800	3.7
$\Lambda_{1.4}$	19.239	4.3
$\Lambda_{1.8}$	5.344	8.2

RMSE on the test set. We also show Root Mean Square relative error.



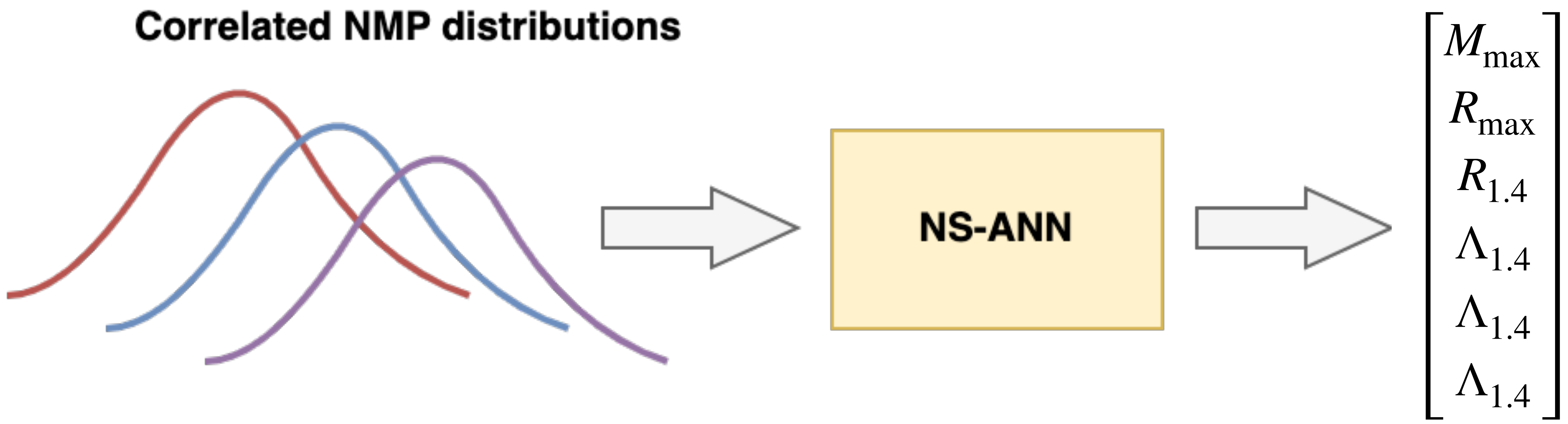
Learning curves for the training (blue, solid) and validation (red, dashed) data sets plotted against the number of elapsed epochs. The shaded region around the solid curve indicates a 1σ or 68% CI region around the central value

Step 2: Verify the Trained Model

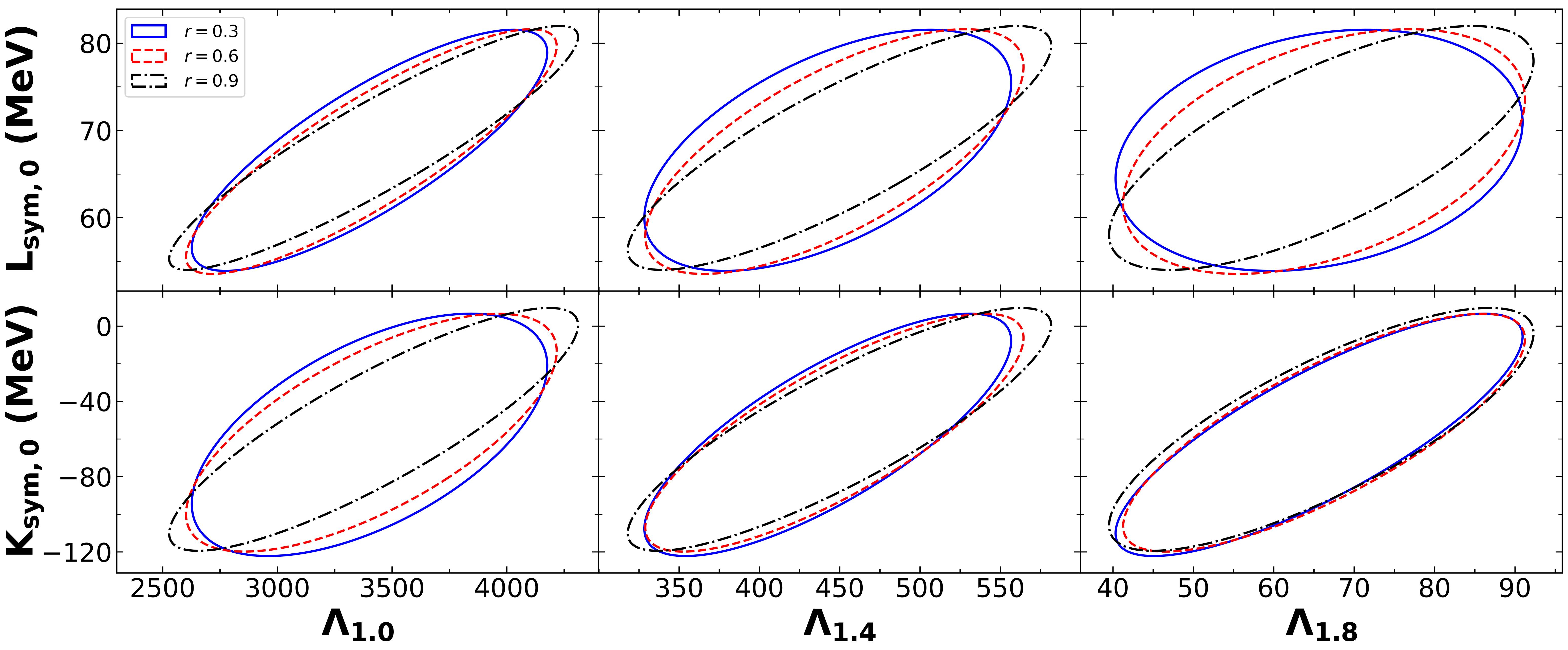


Verifying the NS-ANN

- Given a trained NS-ANN, the first step is to verify the correctness of the model. This is done by tracking the response of the output as the input is varied in a physical manner.
- Once again, pseudo-sets of NMPs are generated, but for three different intercorrelation coefficients, r, between $L_{sym,0}$ and $K_{sym,0}$ set to 0.3, 0.6, and 0.9.
- We plot the 1σ confidence ellipses for Λ_M versus $L_{\text{sym},0}$ and $K_{\text{sym},0}$ for NS mass $M = 1.0, 1.4$, and $1.8 M_\odot$ predicted by the NS-ANN for three different pseudo-sets of NMPs.
- It is known that correlations between tidal deformability and the symmetry energy slope and its curvature are dependent on correlations between these two NMPs

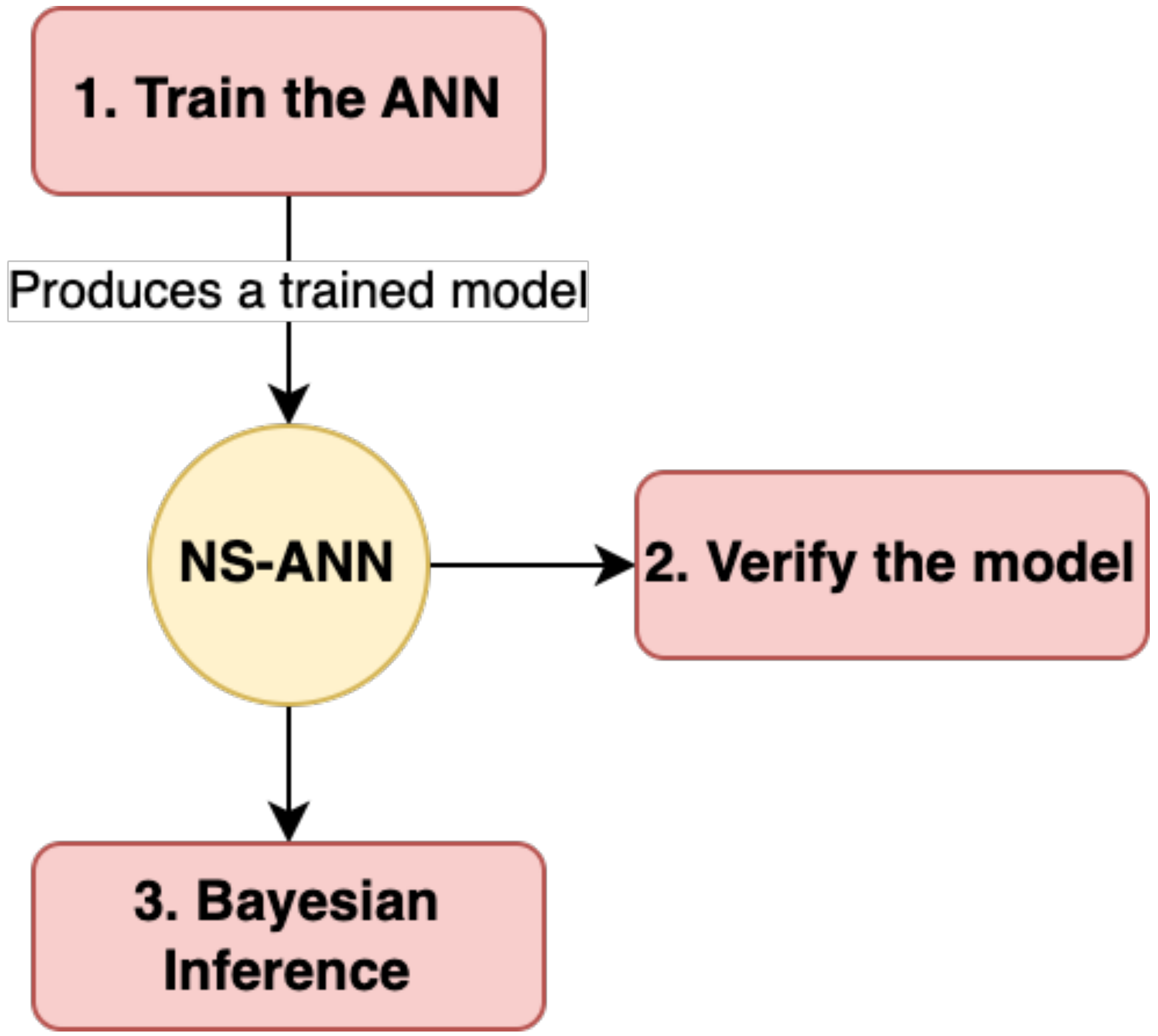


Verifying the NS-ANN



- For the first set, we observe that the correlations of $\Lambda_{1.0,1.4,1.8}$ with $K_{\text{sym},0}$ are $\chi \sim 0.5 - 0.8$, and those with $L_{\text{sym},0}$ are $\chi \sim 0.2 - 0.8$.
- For the second set, we see a non-trivial narrowing of the confidence ellipses, indicating stronger correlations of $\Lambda_{1.0}$ and $\Lambda_{1.4}$ with $L_{\text{sym},0}$ and $K_{\text{sym},0}$, $\chi \sim 0.7 - 0.8$, while these correlations become moderate for $\Lambda_{1.8}$.
- We also observe that the $\Lambda_M - L_{\text{sym},0}$ correlations decrease with increasing mass of NS, while an opposite trend is observed for $\Lambda_M - K_{\text{sym},0}$ correlations.
- These results in excellent agreement with previous studies performed on observing the effect of correlations of the tidal deformability with the slope of the symmetry energy, and its curvature.
- By replacing the models in these studies by the NS-ANN, we are able to obtain almost identical results.

Step 3: Bayesian Inference



Bayesian Inference

- The Bayesian framework allows us to carry out a detailed analysis of the parameters of a model for a given set of fit data.
- The prior knowledge of the model parameters and constraints on them is encoded through the prior distributions. This method uses data points to update the probability of the hypothesis using Bayes' theorem to generate a posterior distribution of the model parameters.
- For a set of parameters $\boldsymbol{\theta}$, the posterior can be obtained as

$$P(\boldsymbol{\theta}\,|\,D) = \frac{\mathscr{L}(D\,|\,\boldsymbol{\theta})P(\boldsymbol{\theta})}{\mathscr{Z}}$$

  Where D denotes the fit data, $P(\boldsymbol{\theta}\,|\,D)$ is the joint posterior probability of the parameters, $\mathscr{L}(D\,|\,\boldsymbol{\theta})$ is the likelihood function, and $\mathscr{Z}$ is the evidence. In our present work, the evidence $\mathscr{Z}$ is not relevant and thus can be ignored.

- The posterior distribution of a given parameter an be obtained by marginalising $P(\boldsymbol{\theta}\,|\,D)$ over the remaining parameters. The marginalised posterior distribution for a parameter is θ_i obtained as

$$P(\theta_i\,|\,D) = \int P(\boldsymbol{\theta}\,|\,D)\prod_{j\neq i} d\theta_j$$

Bayesian Inference

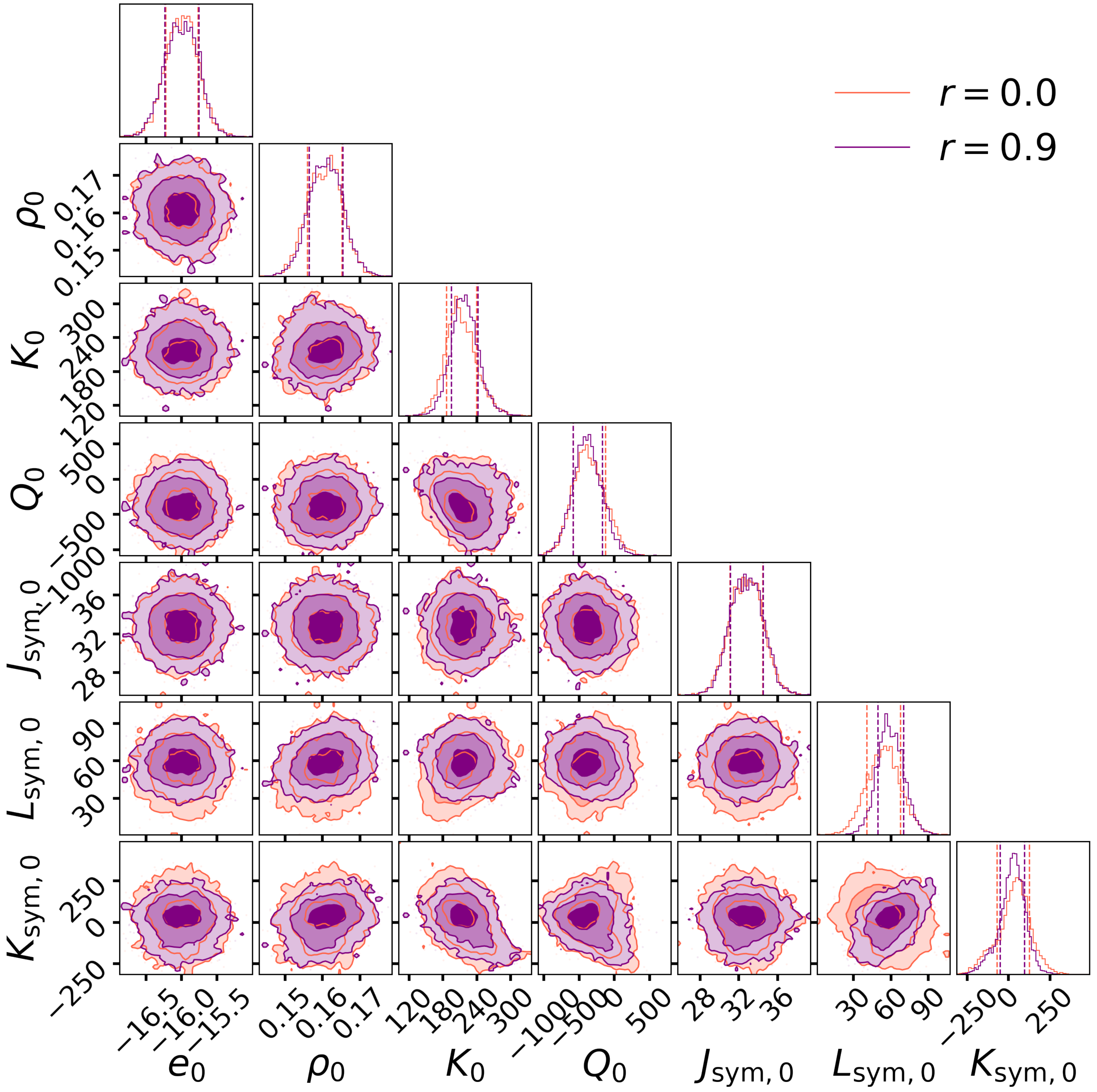
- Given a correct, trained NS-ANN, we finally wish to use the model to perform a statistical analysis of NMPs in light of recent mulitmessenger astronomical data.
- For any Bayesian analysis, we require a prior over the parameters, a likelihood function and a set of fit data.
  - **Prior:** A Gaussian MVGD over the seven NMPs.
  - **Fit Data:** The output of the NS-ANN for a sample drawn from the prior
  - **Likelihood:** A χ^2-type likelihood over four NS observables: maximum mass $M_{\max}$, maximum radius $R_{\max}$, and the radius and tidal deformability for a 1.4$M_\odot$ NS, $(R_{1.4}, \Lambda_{1.4})$. The likelihood has the functional form

$$\mathscr{L}(D|p) = \prod_{j \in \{M_{\max}, R_{\max}, R_{1.4}, \Lambda_{1.4}\}} \left(\frac{1}{\sqrt{2\pi}\sigma_j} \right) \exp\left[-\frac{1}{2} \sum_{i=1}^{N_d} \left(\frac{\mu_j - \text{ANN}(p)_j}{\sigma_j} \right)^2 \right]$$

- These calculations are performed for the millisecond pulsar PSR J0740+6620. The mean and standard deviations for these observables is sourced from NICER and XMM-Newton data.
- Parameter estimation is performed using the Markov Chain Monte Carlo (MCMC) method interfaced in the Bilby library.

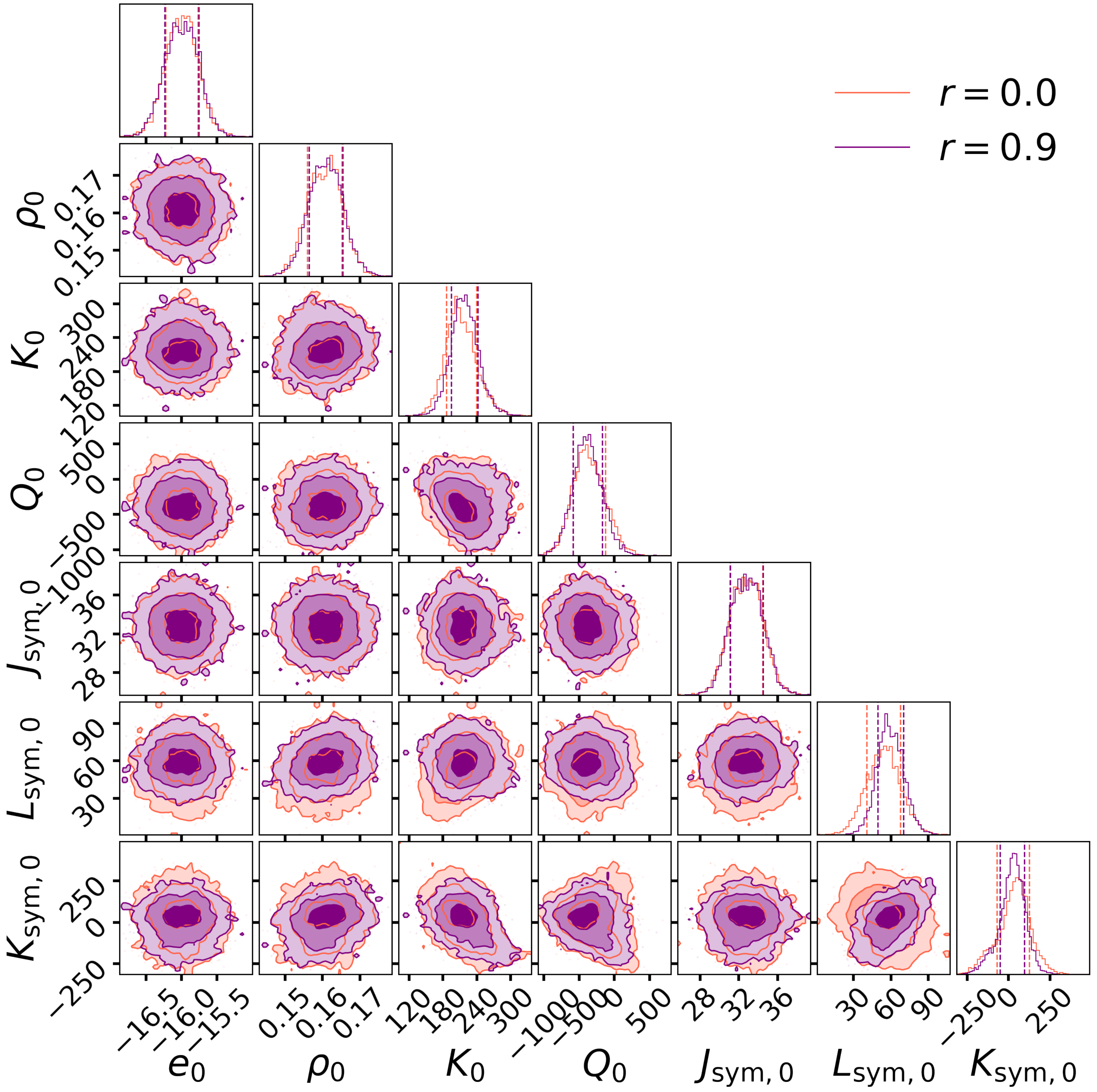
Bayesian Inference

- The aim of this study is to apply a neural network within a Bayesian framework. As a use case, we want to study the how intercorrelations between a pair of NMPs impacts their distribution when constraining them using observational data.
- We perform this analysis for two cases of the prior: (a) an uncorrelated MVGD and (b) a MVGD with the correlation between $L_{sym,0}$ and $K_{sym,0}$ set to 0.9.
- **Correlations implemented in the sampler!**
- We find that if a correlation exists, imposing observational constraints can change the median values of these NMPs by ~9% and -32%, respectively. The corresponding 68% CIs narrow by 23% and 24% respectively.



Bayesian Inference

- Note that the nuclear matter parameters patterning to symmetric nuclear matter are affected by the presence of an interrelation.
- It can be understood that for a given NS matter EoS, there are degeneracies between the symmetric nuclear matter EoS and the symmetry energy.
- Thus, when a correlation is imposed between these two parameters, The uncertainties in higher-order NMPs patterning to symmetry energy propagate to the higher-order NMPs that appear in symmetric nuclear matter EoSs.



Conclusions and Future Work

- As a proof-of-concept, we demonstrate the application of ANNs to predict NS observables like the mass, radius and tidal deformability from a set of Taylor coefficients called nuclear matter parameters obtained from the expansion of the EoS.
- We also show that such a neural network is sensitive to the underlying microphysics information.
- Finally, as a novel contribution, we apply the model within a Bayesian setting. For the use case, we study the effect of intercorrelations among NMPs on the inferred distributions.
- In situation where speed or computational efficiency is desired, a trained neural network can function as a surrogate for traditional physics-based EoS models and possibly provide insights that may be difficult to obtain otherwise.
- For a more realistic application of this framework, empirical uncertainties ought to be considered. This can be done, for example, by using a class of ANNs called Bayesian Neural Networks.
- Moreover, this work is limited to only hadronic NS compositions. In the future, it can be interesting to see whether ML models are able to identify EoSs with different matter compositions and study the effect of these EoSs on higher-order NMPs

Thank you!

Appendix

Parameters	$\boldsymbol{\mu_{p_i}}$	$\sqrt{\boldsymbol{\Sigma_{p_i p_i}}}$
ρ_0	0.16	0.005
e_0	-16	0.26
K_0	230	40
Q_0	-100	200
$J_{\mathrm{sym},0}$	32.5	1.8
$L_{\mathrm{sym},0}$	45	30
$K_{\mathrm{sym},0}$	-100	200

Table 3. The mean value $\boldsymbol{\mu_{p_i}}$ and error $\sqrt{\boldsymbol{\Sigma_{p_i p_i}}}$ for the nuclear matter parameters p_i in the prior multivariate Gaussian distribution. All quantities are in units of MeV except for ρ_0 which is in units of fm^{-3}.

	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}$
M_{max}	2.072	0.066
R_{max}	12.35	0.75
$R_{1.4}$	12.45	0.65
$\Lambda_{1.4}$	255	130

Table 4. Values of the mean, μ, and error σ of the observables used to construct the likelihood function. $\Lambda_{1.4}$ is dimensionless, R_{max} and $R_{1.4}$ are in units of km and M_{max} is in units of $M_\odot$.

Appendix

r		$\Lambda_{1.0}$	$\Lambda_{1.4}$	$\Lambda_{1.8}$
0.3	$L_{\mathrm{sym},0}$	0.81	0.56	0.23
	$K_{\mathrm{sym},0}$	0.57	0.78	0.81
0.6	$L_{\mathrm{sym},0}$	0.86	0.69	0.43
	$K_{\mathrm{sym},0}$	0.69	0.80	0.79
0.9	$L_{\mathrm{sym},0}$	0.91	0.84	0.71
	$K_{\mathrm{sym},0}$	0.86	0.86	0.79

Table 6. The values of the correlation coefficients for $\Lambda_{1.0,1.4,1.8}$ with $L_{\mathrm{sym},0}$ and $K_{\mathrm{sym},0}$ for three different $L_{\mathrm{sym},0} - K_{\mathrm{sym},0}$ correlation coefficients, r

r	NMP	Median	Confidence Interval (CI)					
			50%		68%		90%	
			Min	Max	Min	Max	Min	Max
0	ρ_0	0.161	0.157	0.164	0.156	0.165	0.153	0.168
	e_0	−15.99	−16.15	−15.84	−16.23	−15.77	−16.37	−15.62
	K_0	210.01	193.48	229.44	186.25	239.16	169.97	261.43
	Q_0	−370.18	−513.05	−206.03	−584.49	−125.57	−716.81	29.85
	$J_{\mathrm{sym},0}$	32.80	31.59	33.96	31.09	34.45	30.01	35.53
	$L_{\mathrm{sym},0}$	54.34	44.94	63.27	40.57	67.58	31.88	77.28
	$K_{\mathrm{sym},0}$	40.94	−25.18	98.39	−70.95	124.79	−155.10	189.68
0.9	ρ_0	0.161	0.158	0.164.	0.156	0.165	0.154	0.168
	e_0	−16.00	−16.16	−15.83	−16.24	−15.76	−16.39	−15.62
	K_0	217.62	202.30	233.90	194.77	242.05	179.27	260.65
	Q_0	−385.28	−517.08	−247.22	−581.46	−169.05	−716.75	−33.48
	$J_{\mathrm{sym},0}$	32.82	31.63	33.99	31.13	34.53	30.13	35.54
	$L_{\mathrm{sym},0}$	59.22	52.55	66.58	49.26	69.92	42.55	77.18
	$K_{\mathrm{sym},0}$	27.67	−22.90	74.54	−51.84	95.89	−126.78	142.13

Table 7. The median and the minimum (min) and maximum (max) values of the associated 50%, 68%, and 90% confidence intervals obtained for the marginalized posterior distributions of the NMPs for two multivariate Gaussian priors with parameters as listed in Table 3 and data from Table 4. $r \in \{0, 0.9\}$ is the Pearson correlation coefficient between $L_{\mathrm{sym},0}$ and $K_{\mathrm{sym},0}$ in the prior. All quantities are in units of MeV except for ρ_0 which is in units of fm^{-3}.