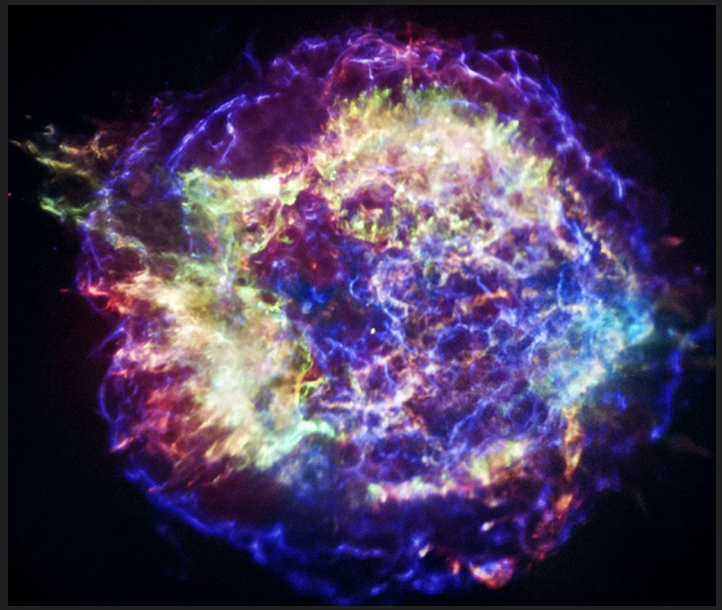
University of Hertfordshire UH

XMM Newton Workshop,
June 7th, 2024

Neutron star population synthesis: an overview and new results

Dr Vanessa Graber
v.graber@herts.ac.uk

in collaboration with Michele Ronchi, Celsa Pardo Araujo, and Nanda Rea



Cassiopeia A supernova remnant
(credit: NASA/CXC/SAO)

Neutron star population synthesis

- We can estimate the **total number of neutron stars in our Galaxy:**

CC supernova rate: ~ 2 per century × **Galaxy age:** ~ 13.6 billion years = **NS number:** ~ 2.8×10^8

- We only **detect** a very **small fraction** of all neutron stars. Population synthesis bridges this gap focusing on the full population of neutron stars (e.g. Faucher-Giguère & Kaspi 2006, Lorimer et al. 2006, Gullón et al. 2014, Cieślar et al. 2020):

Neutron star population synthesis

- We can estimate the **total number of neutron stars in our Galaxy:**

CC supernova rate: ~ 2 per century × **Galaxy age:** ~ 13.6 billion years = **NS number:** ~ 2.8×10^8

- We only **detect** a very **small fraction** of all neutron stars. Population synthesis bridges this gap focusing on the full population of neutron stars (e.g. Faucher-Giguère & Kaspi 2006, Lorimer et al. 2006, Gullón et al. 2014, Cieślar et al. 2020):

model **birth properties** with Monte-Carlo approach → **evolve properties** forward in time

v.graber@herts.ac.uk

Neutron star population synthesis

- We can estimate the **total number of neutron stars in our Galaxy:**

CC supernova rate: ~ 2 per century × **Galaxy age:** ~ 13.6 billion years = **NS number:** ~ 2.8×10^8

- We only **detect** a very **small fraction** of all neutron stars. Population synthesis bridges this gap focusing on the full population of neutron stars (e.g. Faucher-Giguère & Kaspi 2006, Lorimer et al. 2006, Gullón et al. 2014, Cieślar et al. 2020):

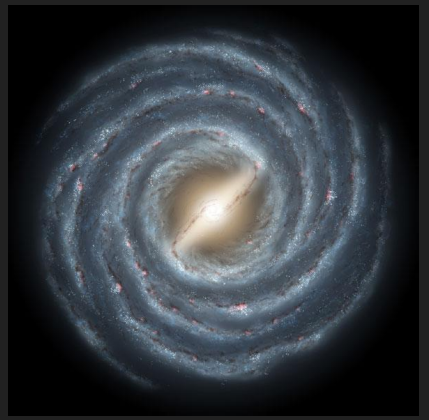
model **birth properties** with Monte-Carlo approach → **evolve properties** forward in time → **apply filters** to mimic observational biases/limits → **compare** mock simulations **to observations** to constrain input

Dynamical evolution I

- **Neutron stars are born in star-forming regions**, i.e., in the Galactic disk along the Milky Way's spiral arms, **and receive kicks** during the supernova explosions.

- We make the following assumptions:
  - Electron-density model (Yao et al., 2017) + rigid rotation with T = 250 Myr.
  - **Exponential disk** with scale height h_c = 0.18 kpc (Wainscoat et al., 1992).
  - Single-component **Maxwell kick-velocity distribution** with dispersion σ_k = 265 km/s (Hobbs et al., 2005).
  - Galactic potential (Marchetti et al., 2019).

Artistic illustration of the Milky Way (credit: NASA JPL)

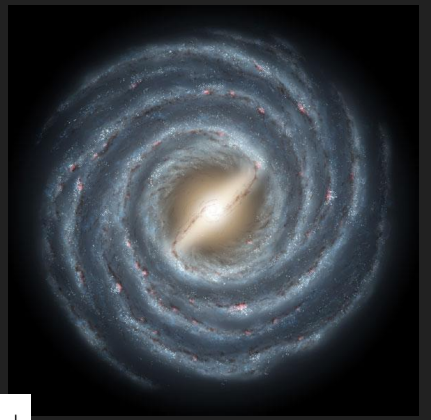


Dynamical evolution I

- **Neutron stars are born in star-forming regions**, i.e., in the Galactic disk along the Milky Way's spiral arms, **and receive kicks** during the supernova explosions.

- We make the following assumptions:
  - Electron-density model (Yao et al., 2017) + rigid rotation with T = 250 Myr.
  - **Exponential disk** with scale height h_c = 0.18 kpc (Wainscoat et al., 1992).
  - Single-component **Maxwell kick-velocity distribution** with dispersion σ_k = 265 km/s (Hobbs et al., 2005).
  - Galactic potential (Marchetti et al., 2019).

Artistic illustration of the Milky Way (credit: NASA JPL)



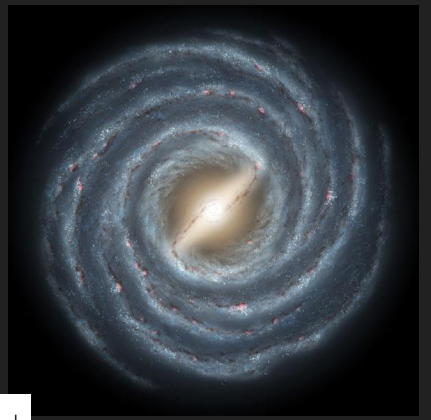
$$\mathcal{P}(z) = \frac{1}{h_c} e^{-\frac{|z|}{h_c}}$$

$$\mathcal{P}(v_k) = \sqrt{\frac{2}{\pi}} \frac{v_k^2}{\sigma_k^3} e^{-\frac{v_k^2}{2\sigma_k^2}}$$

Dynamical evolution I

- **Neutron stars are born in star-forming regions**, i.e., in the Galactic disk along the Milky Way's spiral arms, **and receive kicks** during the supernova explosions.

- We make the following assumptions:
  - Electron-density model (Yao et al., 2017) + rigid rotation with T = 250 Myr.
  - **Exponential disk** with scale height h_c = 0.18 kpc (Wainscoat et al., 1992).
  - Single-component **Maxwell kick-velocity distribution** with dispersion σ_k = 265 km/s (Hobbs et al., 2005).
  - Galactic potential (Marchetti et al., 2019).



Artistic illustration of the Milky Way (credit: NASA JPL)

$$\mathcal{P}(z) = \frac{1}{h_c} e^{-\frac{|z|}{h_c}}$$

$$\mathcal{P}(v_k) = \sqrt{\frac{2}{\pi}} \frac{v_k^2}{\sigma_k^3} e^{-\frac{v_k^2}{2\sigma_k^2}}$$

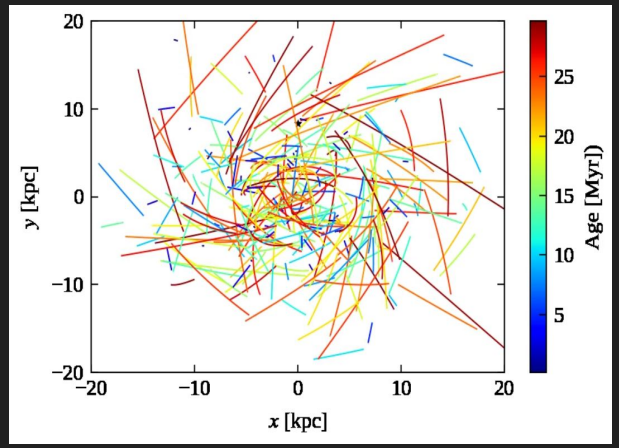
We use this information to determine pulsar positions and velocities.

Dynamical evolution II

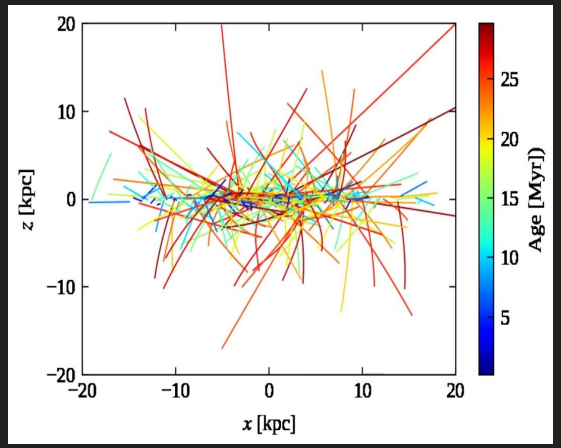
- For our Galactic model Φ_{MW}, we evolve the stars' position & velocity by **solving Newtonian equations of motion** in cylindrical galactocentric coordinates:

$$\ddot{\vec{r}} = -\vec{\nabla}\Phi_{\mathrm{MW}}$$

Top view

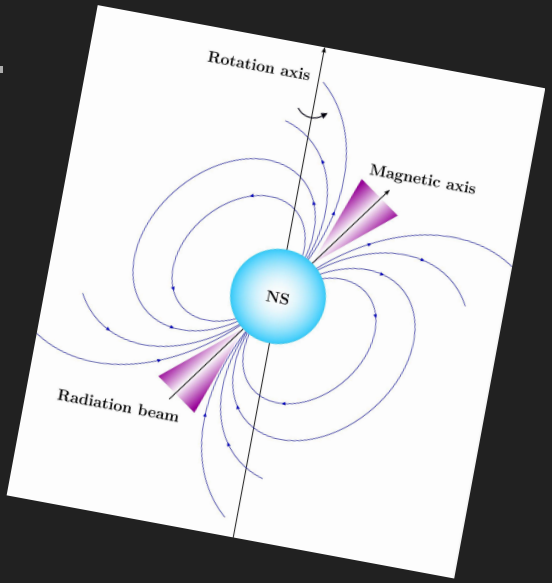


Side view



Galactic evolution tracks for h_c = 0.18 kpc, σ = 265 km/s.

Magneto-rotational evolution I



- The neutron-star magnetosphere exerts a **torque onto the star**. This causes **spin-down** and the **alignment of the magnetic and rotation axes**.

- Neutron star **magnetic fields decay** due to the Hall effect and Ohmic dissipation in the outer stellar layer (crust) (e.g., Viganó et al., 2013 & 2021).

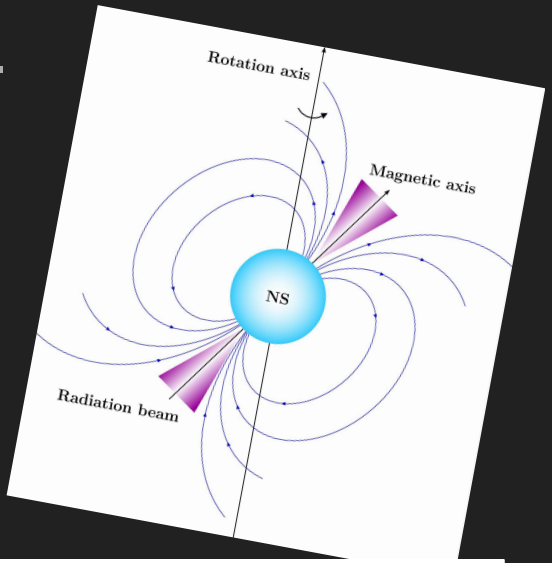
- We make the following assumptions:
  - **Initial periods** follow a log-normal with $\mu_{\log P}$ and $\sigma_{\log P}$ (Igoshev et al., 2022)
  - **Initial fields** follow a log-normal with $\mu_{\log B}$ and $\sigma_{\log B}$ (Gullón et al., 2014)
  - Above $\tau \sim 10^6$ yr, **field decay** follows a power-law with $B(t) \sim B_0 (1 + t/\tau)^a$.

Magneto-rotational evolution I

- The neutron-star magnetosphere exerts a **torque onto the star**. This causes **spin-down** and the **alignment of the magnetic and rotation axes**.

- Neutron star **magnetic fields decay** due to the Hall effect and Ohmic dissipation in the outer stellar layer (crust) (e.g., Viganó et al., 2013 & 2021).

- We make the following assumptions:
  - **Initial periods** follow a log-normal with $\mu_{\log P}$ and $\sigma_{\log P}$ (Igoshev et al., 2022)
  - **Initial fields** follow a log-normal with $\mu_{\log B}$ and $\sigma_{\log B}$ (Gullón et al., 2014)
  - Above $\tau \sim 10^6$ yr, **field decay** follows a power-law with $B(t) \sim B_0 (1 + t/\tau)^a$.



$$\mathcal{P}(\log P_0) = \frac{1}{\sqrt{2\pi}\sigma_{\log P}} \exp\left(-\frac{\log P_0 - \mu_{\log P}}{2\sigma^2_{\log P}}\right)$$

We vary the five free parameters $\mu_{\log P}$, $\mu_{\log B}$, $\sigma_{\log P}$, $\sigma_{\log B}$, and a.

Magneto-rotational evolution II

- To model the magneto-rotational evolution, we numerically **solve two coupled ordinary differential equation**s for the period and the misalignment angle (Aguilera et al., 2008; Philippov et al. 2014).

- We use results from **2D magneto-thermal simulations** to determine the evolution of the magnetic field.

- This allows us to follow the stars' P and $\dot{P}$ evolution in the P$\dot{P}$-plane.

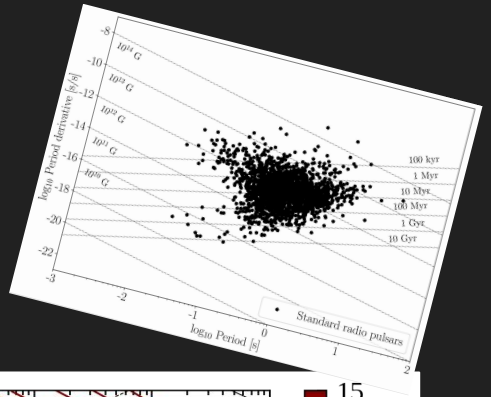
Magneto-rotational evolution II

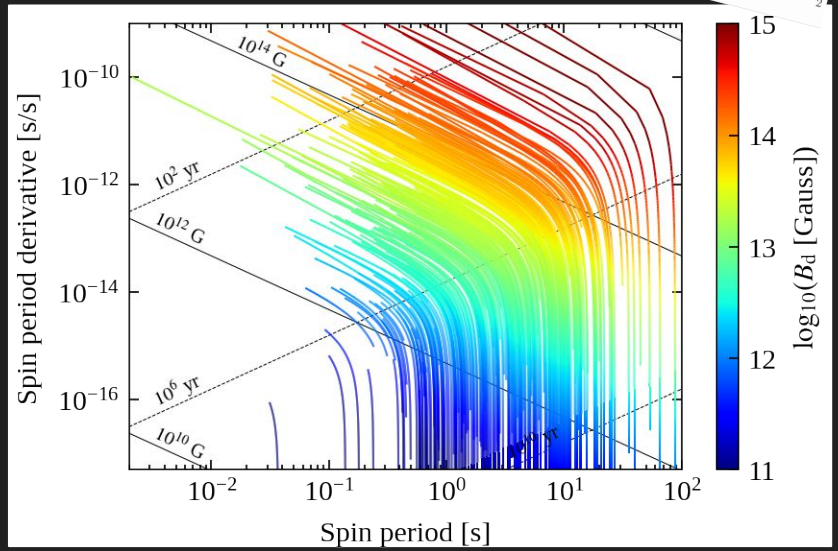
- To model the magneto-rotational evolution, we numerically **solve two coupled ordinary differential equation**s for the period and the misalignment angle (Aguilera et al., 2008; Philippov et al. 2014).

- We use results from **2D magneto-thermal simulations** to determine the evolution of the magnetic field.

- This allows us to follow the stars' P and $\dot{P}$ evolution in the P$\dot{P}$-plane.

Period period-derivative evolution tracks for $\mu_{\log P} = -0.6$, $\sigma_{\log P} = 0.3$, $\mu_{\log B} = 13.25$, $\sigma_{\log B} = 0.75$.





Radio emission and detection

- The stars' **rotational energy E_{rot}** is converted into coherent radio emission. We assume that the corresponding **radio luminosity L_{radio}** is proportional to the loss of E_{rot} (Faucher- Giguère & Kaspi, 2006; Gullón et al., 2014). L_0 is taken from observations.

$$L_{\rm radio} = L_0 \left(\frac{\dot{P}}{P^3} \right)^{1/2} \propto \dot{E}_{\rm dot}^{1/2}$$

- As the radio **emission is beamed**, between ~60-90% of pulsars do not point towards us. For those that do intercept our line of sight, we compute the **radio flux S_{radio}** and the **pulse width W**.

$$S_{\rm radio} = \frac{L_{\rm radio}}{\Omega_{\rm beam} d^2}$$

Radio emission and detection

- The stars' **rotational energy E_{rot}** is converted into coherent radio emission. We assume that the corresponding **radio luminosity L_{radio}** is proportional to the loss of E_{rot} (Faucher- Giguère & Kaspi, 2006; Gullón et al., 2014). L_0 is taken from observations.

$$L_{\mathrm{radio}} = L_0 \left(\frac{\dot{P}}{P^3} \right)^{1/2} \propto \dot{E}_{\mathrm{dot}}^{1/2}$$

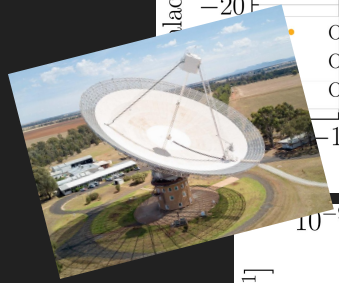
- As the radio **emission is beamed**, between ~60-90% of pulsars do not point towards us. For those that do intercept our line of sight, we compute the **radio flux S_{radio}** and the **pulse width W**.

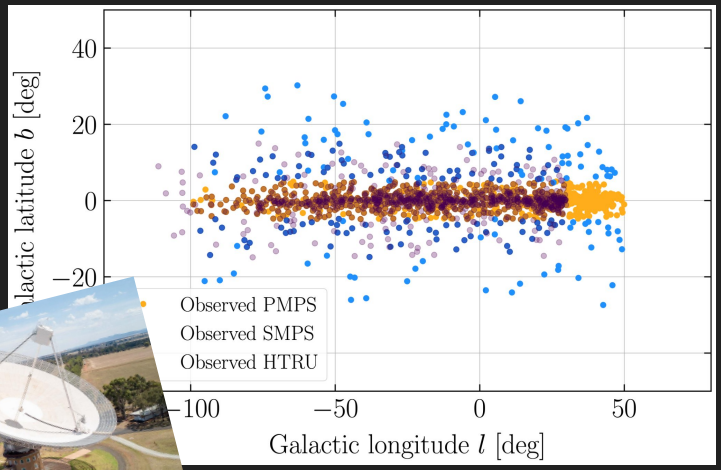
$$S_{\mathrm{radio}} = \frac{L_{\mathrm{radio}}}{\Omega_{\mathrm{beam}} d^2}$$

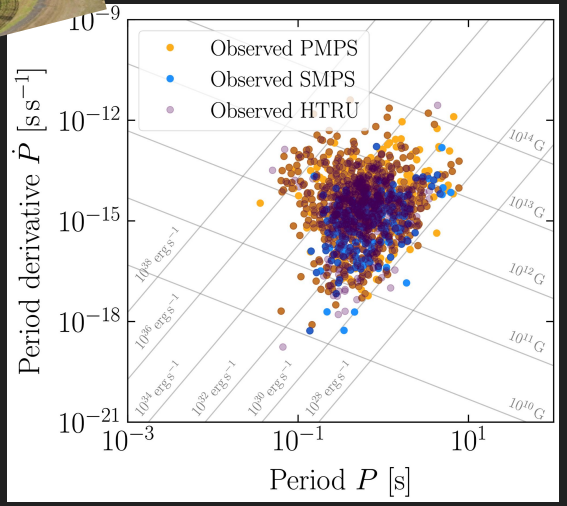
A pulsar counts as detected, if it **exceeds the sensitivity threshold** for a survey recorded with a specific radio telescope.

Three pulsar surveys

- We compare our simulated populations with three surveys from Murriyang (the Parkes Radio Telescope):
  - **Parkes Multibeam Pulsar Survey** (PMPS): *1,009 isolated pulsars*
  - **Swinburne Parkes Multibeam Pulsar Survey** (SMPS): *218 isol. p.*
  - **High Time Resolution Universe Survey** (HTRU): *1,023 isol. pulsars*



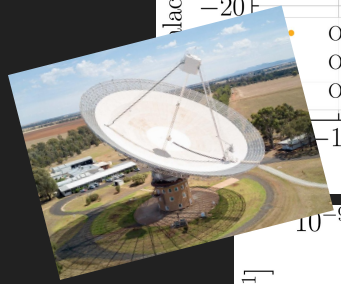


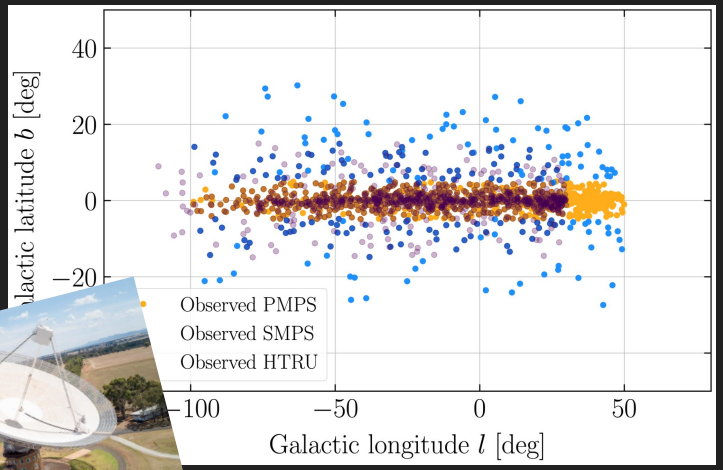


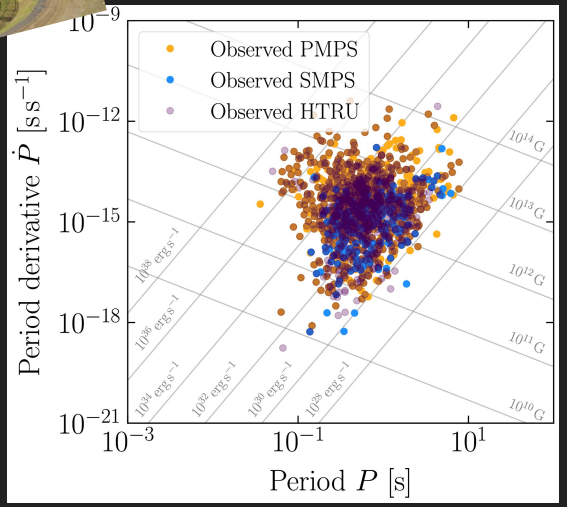
Three pulsar surveys

- We compare our simulated populations with three surveys from Murriyang (the Parkes Radio Telescope):
  - **Parkes Multibeam Pulsar Survey** (PMPS): *1,009 isolated pulsars*
  - **Swinburne Parkes Multibeam Pulsar Survey** (SMPS): *218 isol. p.*
  - **High Time Resolution Universe Survey** (HTRU): *1,023 isol. pulsars*

Can we constrain birth properties by looking at a current snapshot of the pulsar population?







Comparing models and data

- Population synthesis models are complex and have **many free parameters**. Traditional methods (e.g., by-eye studies, χ^2- and KS-tests, annealing methods, MCMC) for parameter inference scale poorly and require simplified models.

- To constrain our free parameters, we want to perform **Bayesian inference**: based on some prior knowledge $\pi(\theta)$, a stochastic model and an observation x, infer the most likely distribution $P(\theta|x)$ for our model parameters θ given:

Comparing models and data

- Population synthesis models are complex and have **many free parameters**. Traditional methods (e.g., by-eye studies, χ^2- and KS-tests, annealing methods, MCMC) for parameter inference scale poorly and require simplified models.

- To constrain our free parameters, we want to perform **Bayesian inference**: based on some prior knowledge $\pi(\theta)$, a stochastic model and an observation x, infer the most likely distribution $P(\theta|x)$ for our model parameters θ given:

$$\underbrace{\mathcal{P}(\theta|x)}_{\text{posterior}} = \frac{\overbrace{\mathcal{P}(\theta)}^{\text{prior } \pi} \overbrace{\mathcal{P}(x|\theta)}^{\text{likelihood } \mathcal{L}}}{\underbrace{\mathcal{P}(x)}_{\text{evidence}}}$$

Comparing models and data

- Population synthesis models are complex and have **many free parameters**. Traditional methods (e.g., by-eye studies, χ^2- and KS-tests, annealing methods, MCMC) for parameter inference scale poorly and require simplified models.

- To constrain our free parameters, we want to perform **Bayesian inference**: based on some prior knowledge $\pi(\theta)$, a stochastic model and an observation x, infer the most likely distribution $P(\theta|x)$ for our model parameters θ given:

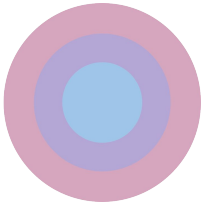
$$\underbrace{\mathcal{P}(\theta|x)}_{\text{posterior}} = \frac{\overbrace{\mathcal{P}(\theta)}^{\text{prior } \pi} \overbrace{\mathcal{P}(x|\theta)}^{\text{likelihood } \mathcal{L}}}{\underbrace{\mathcal{P}(x)}_{\text{evidence}}}$$

For complex simulators, the **likelihood is defined implicitly and often intractable**. This is overcome with **simulation-based** (likelihood-free) **inference** (see e.g. Cranmer et al., 2020).

Simulation-based inference

- To perform **Bayesian inference for any kind of (stochastic) forward model** (e.g. those specified by simulators), we can use the following approach:

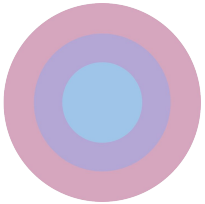
1. sample $\boldsymbol{\theta}_i$
from prior $\pi(\boldsymbol{\theta})$
for $i = 1, \ldots, N$

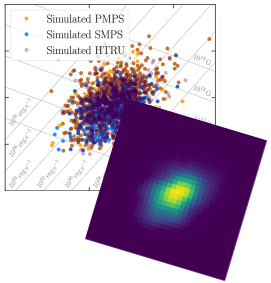


Simulation-based inference

- To perform **Bayesian inference for any kind of (stochastic) forward model** (e.g. those specified by simulators), we can use the following approach:

1. sample $\boldsymbol{\theta}_i$ from prior $\pi(\boldsymbol{\theta})$ for $i = 1, \ldots, N$

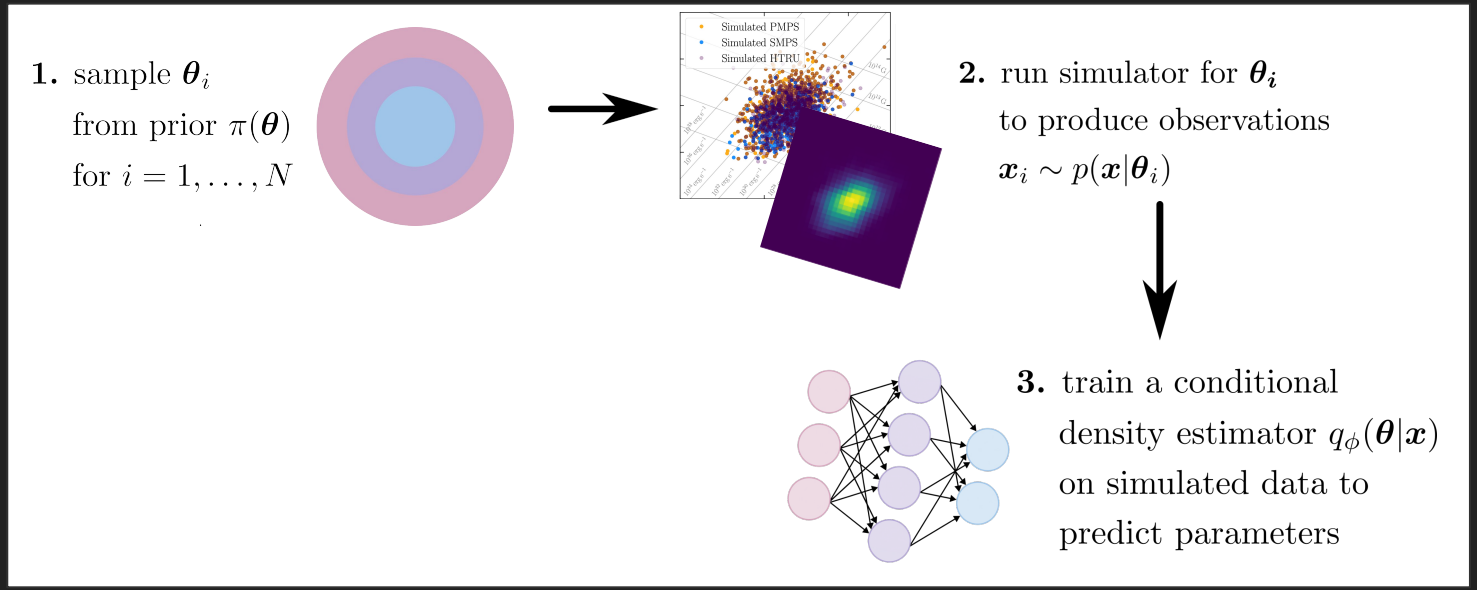




2. run simulator for $\boldsymbol{\theta}_i$ to produce observations $\boldsymbol{x}_i \sim p(\boldsymbol{x}|\boldsymbol{\theta}_i)$

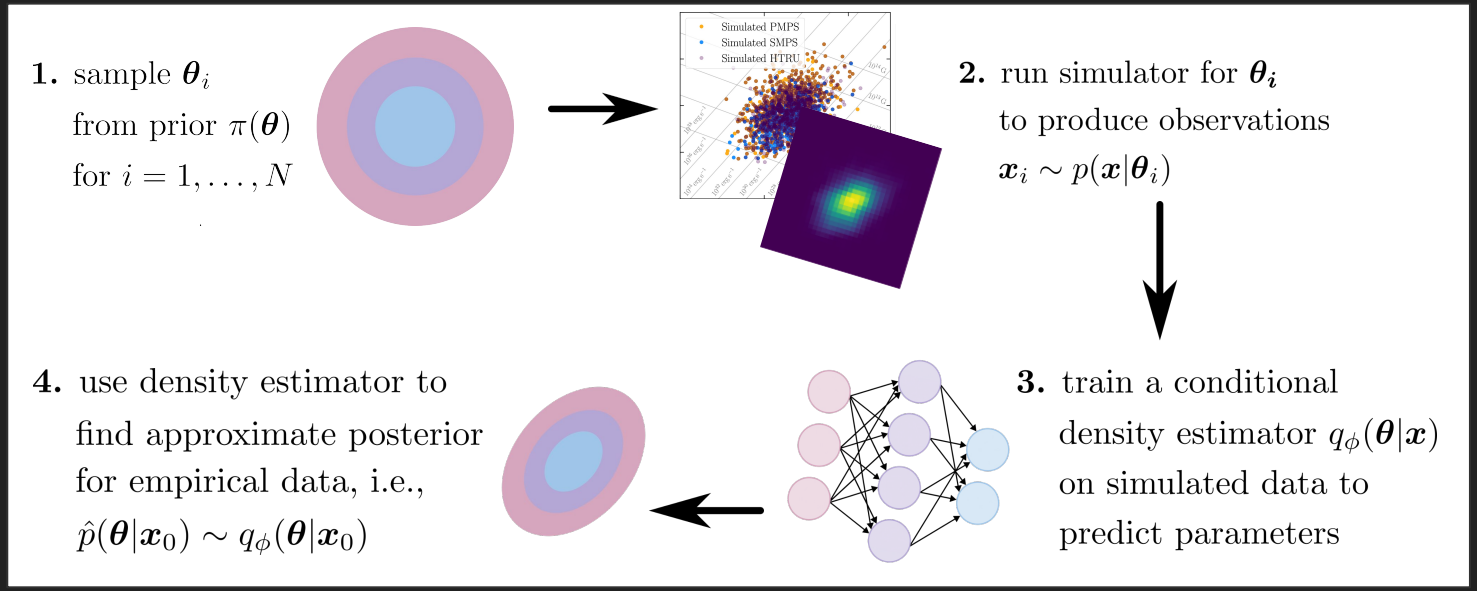
Simulation-based inference

- To perform **Bayesian inference for any kind of (stochastic) forward model** (e.g. those specified by simulators), we can use the following approach:



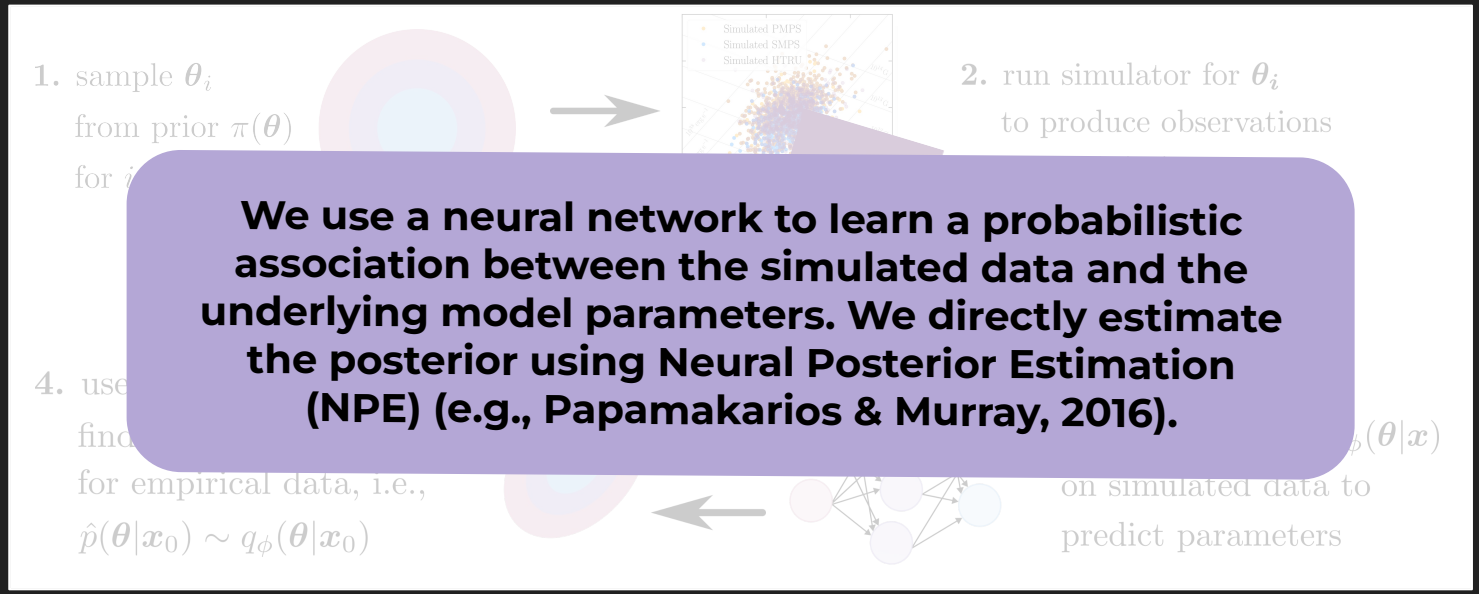
Simulation-based inference

- To perform **Bayesian inference for any kind of (stochastic) forward model** (e.g. those specified by simulators), we can use the following approach:



Simulation-based inference

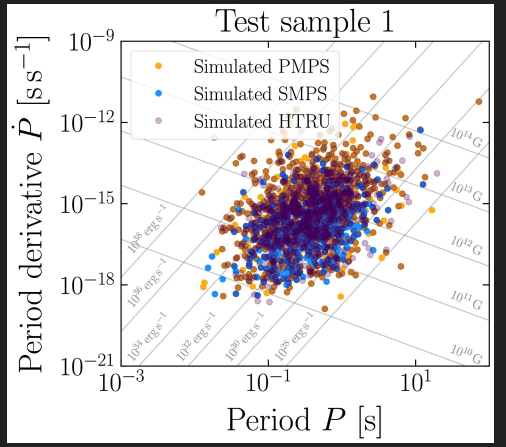
- To perform **Bayesian inference for any kind of (stochastic) forward model** (e.g. those specified by simulators), we can use the following approach:



Simulator output

- We simulate 360,000 populations by varying the 5 magneto-rotational parameters and generate density maps for 3 surveys in P$\dot{P}$-plane (Graber et al., 2024).
  - $\mu_{\log P} \in$ Uniform(-1.5, -0.3)
  - $\sigma_{\log P} \in$ Uniform(0.1, 1.0)
  - $\mu_{\log B} \in$ Uniform(12.0, 14.0)
  - $\sigma_{\log B} \in$ Uniform(0.1, 1.0)
  - $a_{late} \in$ Uniform(-3.0, -0.5)

Test simulation for $\mu_{\log P}$ = -0.85, $\sigma_{\log P}$ = 0.51, $\mu_{\log B}$ = 13.19, $\sigma_{\log B}$ = 0.96 and a_{late} = -0.86.

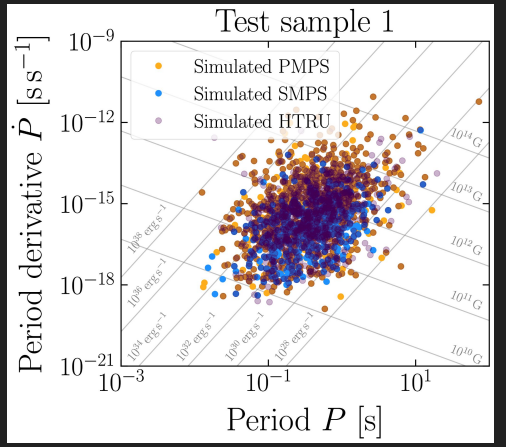


$$\mathcal{P}(\log P_0) = \frac{1}{\sqrt{2\pi}\sigma_{\log P}} \exp\left(-\frac{\log P_0 - \mu_{\log P}}{2\sigma^2_{\log P}}\right)$$

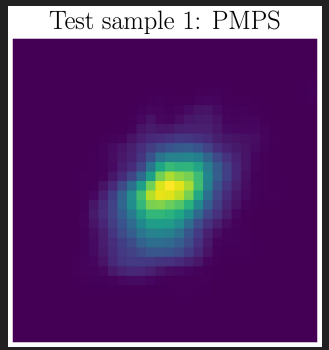
Simulator output

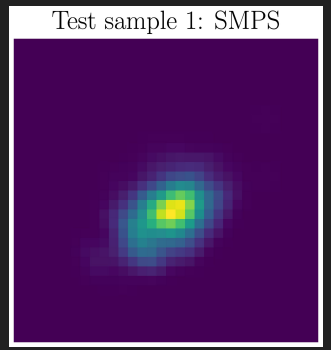
- We simulate 360,000 populations by varying the 5 magneto-rotational parameters and generate density maps for 3 surveys in $P\dot{P}$-plane (Graber et al., 2024).
  - $\mu_{\log P} \in$ Uniform(-1.5, -0.3)
  - $\sigma_{\log P} \in$ Uniform(0.1, 1.0)
  - $\mu_{\log B} \in$ Uniform(12.0, 14.0)
  - $\sigma_{\log B} \in$ Uniform(0.1, 1.0)
  - $a_{late} \in$ Uniform(-3.0, -0.5)

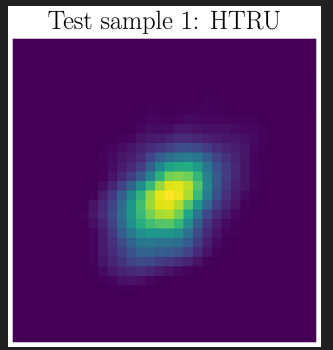
Test simulation for $\mu_{\log P} = -0.85$, $\sigma_{\log P} = 0.51$, $\mu_{\log B} = 13.19$, $\sigma_{\log B} = 0.96$ and $a_{late} = -0.86$.



$$\mathcal{P}(\log P_0) = \frac{1}{\sqrt{2\pi}\sigma_{\log P}} \exp\left(-\frac{\log P_0 - \mu_{\log P}}{2\sigma^2_{\log P}}\right)$$



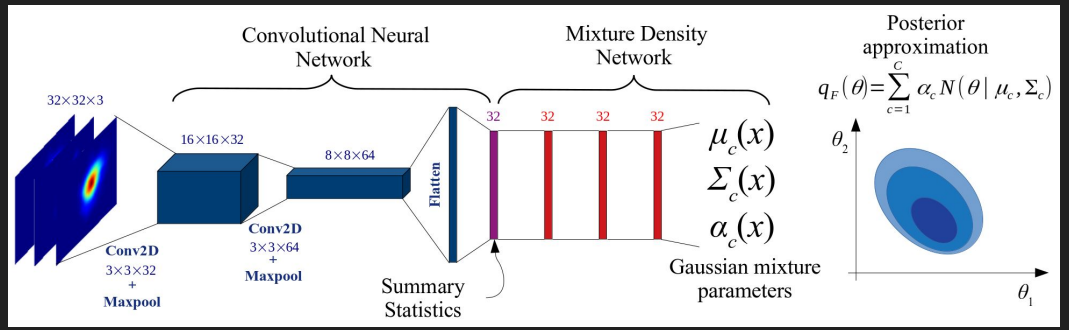




Neural network setup

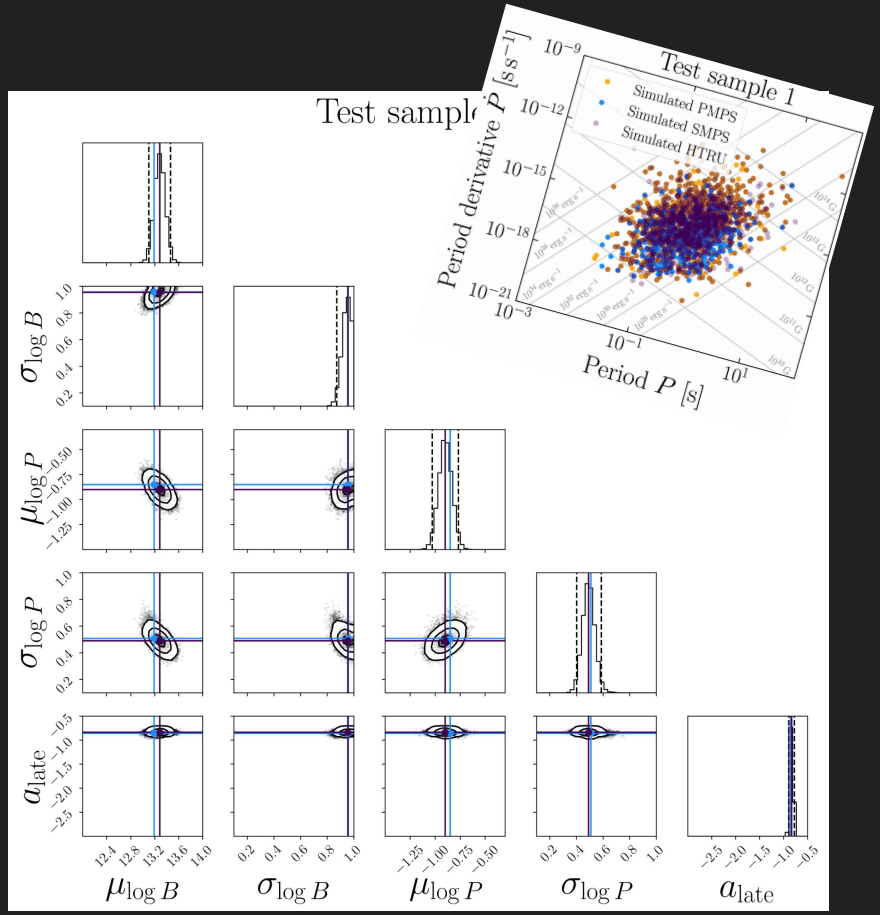
- To perform SBI, we use the **PyTorch package sbi** (Tejero-Cantero et al., 2020; https://www.mackelab.org/sbi/). Our trainable neural network has two parts:
  - **CNN** (see Ronchi et al., 2021): compresses the maps into a latent vector.
  - **Mixture density network (MDN)**: posterior is approximated by a mixture of 10 Gaussians components; we learn the means, stds and coefficients.

- We use Kaiming initialisation for the CNN and Xavier for the MDN, 89% of data for training, 10% for validation and 1% for testing, a batch size of 8 & learning rate of 5×10^{-4}.



Inference on test sample

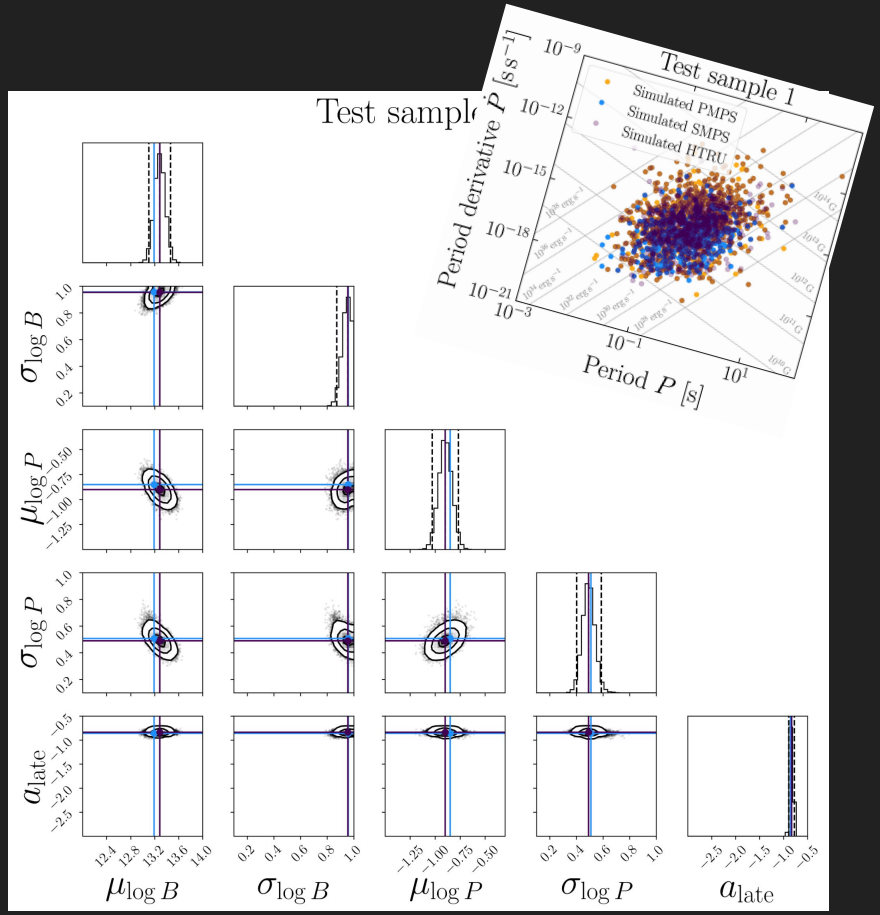
- As our conditional density estimator is represented by a neural network, we can **directly evaluate the posterior distributions for a given (test) observation**.



Inference on test sample

- As our conditional density estimator is represented by a neural network, we can **directly evaluate the posterior distributions for a given (test) observation**.

We recover **narrow and well-defined posteriors** for all five parameters that typically contain the ground truth (parameters used for the forward simulation) at the 95% credibility level.

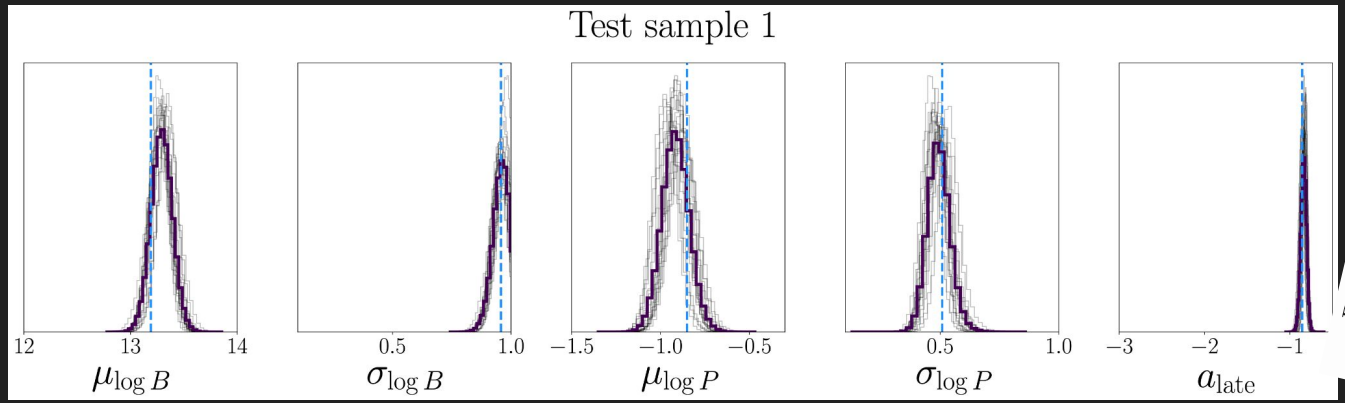


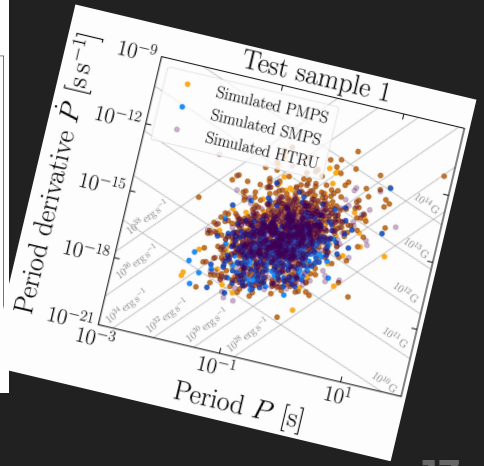
Ensemble of networks

- Our inferred posteriors are **sensitive to neural network setup**. To analyse the robustness of our results, we perform **22 ML experiments** by varying network hyper parameters, survey input, map resolution, preprocessing.
- We combine posteriors from 19 experiments with well-behaved training behaviour into an **ensemble posterior to obtain conservative predictions**.

Ensemble of networks

- Our inferred posteriors are **sensitive to neural network setup**. To analyse the robustness of our results, we perform **22 ML experiments** by varying network hyper parameters, survey input, map resolution, preprocessing.
- We combine posteriors from 19 experiments with well-behaved training behaviour into an **ensemble posterior to obtain conservative predictions**.



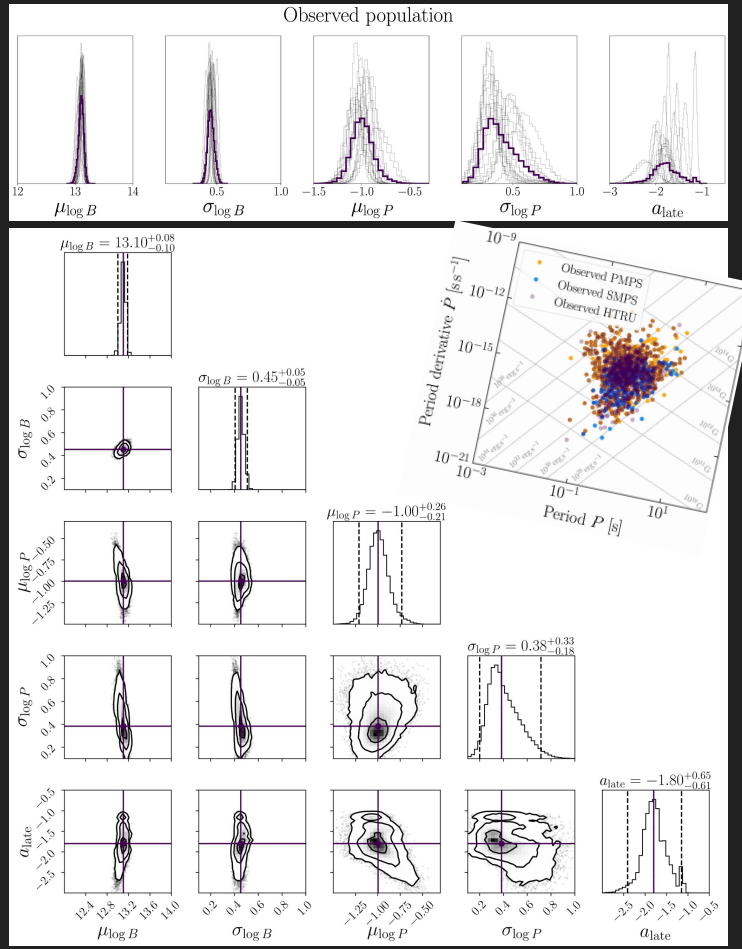


Observed population

- We apply our ensemble network to infer on the observed population of isolated radio pulsars at 95% CI:

$$\mu_{\log B} = 13.10^{+0.08}_{-0.10},$$
$$\sigma_{\log B} = 0.45^{+0.05}_{-0.05},$$

$$\mu_{\log P} = -1.00^{+0.26}_{-0.21},$$
$$\sigma_{\log P} = 0.38^{+0.33}_{-0.18},$$
$$a_{\text{late}} = -1.80^{+0.65}_{-0.61}.$$

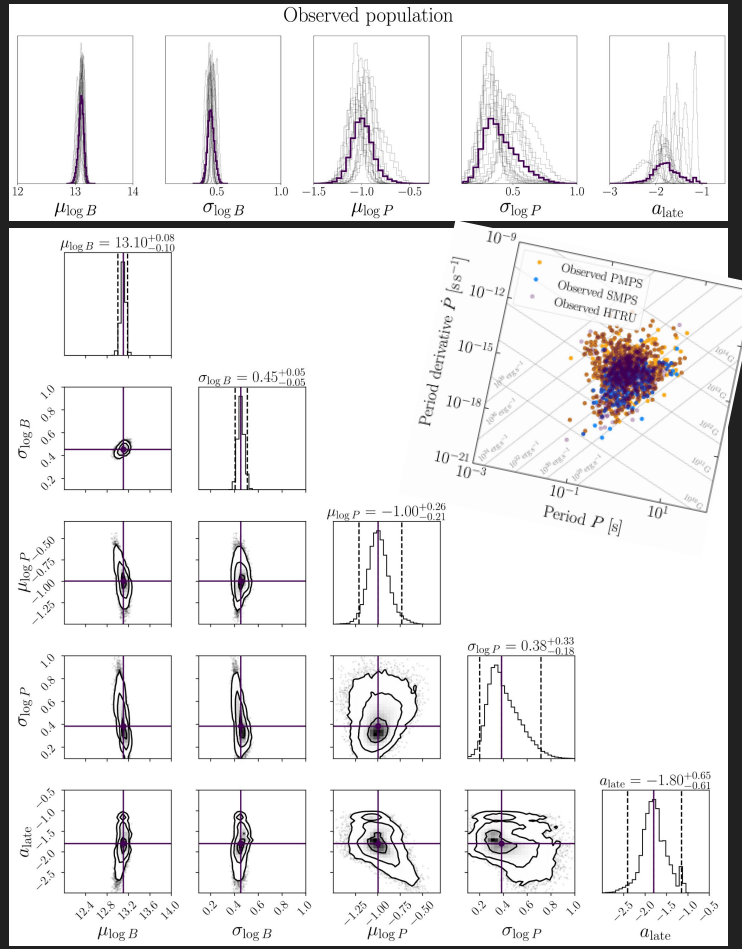


Observed population

- We apply our ensemble network to infer on the observed population of isolated radio pulsars at 95% CI:

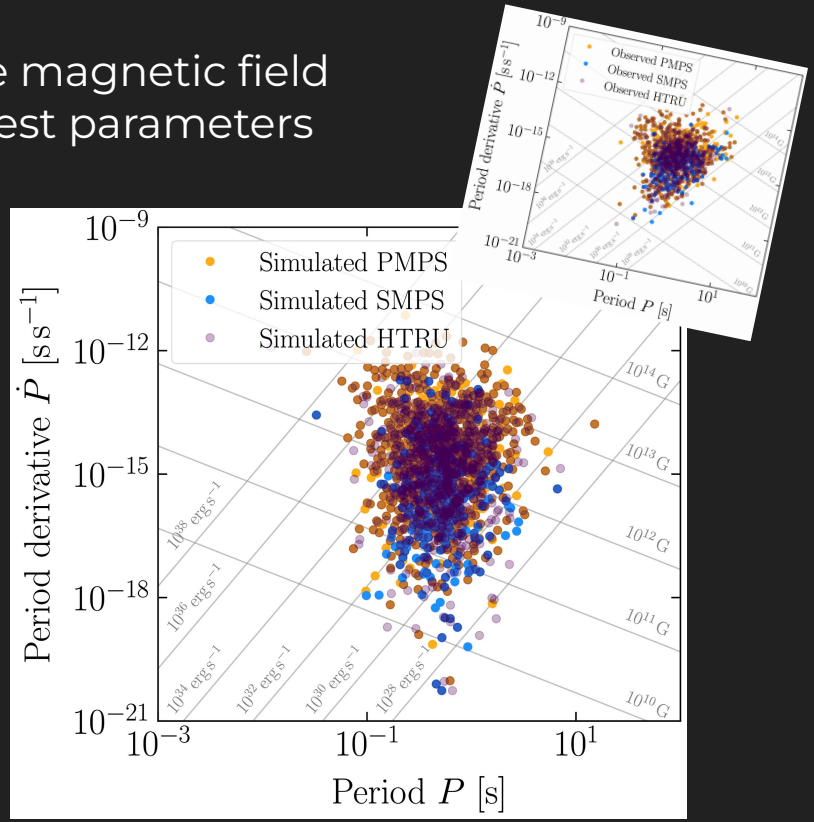
$$\mu_{\log B} = 13.10^{+0.08}_{-0.10},$$
$$\sigma_{\log B} = 0.45^{+0.05}_{-0.05},$$
$$\mu_{\log P} = -1.00^{+0.26}_{-0.21},$$
$$\sigma_{\log P} = 0.38^{+0.33}_{-0.18},$$
$$a_{\text{late}} = -1.80^{+0.65}_{-0.61}.$$

We find **tight constraints for B_0 and P_0** parameters. Larger uncertainties and bi-modal posterior in a_{late} hint at insufficient modelling of late-time B-field decay.



Two sanity checks

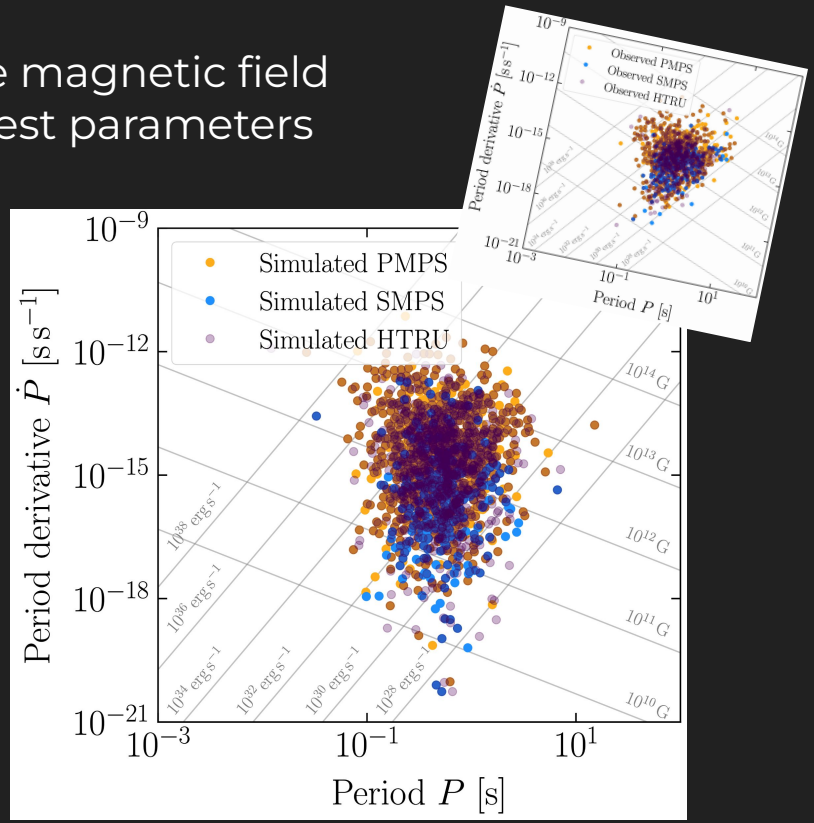
- Despite potential issues with the late-time magnetic field decay, rerunning our simulator with the best parameters produces a **reasonable PṖ diagram**.
- Our framework allows us to constrain the corresponding pulsar **birth rate:**



Two sanity checks

- Despite potential issues with the late-time magnetic field decay, rerunning our simulator with the best parameters produces a **reasonable P$\dot{P}$ diagram**.
- Our framework allows us to constrain the corresponding pulsar **birth rate:**

PMPS: $\sim 2.02 \pm 0.02$ neutron stars per century,
SMPS: $\sim 1.84 \pm 0.03$ neutron stars per century,
HTRU: $\sim 1.66 \pm 0.02$ neutron stars per century,

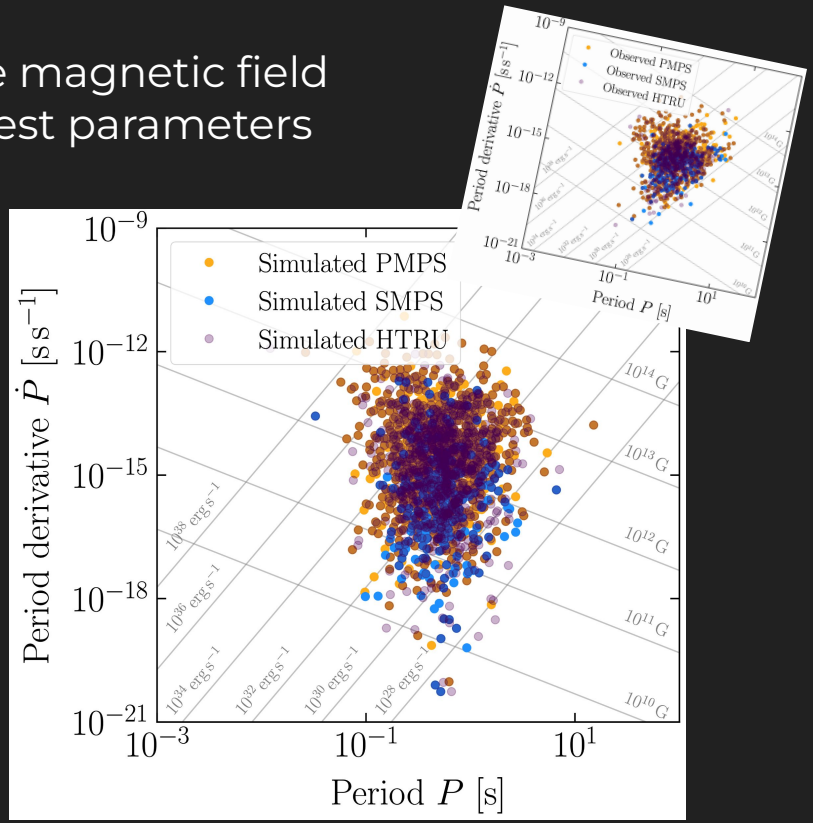


Two sanity checks

- Despite potential issues with the late-time magnetic field decay, rerunning our simulator with the best parameters produces a **reasonable P$\dot{P}$ diagram**.
- Our framework allows us to constrain the corresponding pulsar **birth rate:**

PMPS: $\sim 2.02 \pm 0.02$ neutron stars per century,
SMPS: $\sim 1.84 \pm 0.03$ neutron stars per century,
HTRU: $\sim 1.66 \pm 0.02$ neutron stars per century,

This is in good agreement with core-collapse supernova rates, i.e., 1.63 ± 0.46 per century (Rozwadowska et al., 2021).



Take-home points and outlook

- **Population synthesis bridges gap between detected pulsars and the invisible population.**
- **It allows us to constrain birth rates and NS birth properties.**

Take-home points and outlook

- Population synthesis bridges gap between detected pulsars and the invisible population.
- It allows us to constrain birth rates and NS birth properties.

- SBI has opened up the possibility for robust statistical inference with complex simulators.
- We successfully used SBI to infer magneto-rotational parameters.

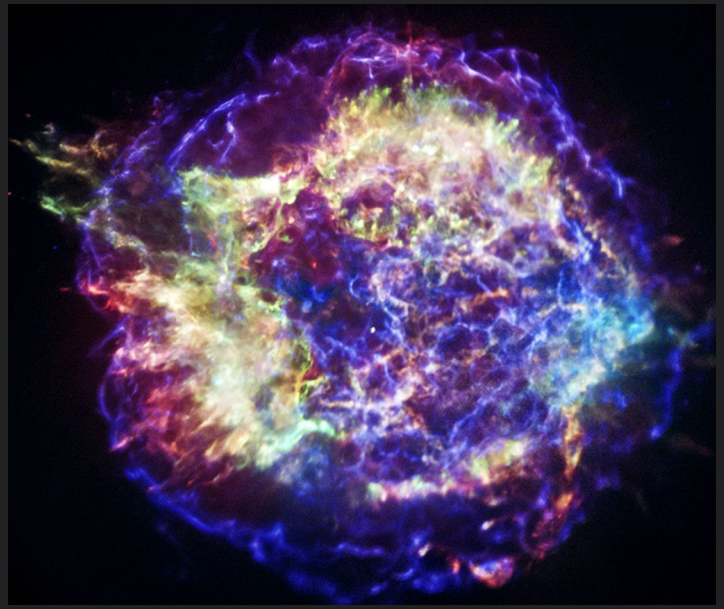
Take-home points and outlook

- Population synthesis bridges gap between detected pulsars and the invisible population.
- It allows us to constrain birth rates and NS birth properties.

- SBI has opened up the possibility for robust statistical inference with complex simulators.
- We successfully used SBI to infer magneto-rotational parameters.

- We are currently working to improve our approach:
  - Model multi-wavelength emission and observations.
  - Remove the need for large training datasets by using an sequential SBI approach that adaptively simulates data.

THANK YOU



Cassiopeia A supernova remnant
(credit: NASA/CXC/SAO)

Magneto-rotational evolution - analytics

- For a given initial rotation period, magnetic field strength and inclination angle (uniformly distributed along the sphere), the evolution of a neutron star is determined by **three ordinary differential equations**. The first two are

$$\dot{P} = \frac{\pi^2}{c^3}\frac{B^2 R^6}{IP}\left(\kappa_0 + \kappa_1 \sin^2\chi\right) \qquad \dot{\chi} = -\frac{\pi^2}{c^3}\frac{B^2 R^6}{IP^2}\left(\kappa_2 \sin\chi\cos\chi\right)$$

  The κ are determined from simulations. For a **realistic pulsar magnetosphere** filled with plasma we have $\kappa_0 \simeq \kappa_1 \simeq \kappa_2 \simeq 1$ (Spitkovsky 2006; Philippov et al., 2014).

- From the induction equation, we deduce for the B-field (Aguilera et al., 2008):

$$\dot{B} = -\frac{c^2 B}{4\pi\sigma L^2} - \frac{cB^2}{4\pi e n_e L^2} \quad \text{with} \quad \tau_{\rm Ohm} = \frac{4\pi\sigma L^2}{c^2} \sim 10^6\ \text{yr}, \quad \tau_{\rm Hall} = \frac{4\pi e n_e L^2}{cB} \sim 10^4 B_{12}\ \text{yr}$$

University of Hertfordshire UH

v.graber@herts.ac.uk

Improved B-field prescription

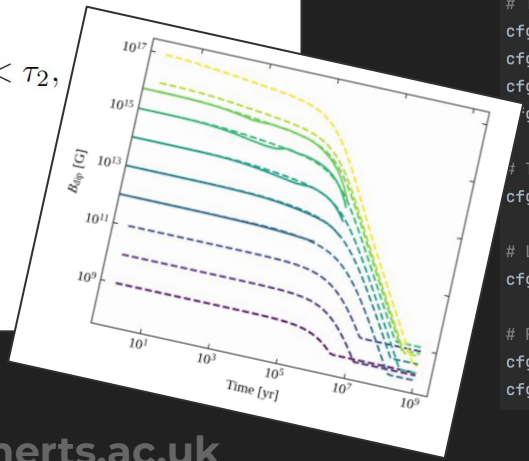
- To provide a more **realistic description of the magnetic field decay**, we **fit an analytical function** to B-field evolution curves obtained from 2D magneto-thermal simulations (Viganó et al., 2021) for various fields. As they are only valid up to 10^6 yr, we extend the **late-time behaviour with a power law** as follows:

$$B(t) = B_0 \left(1 + \frac{t}{\tau_1}\right)^{a_1} \left(1 + \frac{t}{\tau_2}\right)^{a_2 - a_1} \left(1 + \frac{t}{\tau_{\rm late}}\right)^{a_{\rm late} - a_2} \quad \text{for} \quad \tau_1 < \tau_2 < \tau_{\rm late},$$

$$B(t) = B_0 \left(1 + \frac{t}{\tau_1}\right)^{a_1} \left(1 + \frac{t}{\tau_{\rm late}}\right)^{a_{\rm late} - a_1} \quad \text{for} \quad \tau_1 < \tau_{\rm late} < \tau_2,$$

$$B(t) = B_0 \left(1 + \frac{t}{\tau_{\rm late}}\right)^{a_{\rm late}} \quad \text{for} \quad \tau_{\rm late} < \tau_1 < \tau_2.$$

with $\tau_1 \equiv A_1 B_0^{b_1} \quad \tau_2 \equiv A_2 B_0^{b_2}$




```
# Power-law indices.
cfg["a1"]: float = -0.13
cfg["a2"]: float = -3.0

# Timescale parameters, normalizatio
cfg["A1"]: float = 1.0e14
cfg["b1"]: float = -0.8
cfg["A2"]: float = 6.0e8
cfg["b2"]: float = -0.2

# Timescale in [yr] when transitioni
cfg["tau_late"]: float = 2.0e6

# Late time power-law index.
cfg["a_late"]: float = -2.0

# Parameters for a log-normal distri
cfg["B_millisec_mean"] = 8.5
cfg["B_millisec_sigma"] = 0.5
```

University of Hertfordshire UH

v.graber@herts.ac.uk

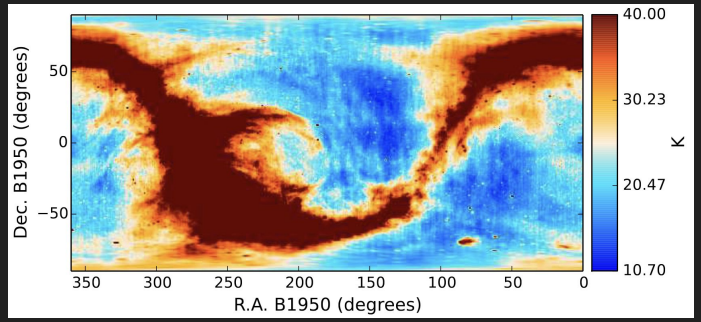
Radiometer equation

$$S_{\mathrm{radio}} = \frac{L_{\mathrm{radio}}}{\Omega_{\mathrm{beam}} d^2}$$

- Following Lorimer & Kramer (2005), we convert the intrinsic (bolometric) radio flux S_{radio} into a flux density measured at a given frequency. We then account for the fact that the intrinsic pulse width is broadened to w_{eff} (due to inter-stellar scattering, dispersion and finite sampling resolution of a detector) and compute the period-averaged flux S_{mean} observed by a radio telescope.

- With these estimates and information specific to the PMPS, SMPS and HTRU surveys, we **determine the signal-to-noise-ratio** for a given simulated pulsar using the **radiometer equation**:

$$S/N = \frac{S_{\mathrm{mean}} G \sqrt{N_{\mathrm{pol}} \Delta\nu \Delta t_{\mathrm{obs}}}}{\beta \left(T_{\mathrm{rec}} + T_{\mathrm{sky}}(l,b)\right)} \sqrt{\frac{P - w_{\mathrm{eff}}}{w_{\mathrm{eff}}}}$$



Haslam sky temperature map taken from Remazeilles et al., (2015).

University of Hertfordshire UH

v.graber@herts.ac.uk

SBI approaches

- Different approaches (all relying on deep learning) exist to **learn a probabilistic association** between the simulated data and the underlying parameters. These algorithms essentially focus on different pieces of Bayes' theorem:
  - Neural Posterior Estimation (NPE) (e.g., Papamakarios & Murray, 2016)
  - Neural Likelihood Estimation (NLE) (e.g., Papamakarios et al., 2019)
  - Neural Ratio Estimation (NRE) (e.g., Hermans et al., 2020; Delaunoy et al., 2022)

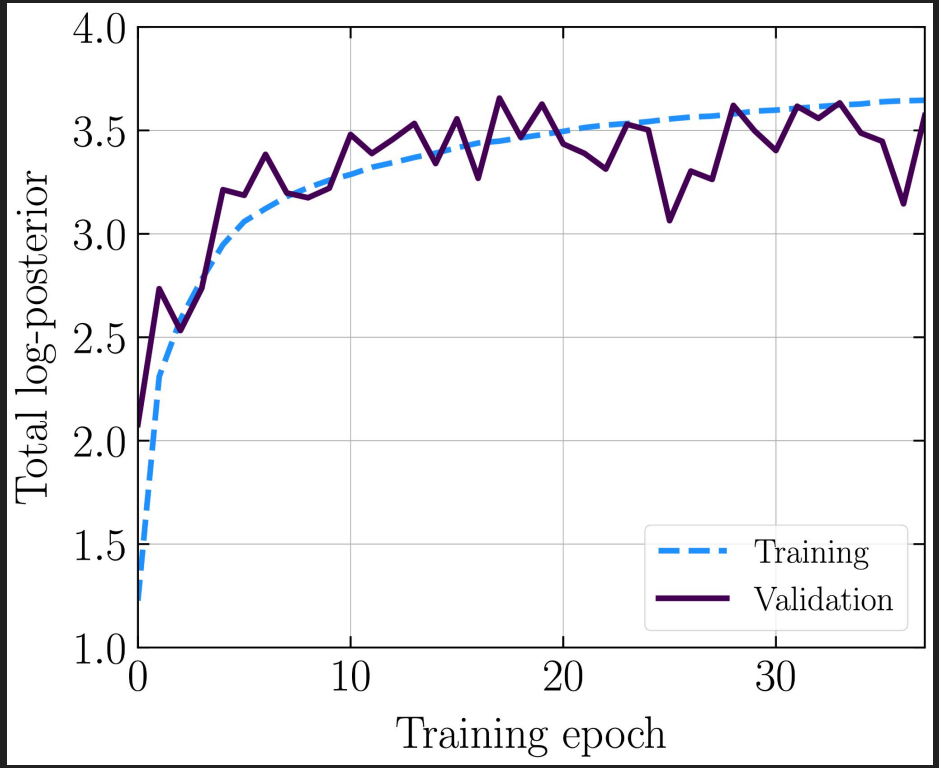
We focus on NPE. This allows us to **directly learn the posterior distribution**. In contrast, NLE and NRE need an extra (potentially time consuming) MCMC sampling step to construct a posterior.

- All methods exist in **sequential form** (SNPE, SNLE, SNRE), **which adds a fifth step to workflow**. Instead of sampling from the prior, we adaptively generate simulations from the posterior. This **typically requires fewer simulations**.

v.graber@herts.ac.uk

Neural network training behaviour

- **Typical learning behaviour** for our simulation-based inference experiments.
- The **log-probability increases** as a function of training epochs for the training and validation loss.
- We see some variation in the validation loss but **little overfitting**.



University of Hertfordshire UH
v.graber@herts.ac.uk

22 different ML experiments

No.	Res	Surveys	Frac (%)	Input	Comp	BS	LR	CNN	VM	Epochs	Time (s)	TM
1	32	PMPS, SMPS, HTRU	100	std	10	8	0.0005	baseline	3.65	38	9,373	3.64
2	32	PMPS, SMPS, HTRU	100	std	10	8	0.0005	**deep**	3.71	49	14,292	3.71
3	**64**	PMPS, SMPS, HTRU	100	std	10	8	0.0005	baseline	3.55	55	78,837	3.54
4	**64**	PMPS, SMPS, HTRU	100	std	10	8	0.0005	**deep**	3.64	89	128,119	3.64
5	32	PMPS, SMPS, HTRU	**75**	std	10	8	0.0005	baseline	3.74	71	13,232	3.78
6	32	PMPS, SMPS, HTRU	**50**	std	10	8	0.0005	baseline	3.56	58	7,000	3.55
7⋆	32	PMPS, SMPS, HTRU	100	**norm**	10	8	**0.01**	baseline	3.47	30	7,445	3.73
8	32	PMPS, SMPS, HTRU	100	**norm**	10	8	**0.001**	baseline	9.66	54	13,015	9.60
9	32	PMPS, SMPS, HTRU	100	std	**8**	8	0.0005	baseline	3.74	52	12,389	3.73
10	32	PMPS, SMPS, HTRU	100	std	**5**	8	0.0005	baseline	3.83	118	27,973	3.86
11	32	PMPS, SMPS, HTRU	100	std	10	**16**	0.0005	baseline	3.99	85	10,476	3.97
12	32	PMPS, SMPS, HTRU	100	std	10	**32**	0.0005	baseline	4.11	79	5,346	4.06
13	32	PMPS, SMPS, HTRU	100	std	10	8	**0.001**	baseline	3.36	61	14,785	3.33
14	32	PMPS, SMPS, HTRU	100	std	10	8	**0.0001**	baseline	4.22	75	18,369	4.22
15	32	**HTRU**	100	std	10	8	0.0005	baseline	3.43	63	15,568	3.42
16	32	**SMPS, HTRU**	100	std	10	8	0.0005	baseline	3.58	40	9,979	3.59
17	32	**PMPS, SMPS**	100	std	10	8	0.0005	baseline	3.41	69	16,937	3.41
18⋆	**64**	PMPS, SMPS, HTRU	**50**	std	10	8	0.0005	baseline	3.45	47	5,766	3.44
19	32	PMPS, SMPS, HTRU	100	**norm**	10	**32**	**0.001**	baseline	10.05	44	2,864	10.20
20	32	PMPS, SMPS, HTRU	100	**norm**	10	**32**	**0.0001**	baseline	10.31	90	5,815	10.49
21	32	PMPS, SMPS, HTRU	100	**norm**	10	**16**	**0.001**	baseline	9.82	77	9,901	9.98
22⋆	32	PMPS, SMPS, HTRU	100	**norm**	10	**16**	**0.0001**	baseline	10.45	124	15,603	10.55

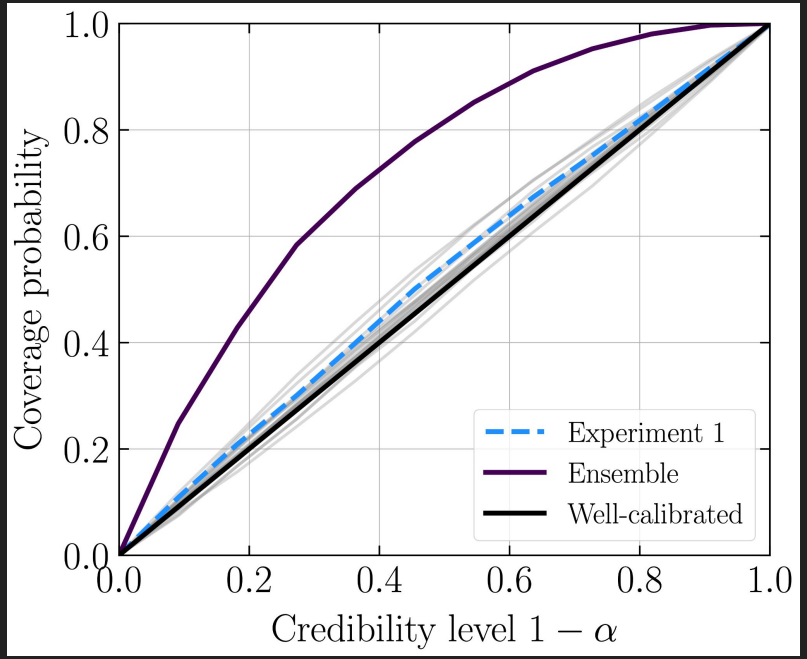
University of Hertfordshire UH

v.graber@herts.ac.uk

Quality check: coverage calculation

- A measure that provides information on the quality of our posterior is the **coverage**. I.e., we take a frequentist approach and determine how often our true parameters fall into the (1-α)% credible interval under the model.

- By definition, a **(1-α)% credibility level covers the true parameter in (1-α)% of our analyses**, resulting in a diagonal line. In this case, the **posterior is neither conservative (too wide; above diagonal) nor overconfident (too narrow; below diagonal)**.

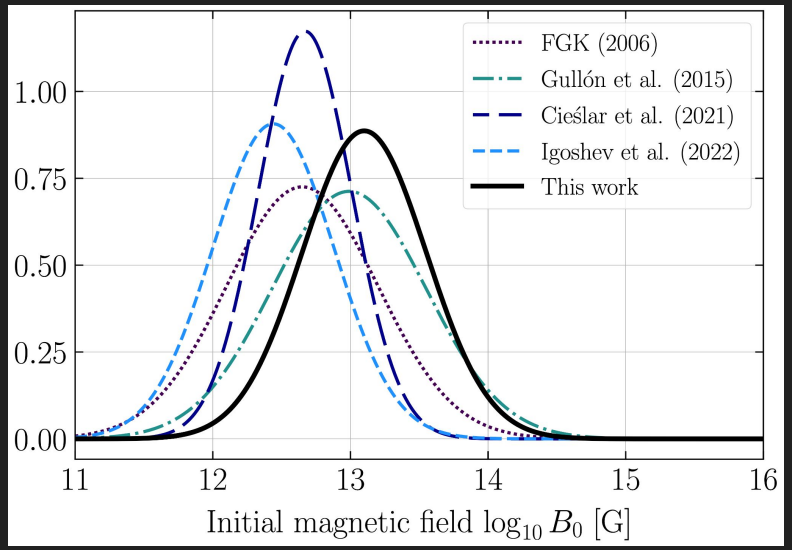


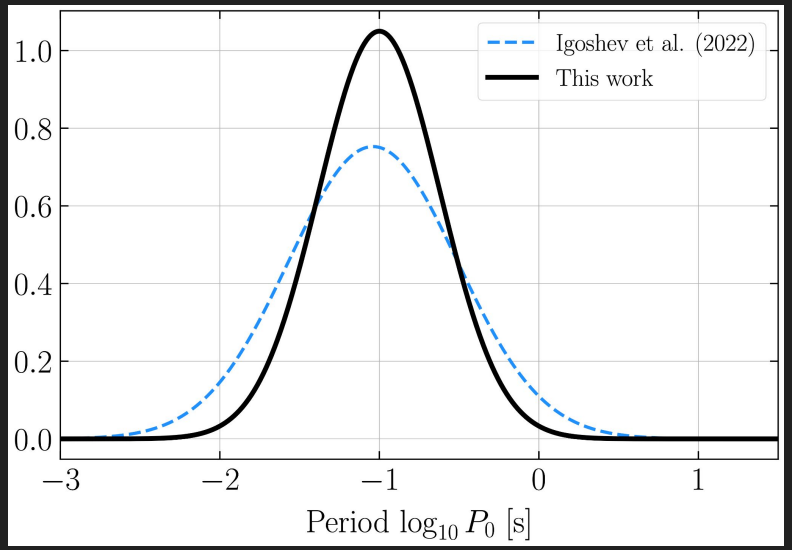
University of Hertfordshire UH
v.graber@herts.ac.uk

Comparison to other studies

- A comparison with earlier studies is difficult due to different simulation models and inference methods.

References	$\mu_{\log B}$	$\sigma_{\log B}$	$\mu_{\log P}$	$\sigma_{\log P}$
Faucher-Giguère & Kaspi (2006)	12.65	0.55	–	–
Gullón et al. (2015)	12.99	0.56	–	–
Cieślar et al. (2020)	$12.67^{+0.01}_{-0.02}$	$0.34^{+0.02}_{-0.01}$	–	–
Igoshev et al. (2022)	12.44	0.44	$-1.04^{+0.15}_{-0.20}$	$0.53^{+0.12}_{-0.08}$
This work	$13.10^{+0.04}_{-0.05}$	$0.45^{+0.03}_{-0.02}$	$-1.00^{+0.11}_{-0.10}$	$0.38^{+0.16}_{-0.10}$





University of Hertfordshire UH
v.graber@herts.ac.uk