Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

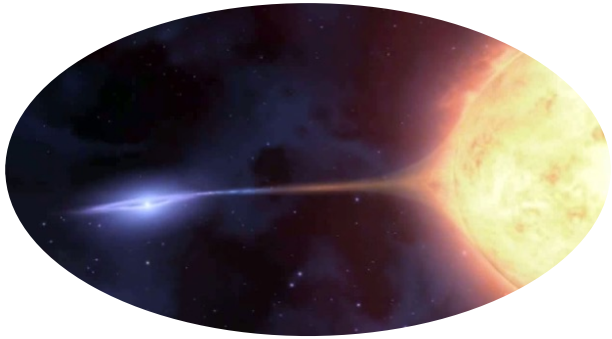
Simulation-Based Inference for realistic Supernova Type Ia data

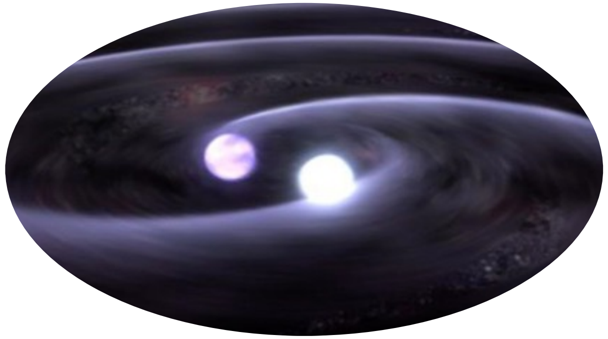
Roberto Trotta, SISSA & Imperial

Spoke 3 General Meeting, Elba 5-9 / 05, 2024

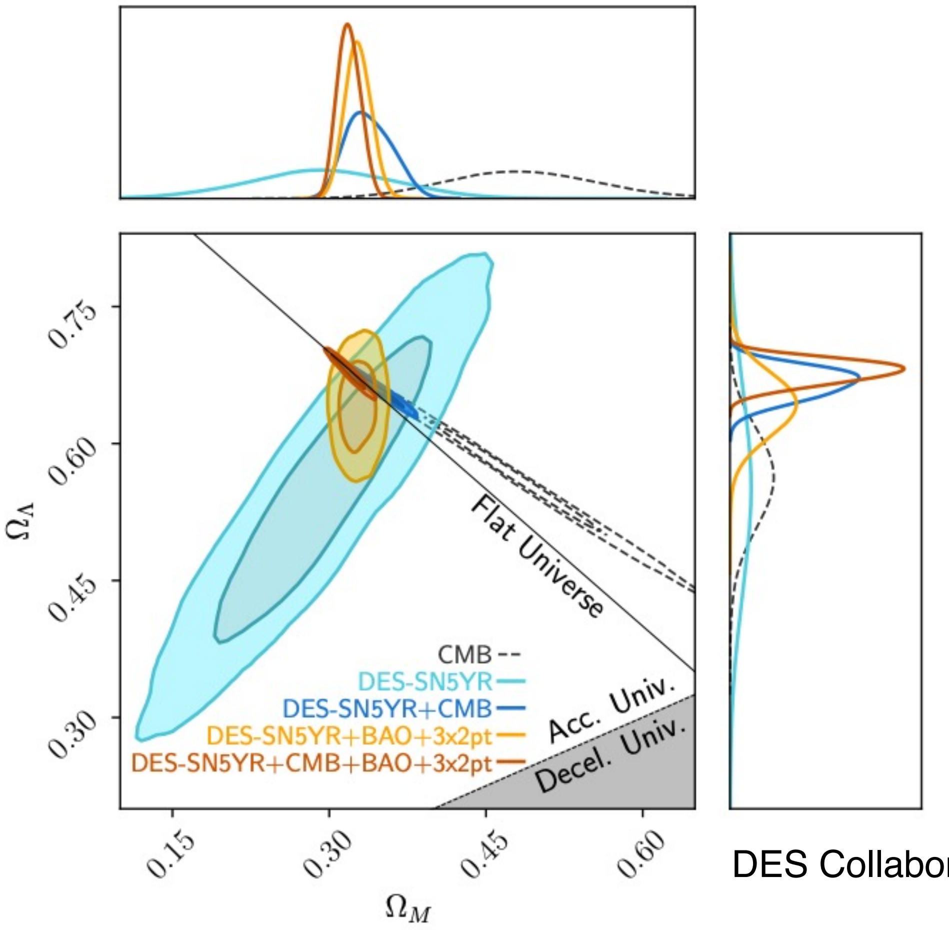
Scientific Rationale

Supernova Type Ia (SNIa): thermonuclear explosion of a CO white dwarf in a binary system; can be standardized from its lightcurve to measure the redshift-distance relation, and from this cosmology.





Fraction of dark energy content



DES Collaboration (2024)

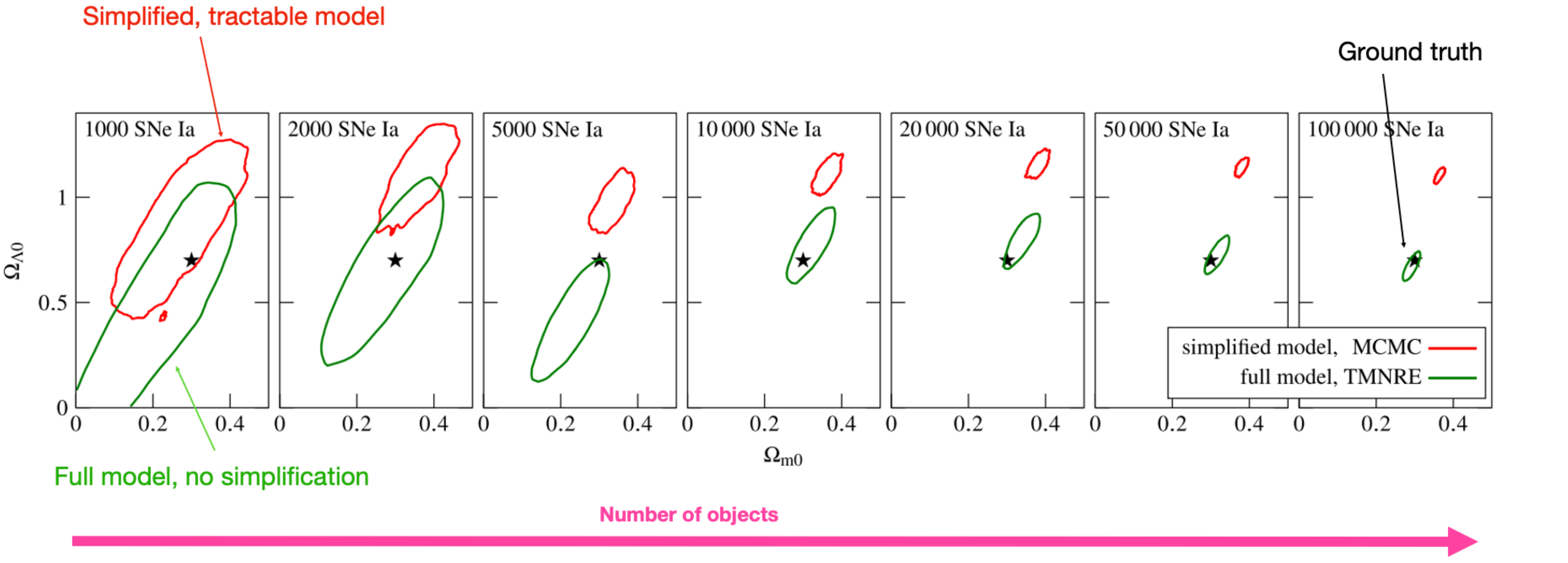
Fraction of matter (baryonic+dark) content

Scientific Rationale

- With the advent of large surveys such as LSST/Vera Rubin, the number of supernova Type Ia (SNIa) candidates will increase by orders of magnitude, from ~ 2000 to O(10^5) .
- Existing statistical methods are insufficient to deal with the resulting data complexity.
- Simplified models (eg linearized Gaussian likelihood) won't be sufficiently accurate to deal with the increased statistical accuracy from a O(10^5) sample size, and systematics will dominate (and potentially bias cosmological inference!).
- Other challenges:
- spectroscopy available only for a small subset of objects, leading to non-Ia contamination
- redshift needs to be estimated from lightcurve alone
- object-level modelling of local dust fundamental to account for absorption/reddening
- inclusion of NIR data where available
- observational selection effects need to be addressed (see **StratLearn** project; **Chiara Moretti**)

Scientific Rationale

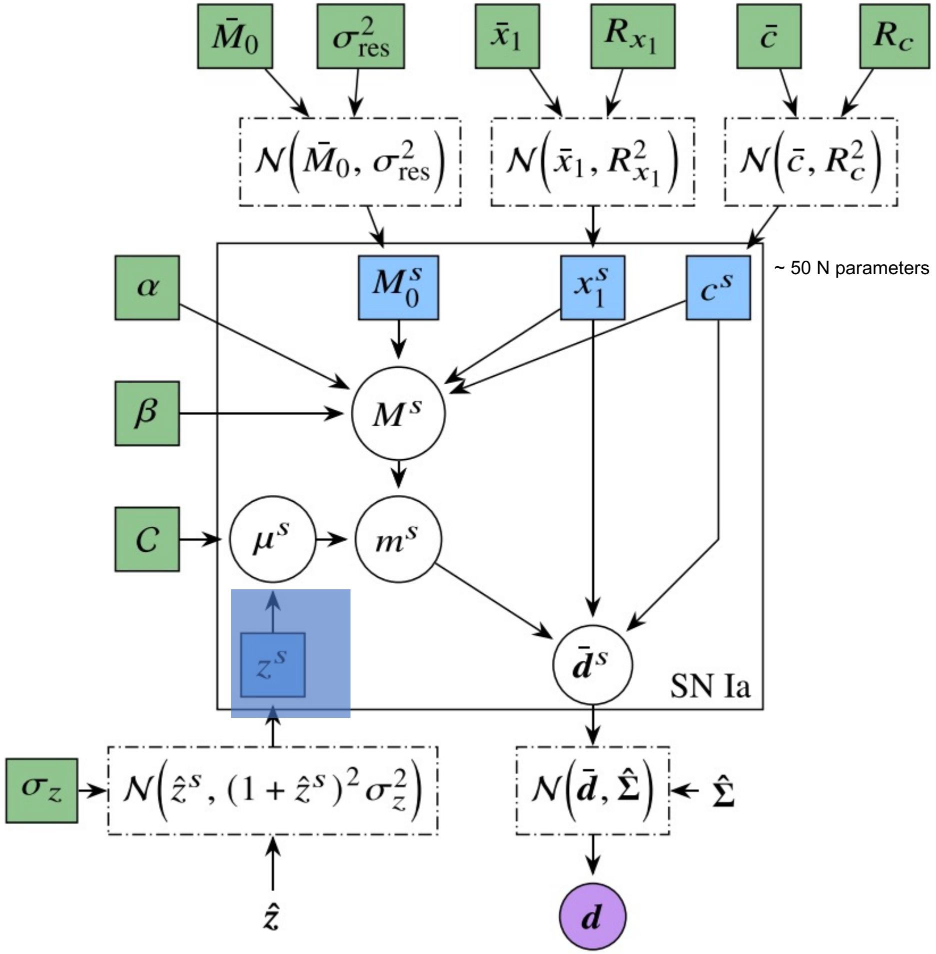
As data grow in size, precision increases but accuracy might decrease as simplifications that make the model tractable (linear propagation of errors, Gaussianity) become increasingly poor.



Karchev, RT & Weniger (2022)

Technical Objectives, Methodologies and Solutions

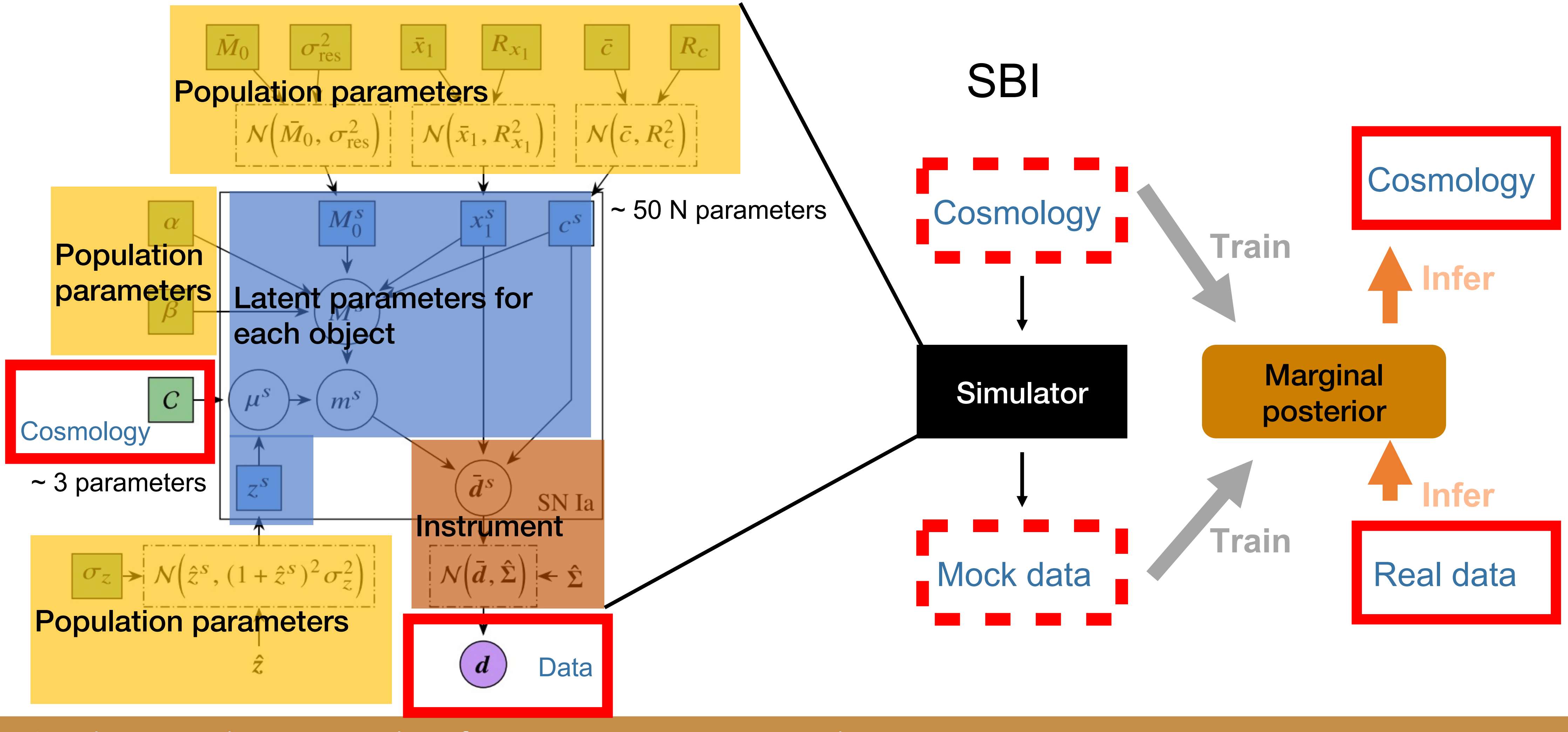
BHM:



- We propose to replace classical Bayesian Hierarchical modelling (BHM) with state-of-the art **Simulation Based Inference** (SBI).
- Idea: **replace** complex astrophysics with explicit forward modeling inside a fast simulator, going Cosmology → data
- **Exploit** neural network techniques to perform a modern version of ABC (Approximate Bayesian Computation): **Truncated Marginal Neural Ratio Estimation** (TMNRE)

~ 200 N data points

Technical Objectives, Methodologies and Solutions

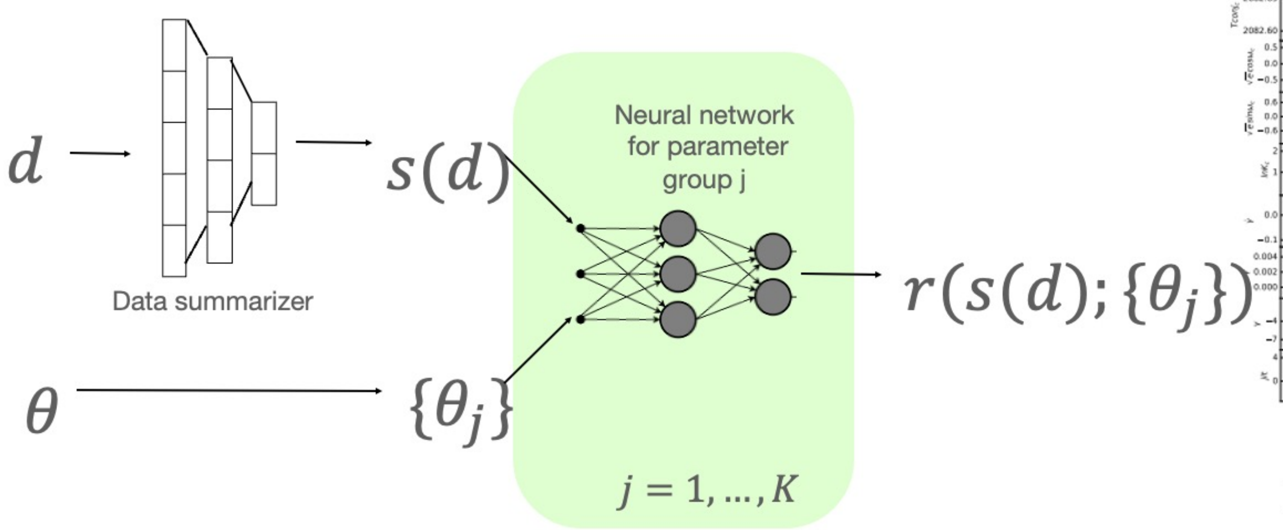


Technical Objectives, Methodologies and Solutions

We are interested in marginal inference on a small-dimensional (1- or 2-d) subset of parameters, $\{\theta_j\}$, i.e. we seek the marginal posterior:

$$p(\{\theta_j\}|d) = \int d\{\theta_{i\neq j}\}p(\theta|d)$$

In marginal NRE, we train a separate network to learn each marginal of interest from learnt data summaries:



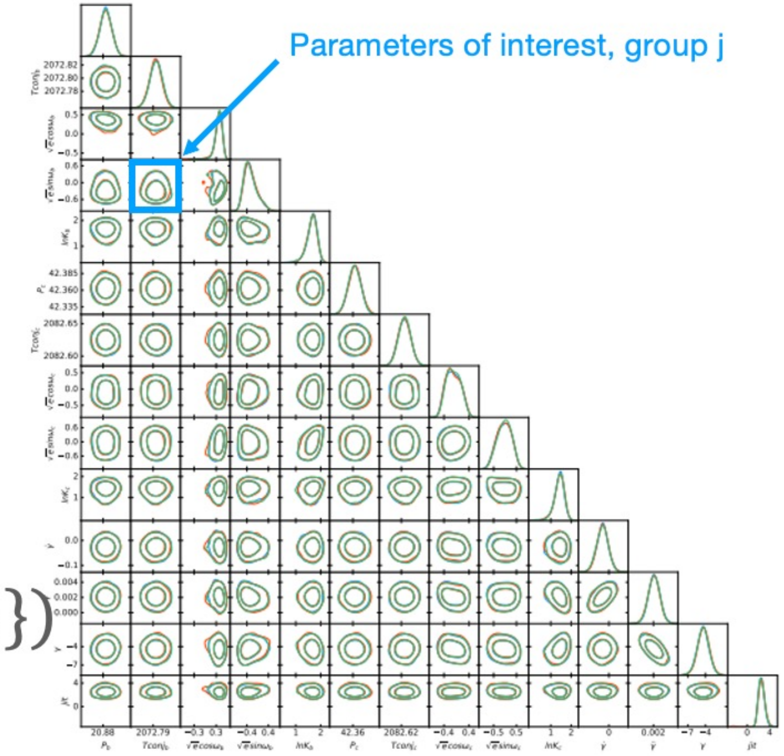


Figure from Karamanis (2022)

Accomplished Work

Kosio Karchev
PhD student

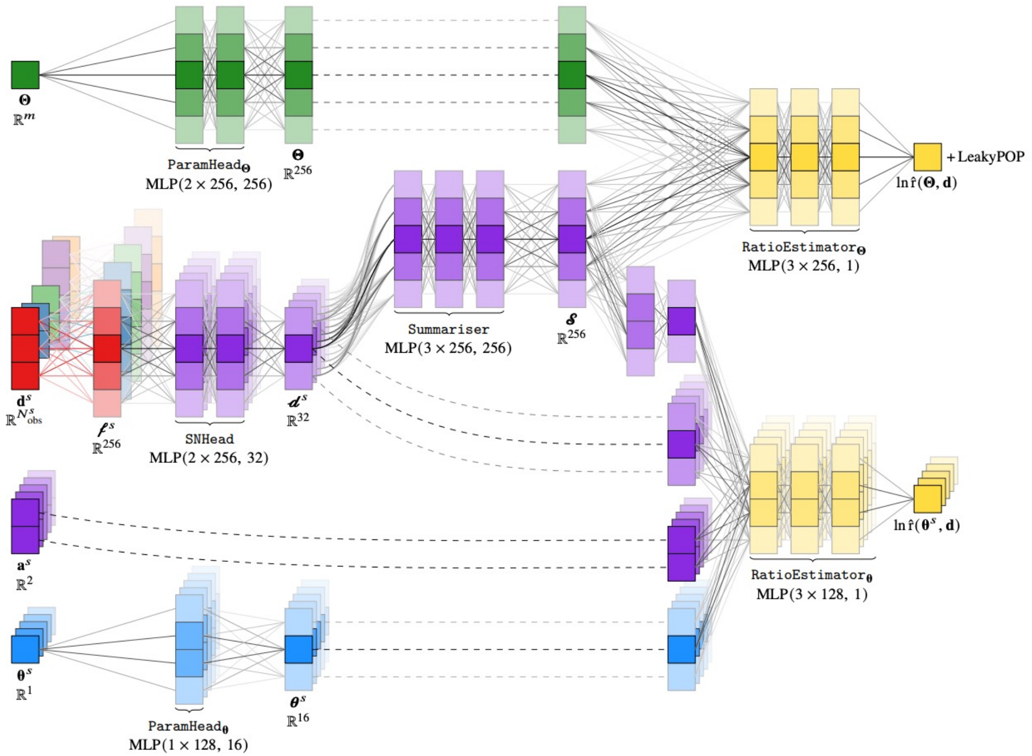
- Proof-of-concept paper on simplified simulations published; demonstrated scalability of our method to 100,000 SNIe:
  Karchev, Trotta & Weniger, «SICRET: Supernova Ia Cosmology with truncated marginal neural Ratio EsTimation», MNRAS 520 (2023) 1056-1072, arXiv:2209.06733;
  codes published:
  https:// github.com/kosiokarchev/clipppy
  https:// github.com/kosiokarchev/phytorch

- Demonstration of feasibility of Bayesian model comparison with TMNRE in models with over 4,000 latent variables, application to real data:
  Karchev, Trotta & Weniger (2023), «SimSIMS: Simulation-based Supernova Ia Model Selection with thousands of latent variables», NeurIPS 2023 workshop Machine Learning and the Physical Science, arXiv: 2311.15650

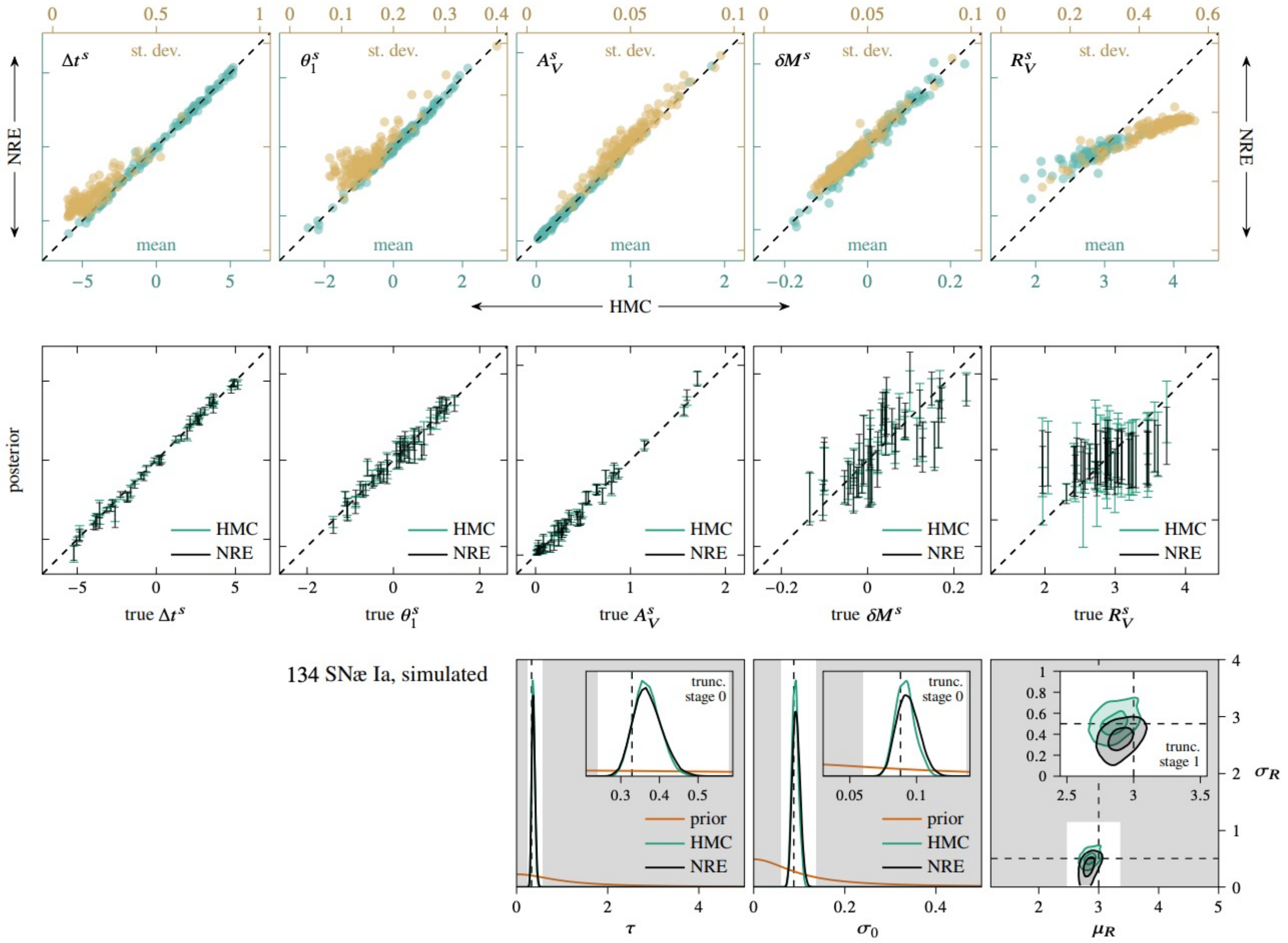
- Upgraded to realistic simulations and application to real data, validated via HMC:
  Karchev, Grayling, Boyd, Trotta, Mandel & Weniger (2024), «SIDE-real: Truncated marginal neural ratio estimation for Supernova Ia Dust Extinction with real data», MNRAS in print, arxiv: 2403.07871, code published: https://github.com/kosiokarchev/slicsim

Results

- Custom-designed neural network for TMNRE with SNIa LC data:



- Full validation on realistic simulations against HMC posteriors, application to CSP data:

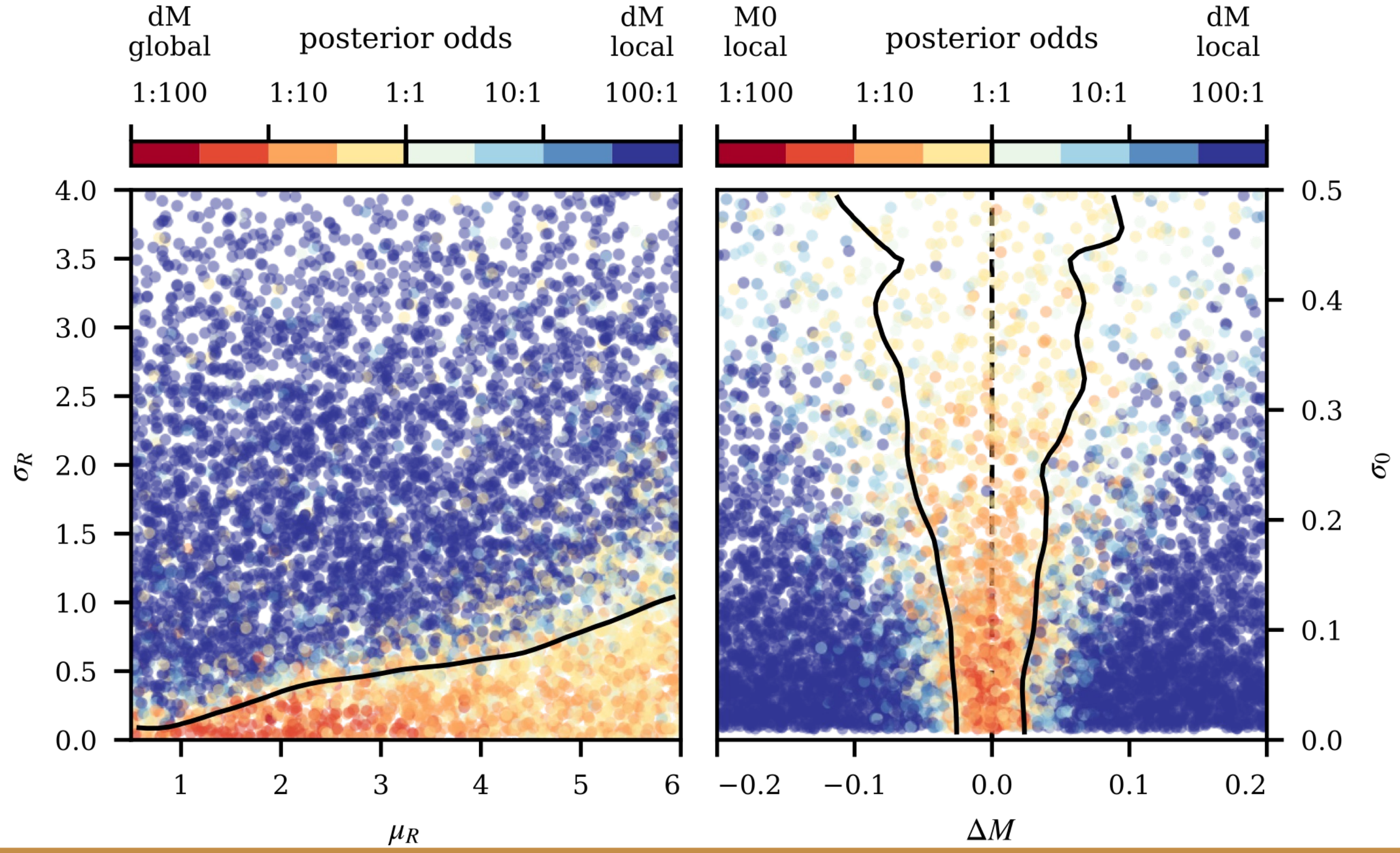


Results

- Trained a fully connected NN on multi-class labels using a cross-entropy loss, with thousands of latent variables. Validation matrix:

true class \ predicted class	M0 global	dM global	M0 local	dM local	M0 split	dM split
M0 global	64 %	16 %	13 %	3.3 %	3.2 %	0.79 %
dM global	17 %	63 %	3.5 %	13 %	0.83 %	3.1 %
M0 local	17 %	4.2 %	49 %	12 %	15 %	3.6 %
dM local	4.5 %	17 %	13 %	46 %	4 %	15 %
M0 split	4.1 %	1 %	16 %	3.8 %	61 %	14 %
dM split	1.1 %	4.6 %	4.2 %	15 %	16 %	59 %

- Model selection outcome maps in parameter space



Next Steps and Expected Results

- Integration of **observational selection effects** for high-redshift SNIa data analysis: use of deep sets to deal with variable number of objects, and different length of LC data. Requires self-consistent modelling of underlying SNIa rates as a function of redshift.

- Extension to **photometric-only SNIa data**: requires redshift estimation and inclusion of contamination (non Ia's) into the forward simulator.

- Joint estimation with **host galaxy data** and modelling of **peculiar velocities** at the same time as cosmology (relevant for low-z anchor). Requires integration of stellar population synthesis codes into host simulator, and parameterized model for peculiar velocities (Rahman et al, 2022).

- **Expected result:** a fully-integrated, end-to-end cosmological SNIa data analysis pipeline capable of handling up to 100,000 objects, including systematics in a principled manner.

Timescale, Milestones and KPIs (by Dec 2024)

- Integration of **observational selection effects** in proof-of-concept paper: **June 2024**. Paper and code published.

- Extension **to realistic (DES-like) selection effects** and application to real data: **October 2024**. Paper and code published.

- Proof-of-concept paper on joint analysis of galaxy data: **Nov 2024.** Paper and code published.

- Three outreach events at science festivals etc by **Dec 2024**.

Thank you!

www.datascience.sissa.it
www.robertotrotta.com