Generative models in neutron star physics

Márcio Ferreira

Hadrons and Hadron Interactions in QCD

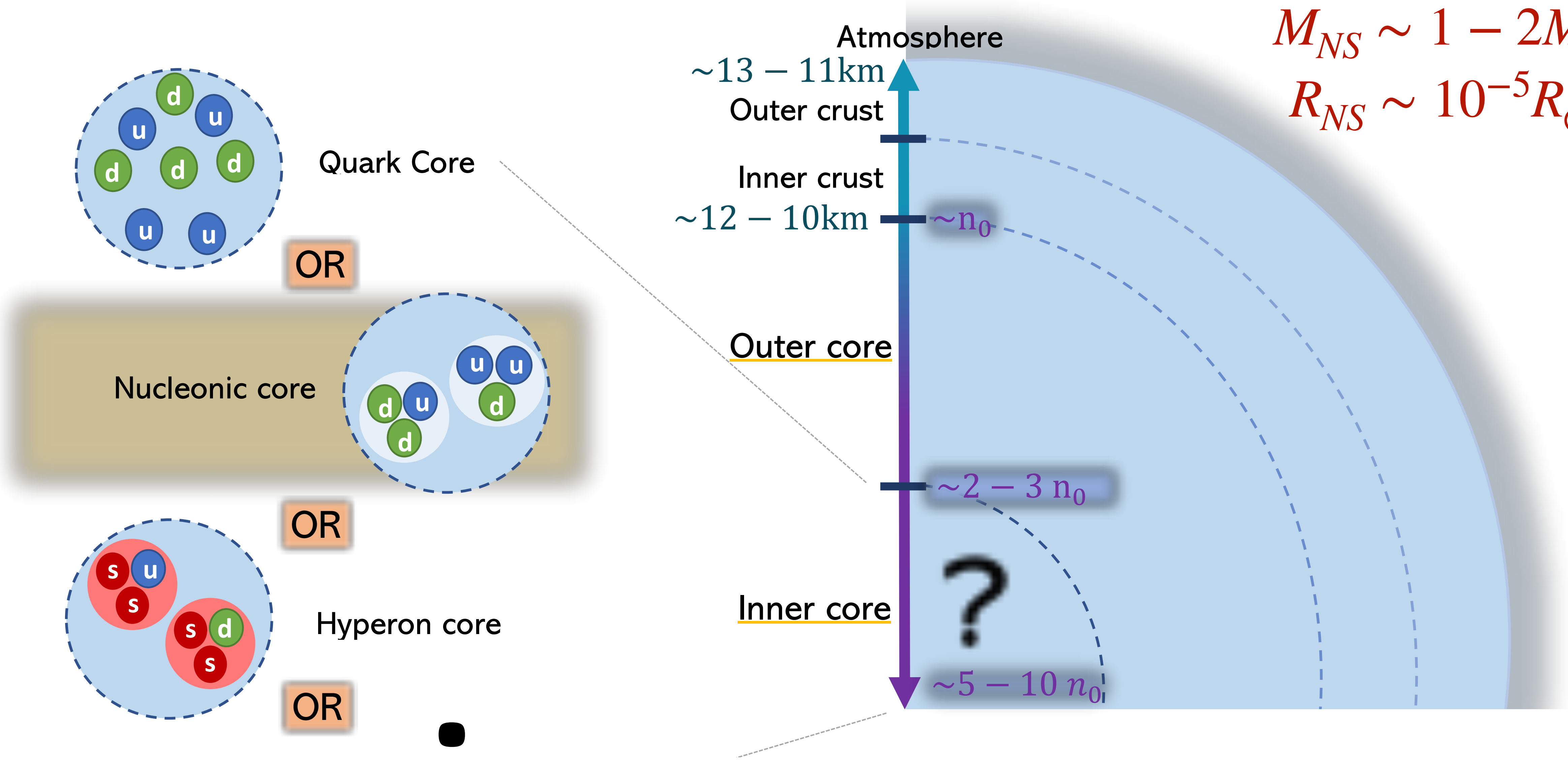
Yukawa Institute for Theoretical Physics

Kyoto University

14Oct - 15Nov (2024)

1 2 9 0 UNIVERSIDADE D COIMBRA

Neutron star physics



$M_{NS} \sim 1 - 2M_{\odot}$

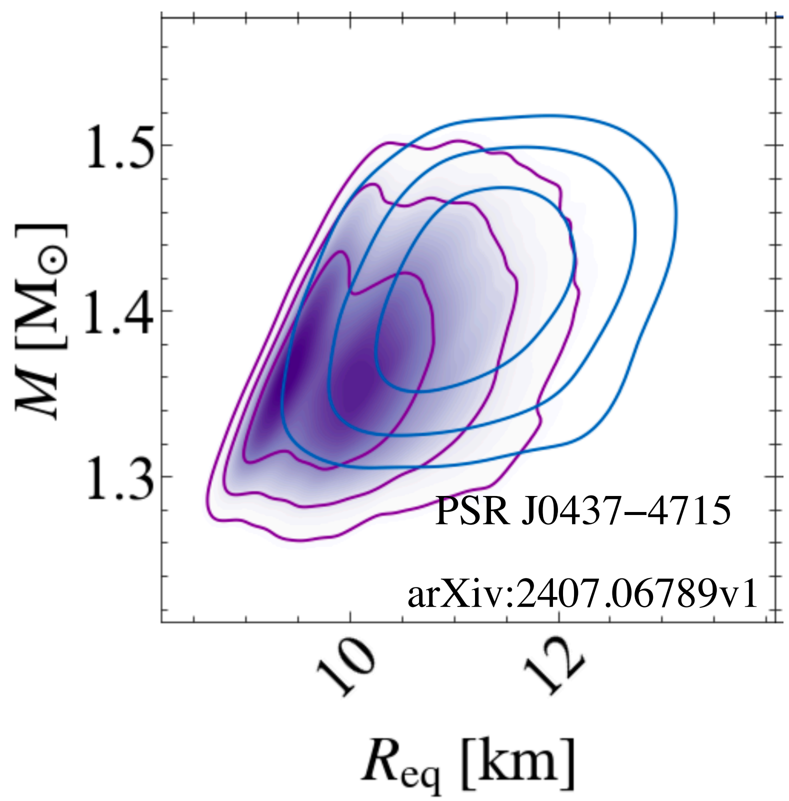
$R_{NS} \sim 10^{-5}R_{\odot}$

The neutron star equation of state remains unknown.

Constraining the EOS

Pulsar Observations

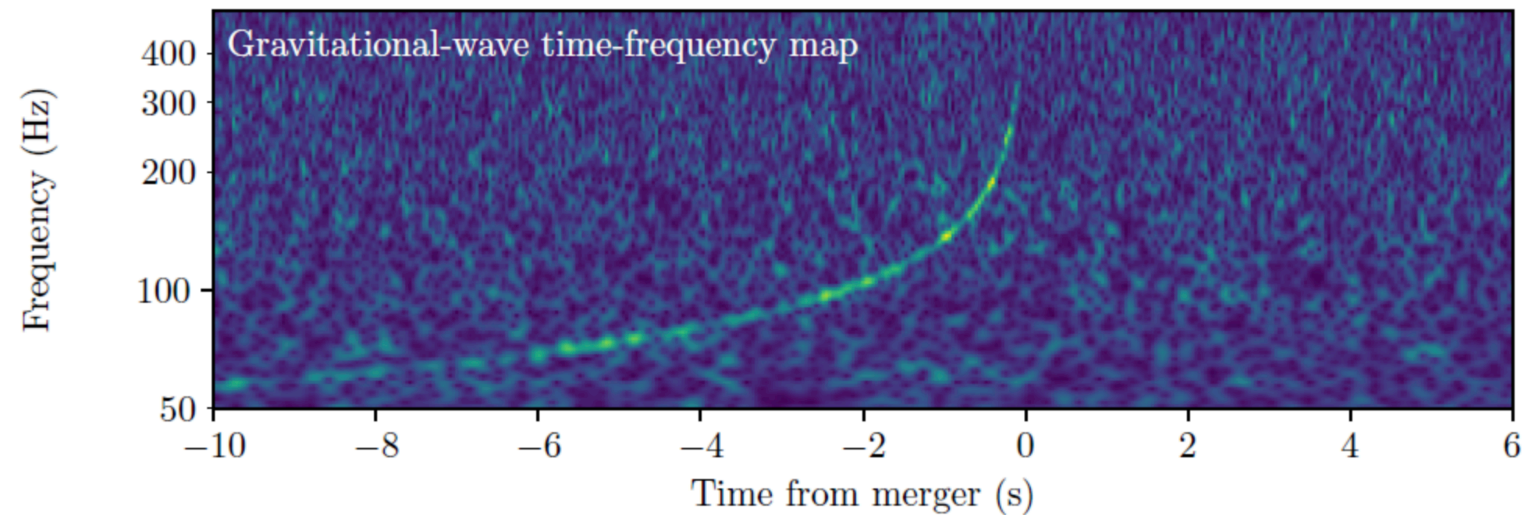
NICER and XMM-Newton
(X-ray observatories)



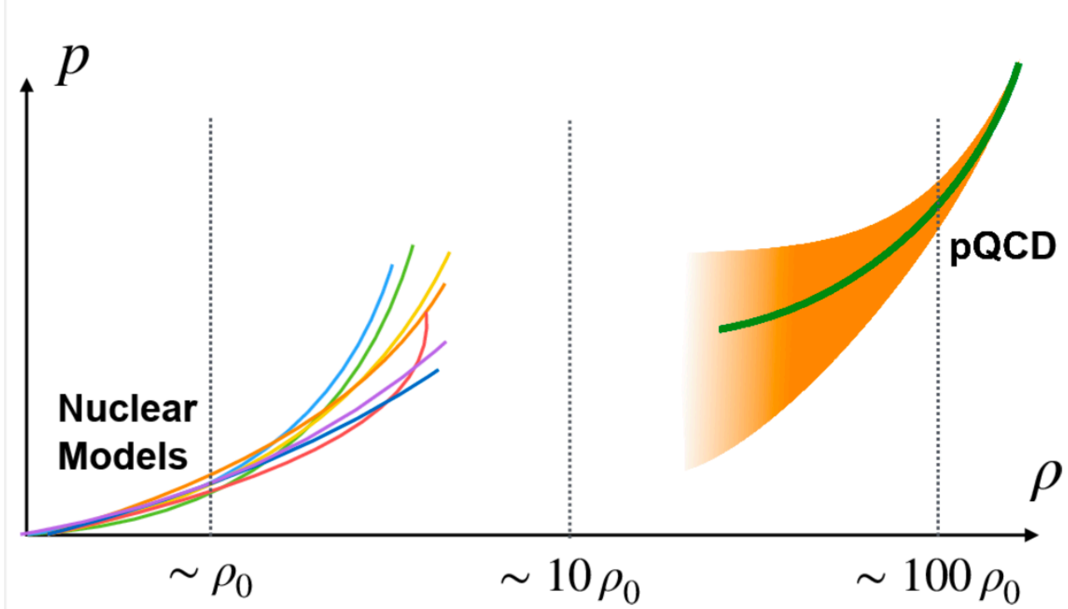
Gravitational Waves

LIGO-Virgo collaboration

GW170817



Strong interactions theory



Combine these data to learn about NS physics

Generative machine learning models in NS physics

- **Conditional variational autoencoders (cVAE):** deep conditional generative model for structured output variables using Gaussian latent variables

  - Fast prediction using stochastic feed-forward inference (efficient and flexible approach to perform Bayesian inference)

$$P(\text{EOS}\,|\,d), \quad d = [(M,R)_1, \ldots, (M,R)_N]$$

MF, Michał Bejger, 2403.14266 [nucl-th]

- **Normalizing flows (NF):** define complex distributions by transforming a simple probability density through a series of invertible mappings.

  - Both sampling and density evaluation can be efficiently performed

$$P(d), \quad d = [(M,R)_1, \ldots, (M,R)_N]$$

Probability of d belonging to a continuous $M(R)$ solution

(anomaly detection framework)

Valéria Carvalho, MF, et. al, PRD109 (2024) 103032

Parameter estimation using simulated-based inference

- Bayesian Inference:

$$p(\theta | d) \propto p(d | \theta)\, p(\theta)$$

posterior — likelihood — prior

- Markov Chain Monte Carlo is used to populate the posterior $\theta \sim p(\theta | d)$
  - Requires a huge number of likelihood evaluations
  - Computationally expensive (namely, high-dimensional θ)
  - Convergence issues

- **Simulated-based inference:** no likelihood evaluation needed but just samples

$$d_i \sim p(d | \theta_i)$$

Learn the posterior $q(\theta | d) \approx p(\theta | d)$ using the cVAE
(Neural posterior estimation)

(θ_i, d_i)

Simulated inference - dataset

$$p(d\,|\,\theta) \propto \exp(-(d^{EOS}(\theta) - d)^2/(2\sigma^2))$$ **Likelihood**

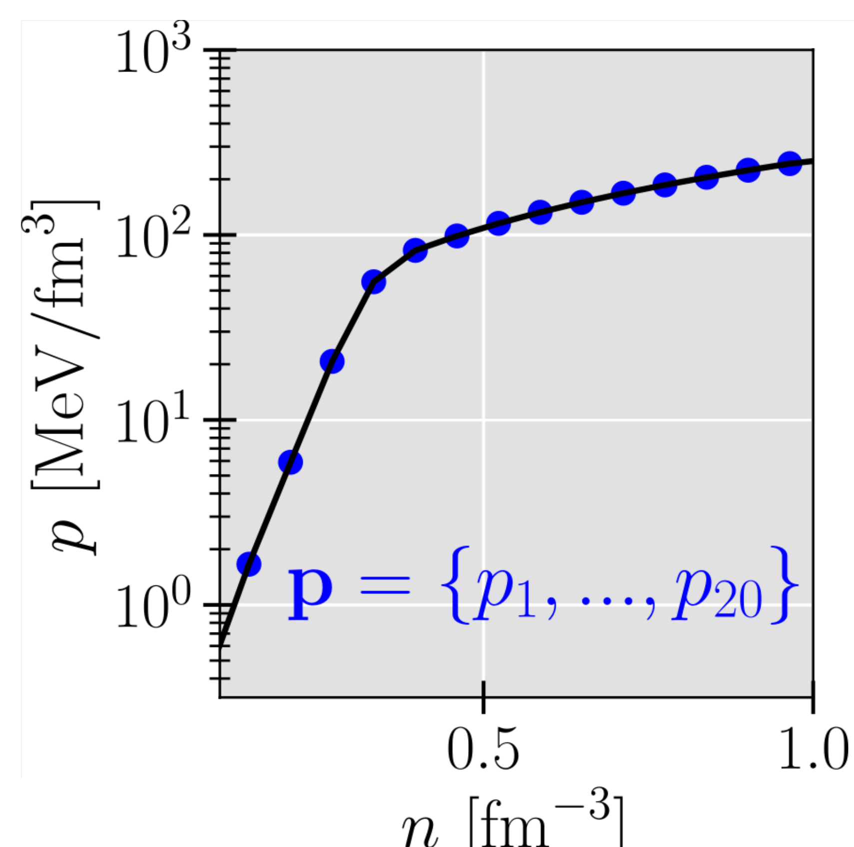
- Dataset construction

Piecewise polytropes: $p(n) = K_i n^{\Gamma_i}$

$$\theta_i \sim p(\theta) \sim \{\Gamma_0, \Gamma_1, n_1, \Gamma_2, n_2, \Gamma_3, n_3, \Gamma_4, n_4\}$$ **Sample from the prior**

- Likelihood samples: uniformly select 5 points over the $TOV(\theta_i)$

$$M_i \sim \mathcal{N}(M_i^{TOV(\theta_i)}, \sigma_M)$$
$$R_i \sim \mathcal{N}(R_i^{TOV(\theta_i)}, \sigma_R)$$
$$d_i = (M_i, R_i)$$



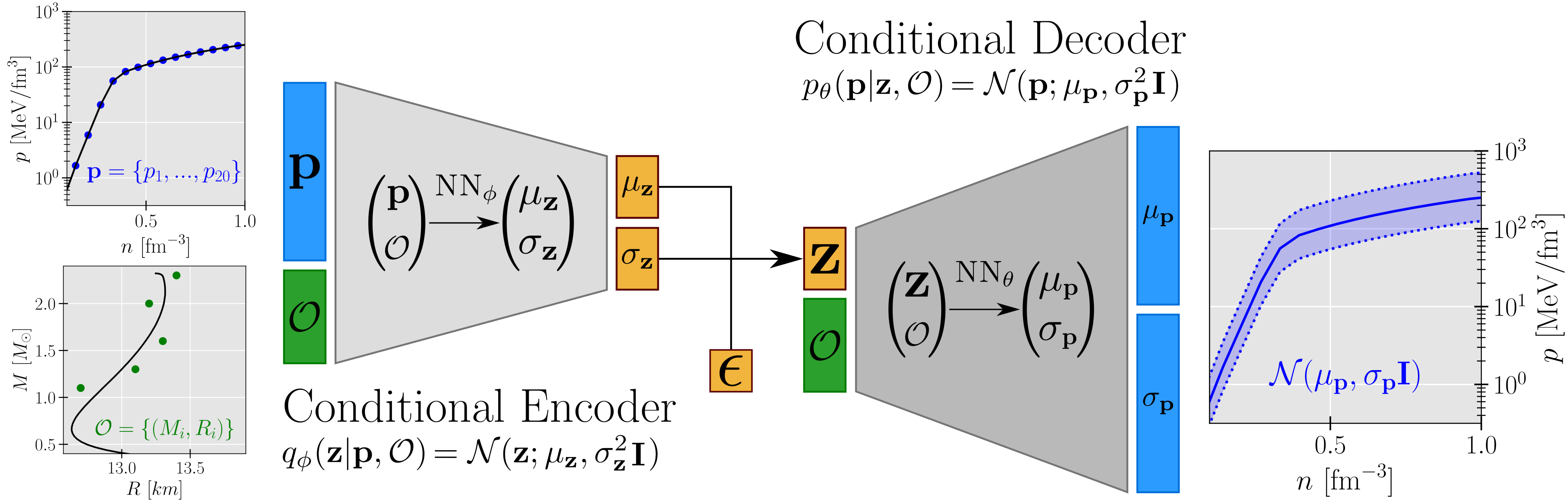
- Training dataset: $\{(\theta_1, d_1), (\theta_2, d_2), \ldots, (\theta_5, d_5)\}$

$$\{(p_1, d_1), (p_2, d_2), \ldots, (p_5, d_5)\}$$

Conditional Variational Autoencoders (cVAE)

- Reconstruct the EOS **given a set of noisy observations**

$$x = [\mathbf{p}, \mathcal{O}] = [p_1, p_2, p_3, \ldots p_{20}, M_1, R_1, M_2, R_2, \ldots, M_5, R_5]$$



$$z = \mu_z + \sigma_z \odot \epsilon, \ \ \epsilon = \mathcal{N}(0,1)$$

Loss function:

$$\mathcal{L}_{\text{VAE}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{x}, \mathcal{O}) = \mathrm{E}_{\boldsymbol{z}\sim q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})}\left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}, \mathcal{O})\right] - \mathrm{D}_{\text{KL}}\left(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x}, \mathcal{O})||p_0(\boldsymbol{z})\right)$$

Reconstruction term

Regularizer

Forces the encoder probability to be close to $p_0(z) = \mathcal{N}(0,1)$

Kullback-Leibler divergence $D_{\text{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} p(x)\, \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x$

Training - cVAE

- During training a single sample from the latent space is used for each EOS

$$\mathbb{E}_{\boldsymbol{z}\sim q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{p})}\left[\log p_{\boldsymbol{\theta}}(\boldsymbol{p}|\boldsymbol{z},\mathcal{O})\right] \approx \log p_{\boldsymbol{\theta}}(\boldsymbol{p}|\boldsymbol{z}_1,\mathcal{O})$$

$$\mathrm{D_{KL}}\left(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{p},\mathcal{O})||p(\boldsymbol{z})\right) = q_{\boldsymbol{\phi}}(\boldsymbol{z}_1|\boldsymbol{p},\mathcal{O})\log\left(\frac{q_{\boldsymbol{\phi}}(\boldsymbol{z}_1|\boldsymbol{p},\mathcal{O})}{p(\boldsymbol{z}_1)}\right)$$

The encoder/decoder PDFs are Gaussian:

$$p(\boldsymbol{z}_1) = \mathcal{N}(\boldsymbol{z}_1;\mathbf{0},\boldsymbol{I})$$

$$q_{\boldsymbol{\phi}}(\boldsymbol{z}_1|\boldsymbol{p},\mathcal{O}) = \mathcal{N}\left(\boldsymbol{z}_1;\boldsymbol{\mu}_{\boldsymbol{z}},\boldsymbol{\sigma}_{\boldsymbol{z}}^2\boldsymbol{I}\right)$$

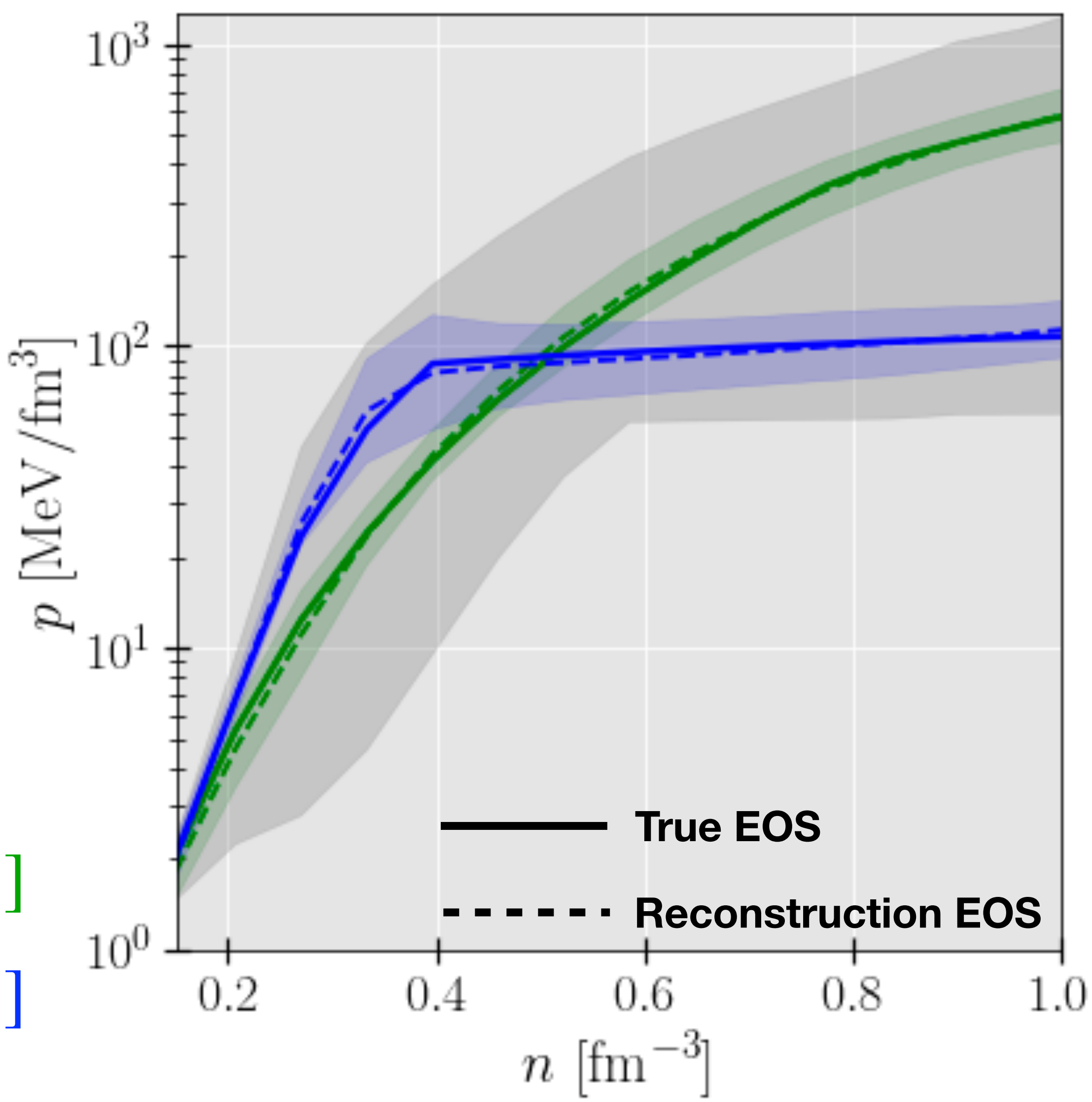
$$p_{\boldsymbol{\theta}}(\boldsymbol{p}|\boldsymbol{z}_1,\mathcal{O}) = \mathcal{N}\left(\boldsymbol{p};\boldsymbol{z}_1,{\boldsymbol{z}_1}^2\boldsymbol{I}\right)$$

The loss function has a closed form solution

Reconstruction of two EOS during training:

$[(p_1, p_2, p_3, \ldots p_{20}), ((M_1, R_1), (M_2, R_2), \ldots, (M_5, R_5))]$

$[(p_1, p_2, p_3, \ldots p_{20}), ((M_1, R_1), (M_2, R_2), \ldots, (M_5, R_5))]$

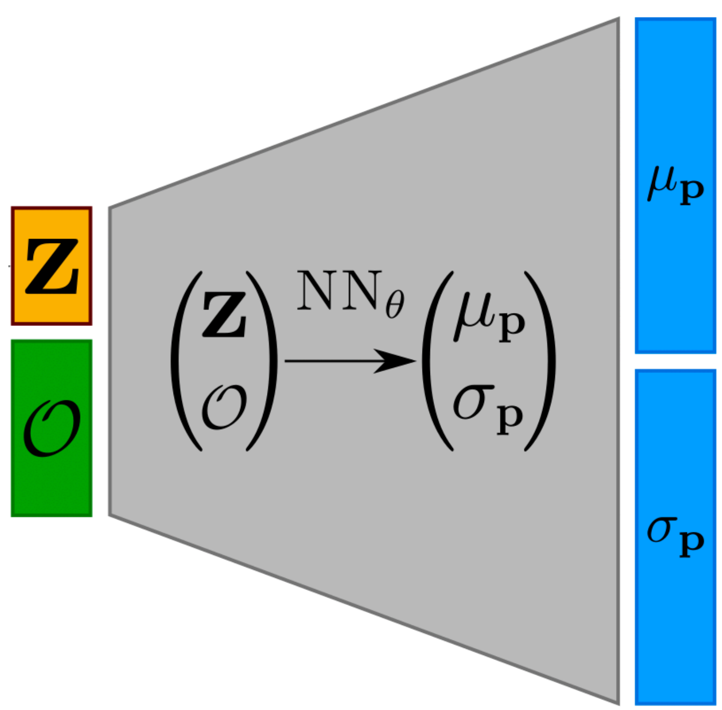


Generative block (cVAE) - Inference model

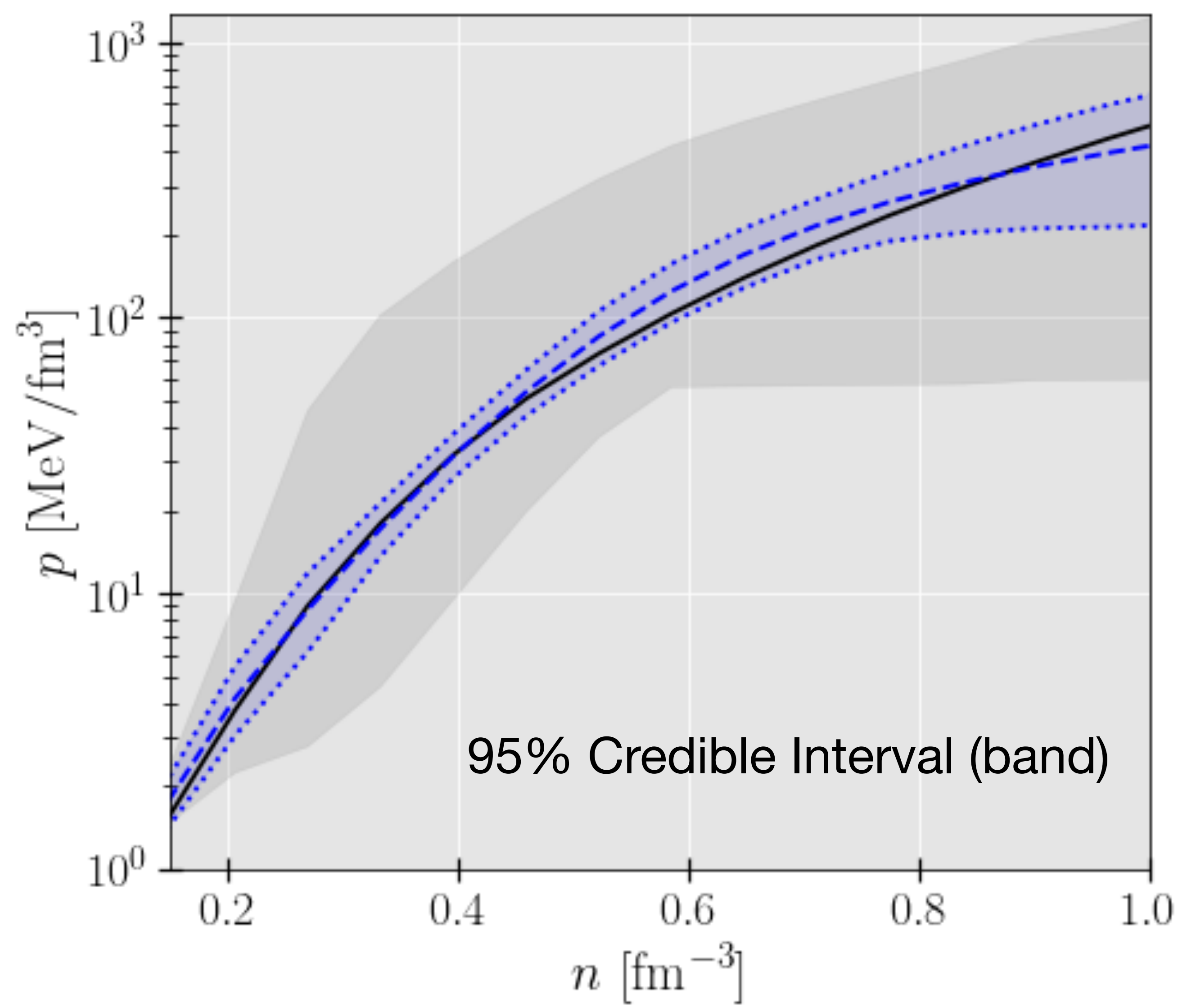
- The generative block as an inference framework: $P(\text{EOS}\,|\,\text{Observation set})$
- During inference phase, we must marginalize over the latent space

$$p_{\boldsymbol{\theta}}(\boldsymbol{p}|\mathcal{O}) \approx \frac{1}{L}\sum_{l=1}^{L}\mathcal{N}_{\boldsymbol{\theta}}(\boldsymbol{p}|\boldsymbol{z}^l, \mathcal{O})$$

Monte Carlo approximation
with L=2000



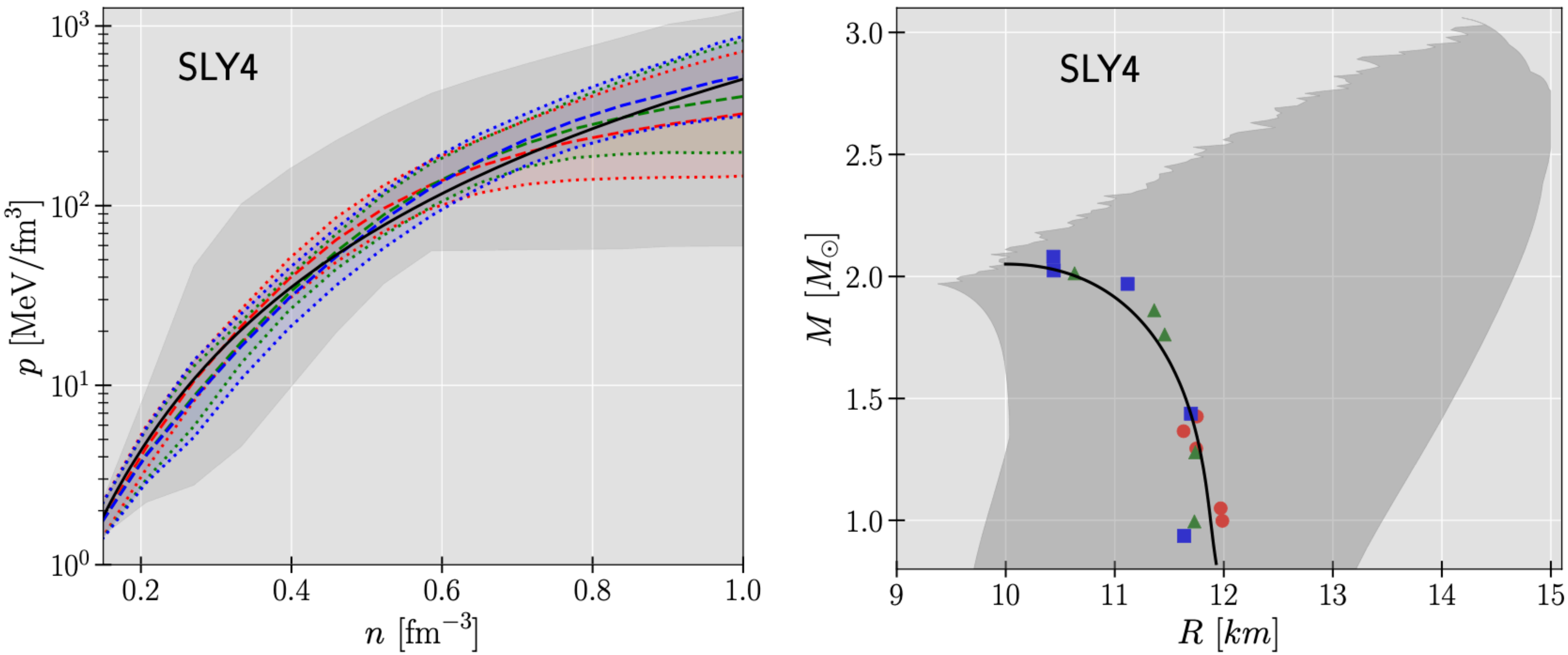
$$\mathcal{O} = [(M_1, R_1), \ldots, (M_5, R_5)]$$



cVAE reconstruction of a nuclear model

- Reconstruction of SLy4 model from a set of noisy observations (5 NS)

$$\mathcal{O} = \{(M_1, R_1), \ldots, (M_5, R_5)\}, \quad \mathcal{O} = \ldots, \quad \mathcal{O} = \ldots.$$



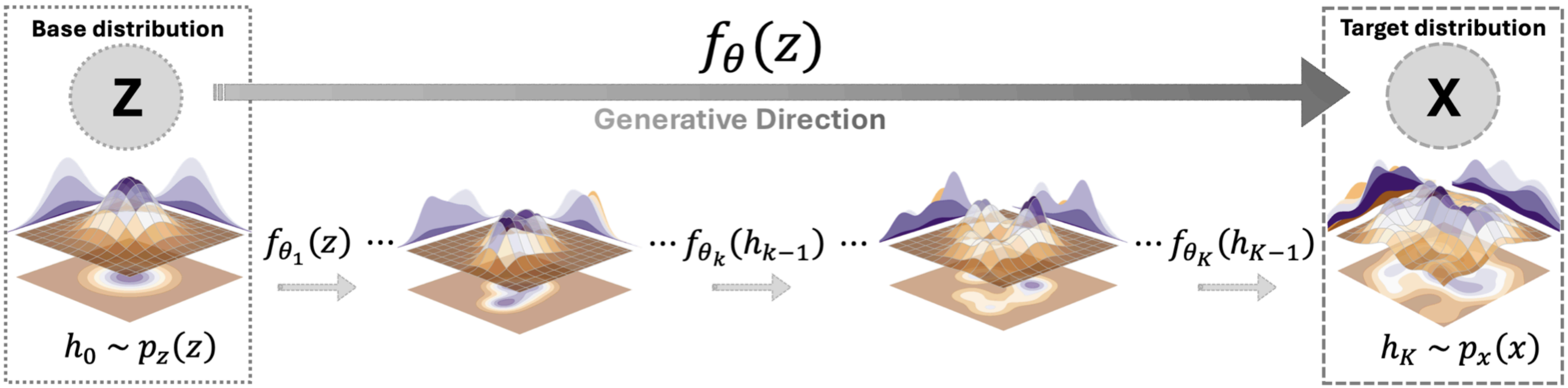
- High density region is restricted by the presence of massive NS
- Low density region more constrained for $\mathcal{O}$ (presence of 5 light NS)
- Good reconstruction (within uncertainty bands)

Conclusions - cVAE

- New generative framework for the equation of state of NS
- Instantaneous inference for any observation set
- The posterior dimension could be easily increased
- The dimension of the conditional vector is flexible (tidal deformability)

Normalizing flows

- Model $p_x(x)$ from multiple **invertible and differentiable transformations** f_{θ_k} of a simple base distribution $p_z(z)$
- Capable of generating new samples $x \sim p_x(x)$ and density estimation $p_x(x)$



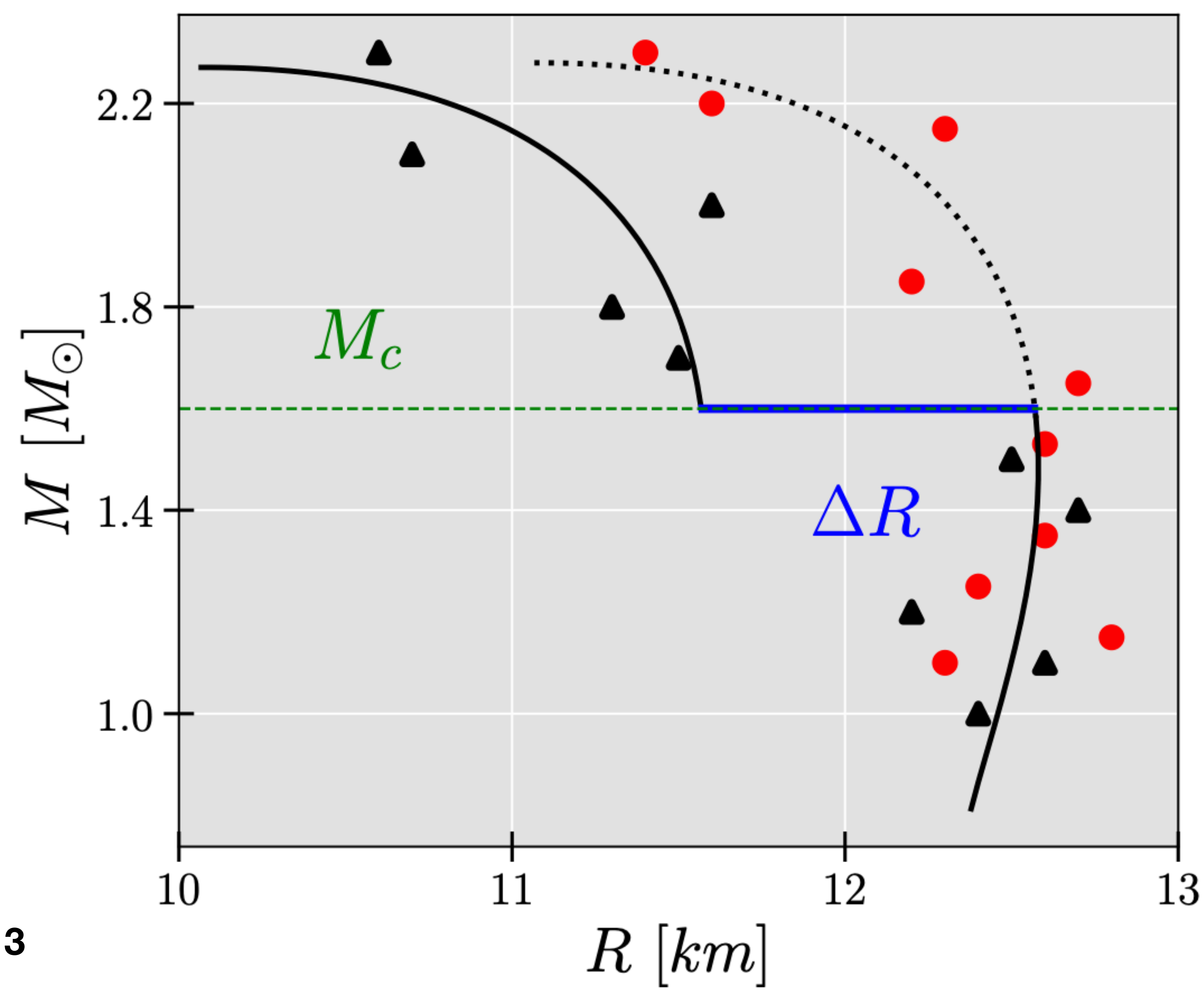
Loss function: minimizing the negative log-likelihood of the data under the NF

$$\mathcal{L}(\mathcal{D}) = -\frac{1}{|\mathcal{D}|}\sum_{\mathbf{x}\epsilon\mathcal{D}} \log\, p_x(\mathbf{x})$$

$$\log\, p_x(\mathbf{x}) = \log\, p_z(\mathbf{f}_{\theta_k}^{-1}(h_{k-1})) + \sum_{k=1}^{K} \log\left|\det \frac{\partial \mathbf{f}_{\theta_k}^{-1}(h_{k-1})}{\partial h_{k-1}}\right|$$

Normalizing flows: detecting a phase transition

- Train on continuous $M(R)$ samples
  - **Training set:** 25287 hadronic models (RMF framework)
- Estimate the likelihood of a non-continuous solutions (two-branch solution)



- Low likelihood for any sample that deviates considerably from the train statistics

$$P(d) \ll P(d)$$

- **Test sets** are parametrized by

$$M_C \quad \Delta R$$

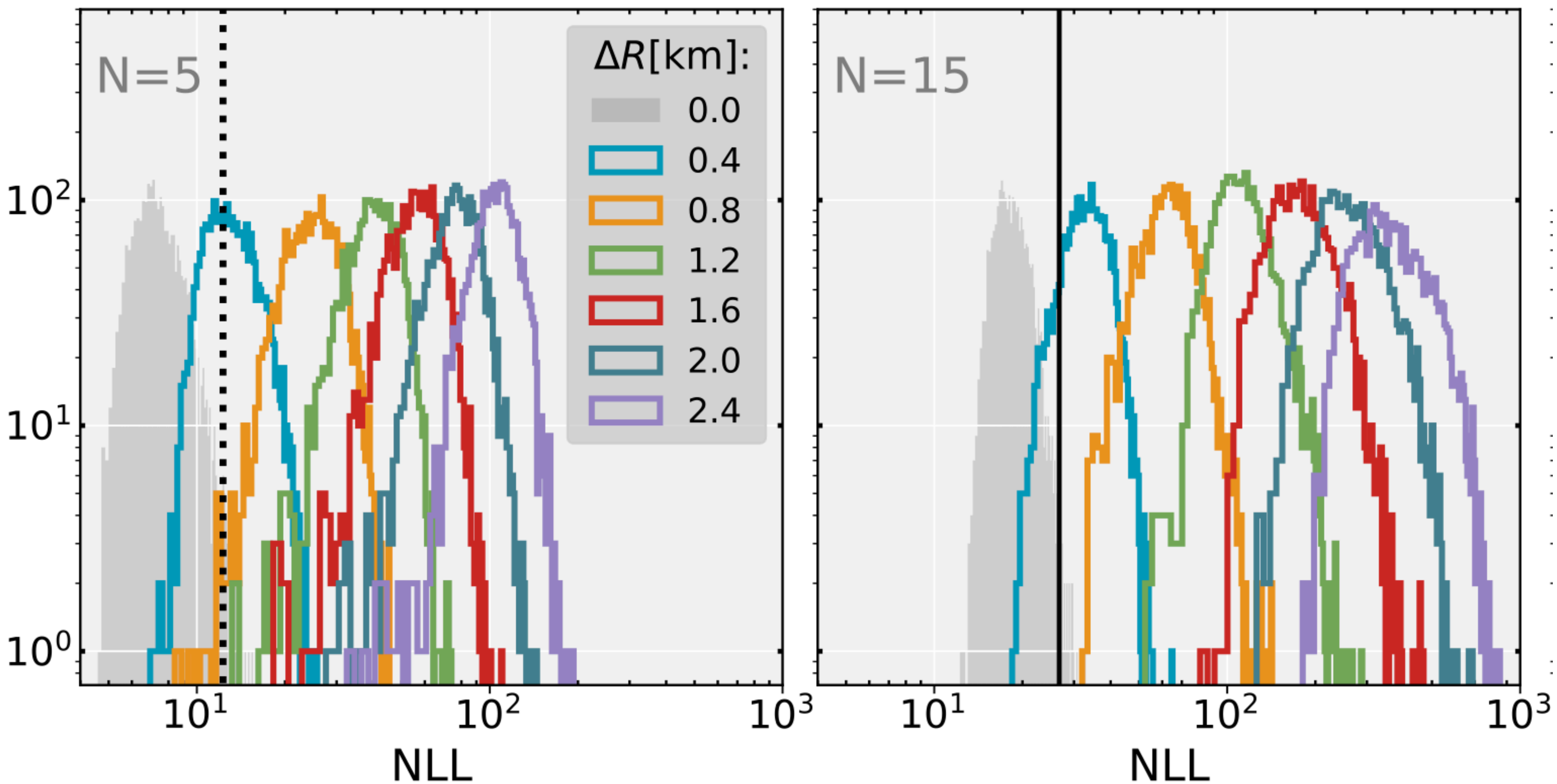
Results: two-branch detection

$M_C = 1.5 M_\odot$

Generation parameters for each dataset.

Dataset	σ_M $[M_\odot]$	σ_R [km]
Set 0	0	0
Set 1	0.1	0.2
Set 2	0.136	0.626

Threshold: False Positive Rate at 1%

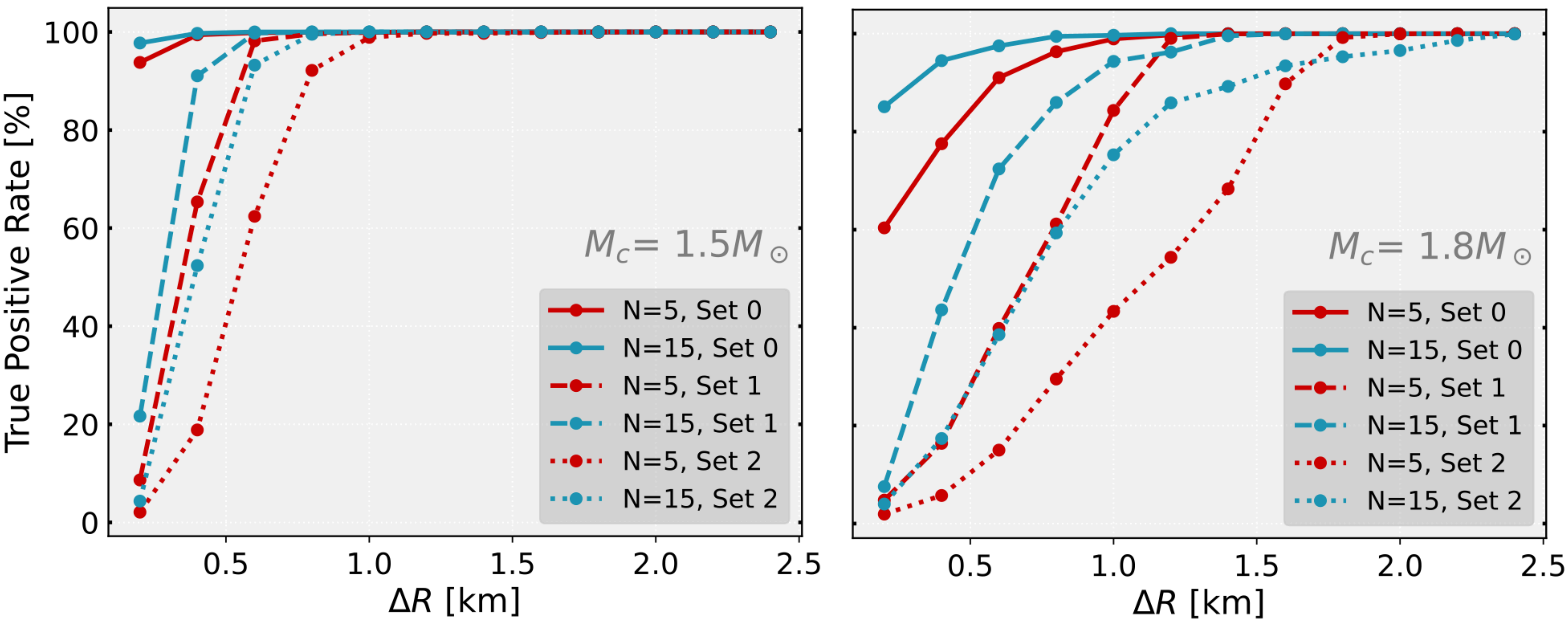


- Detection rate for a two-branch solution is 100% for N=15 stars if $\Delta R \geq 0.8$ km

Results: impact of noise level

Generation parameters for each dataset.

Dataset	σ_M [$M_\odot$]	σ_R [km]
Set 0	0	0
Set 1	0.1	0.2
Set 2	0.136	0.626

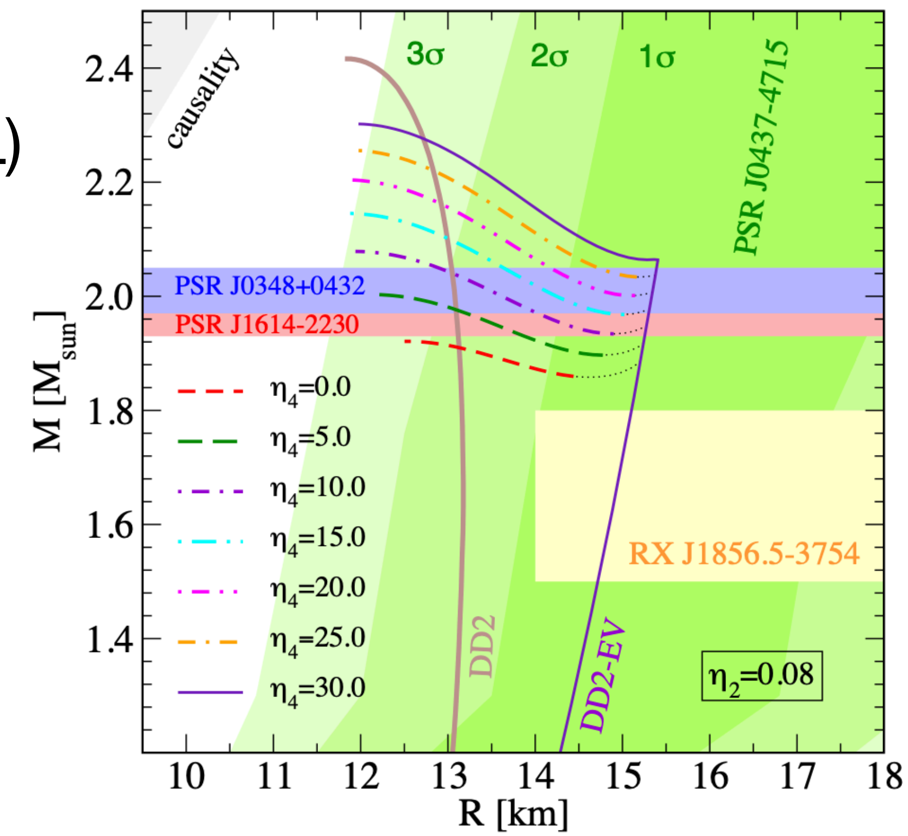


- TPR: lower for $M_c = 1.5M_{\odot}$, increases with N, decreases with noise level
- 100% TRP is reached for 15 neutron stars if

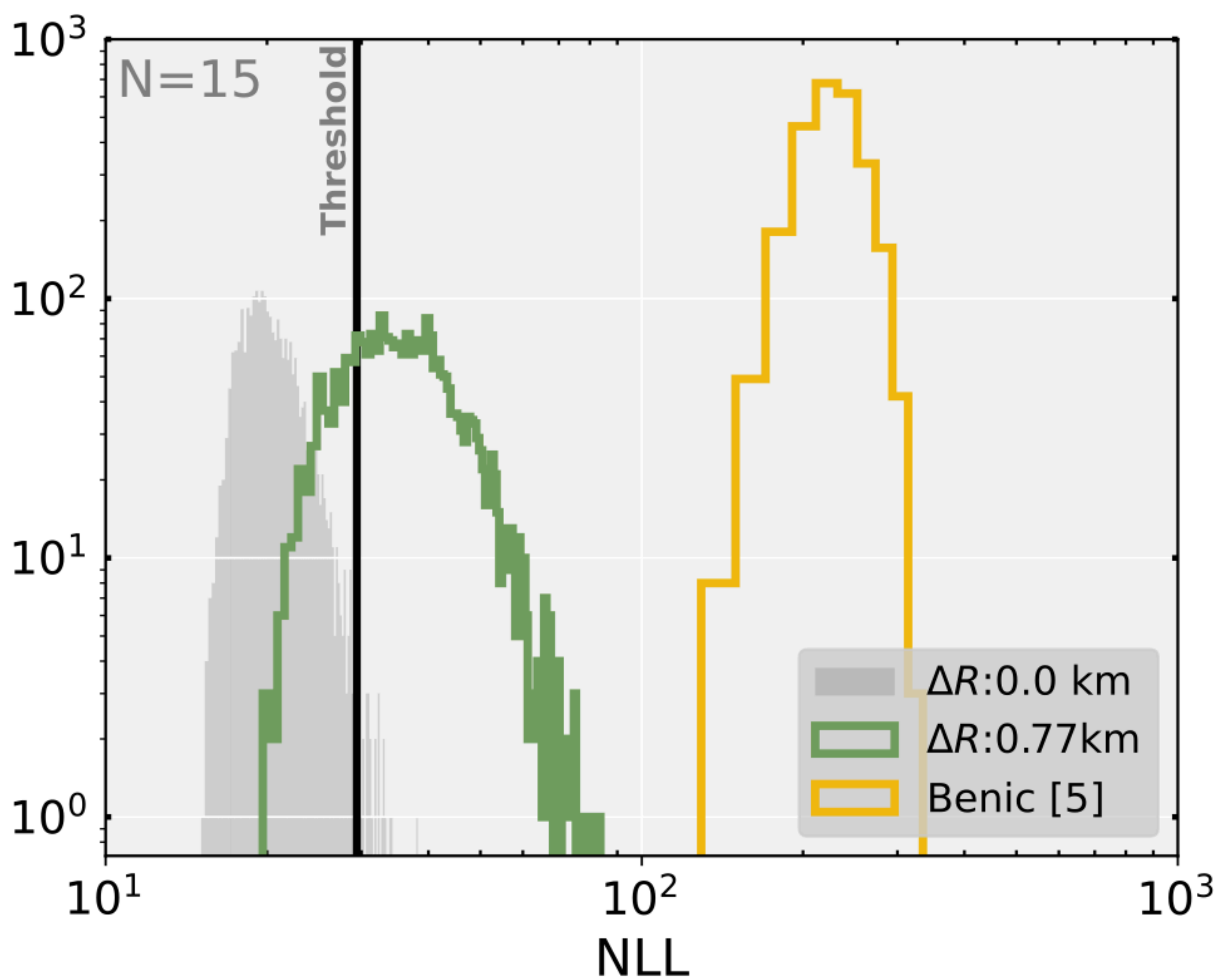
$$\Delta R > 0.6 \text{ km (Set 1) and } \Delta R > 0.8 \text{ km (Set 2)}$$

Test with a microscopic two-branch solution

- Two-branch solution from microscopic models (DD2+NJL) [S. Benic, et. al., Astron. Astrophys. 577, A40 (2015)]



- Test set with $(M_c, \Delta R) = (1.91\, M_\odot, 0.77\,\text{km})$



- Our test sets give a lower bound on the model performance

Conclusions - NF

- NF are flexible models for density estimation (and sampling)
- Differentiate the NS composition from the $M(R)$ curve
- Capable of detecting deviations (anomalies) from a continuous $M(R)$ curve
- For a given observation uncertainty σ_R, we can estimate the ΔR below which the two-branch solution has a low detection probability