Simulation-based Inference for Stellar Parameter Estimation

Summer Undergraduate Data Science (SUDS)

Yihan Wang

Supervised by Prof. Joshua Speagle
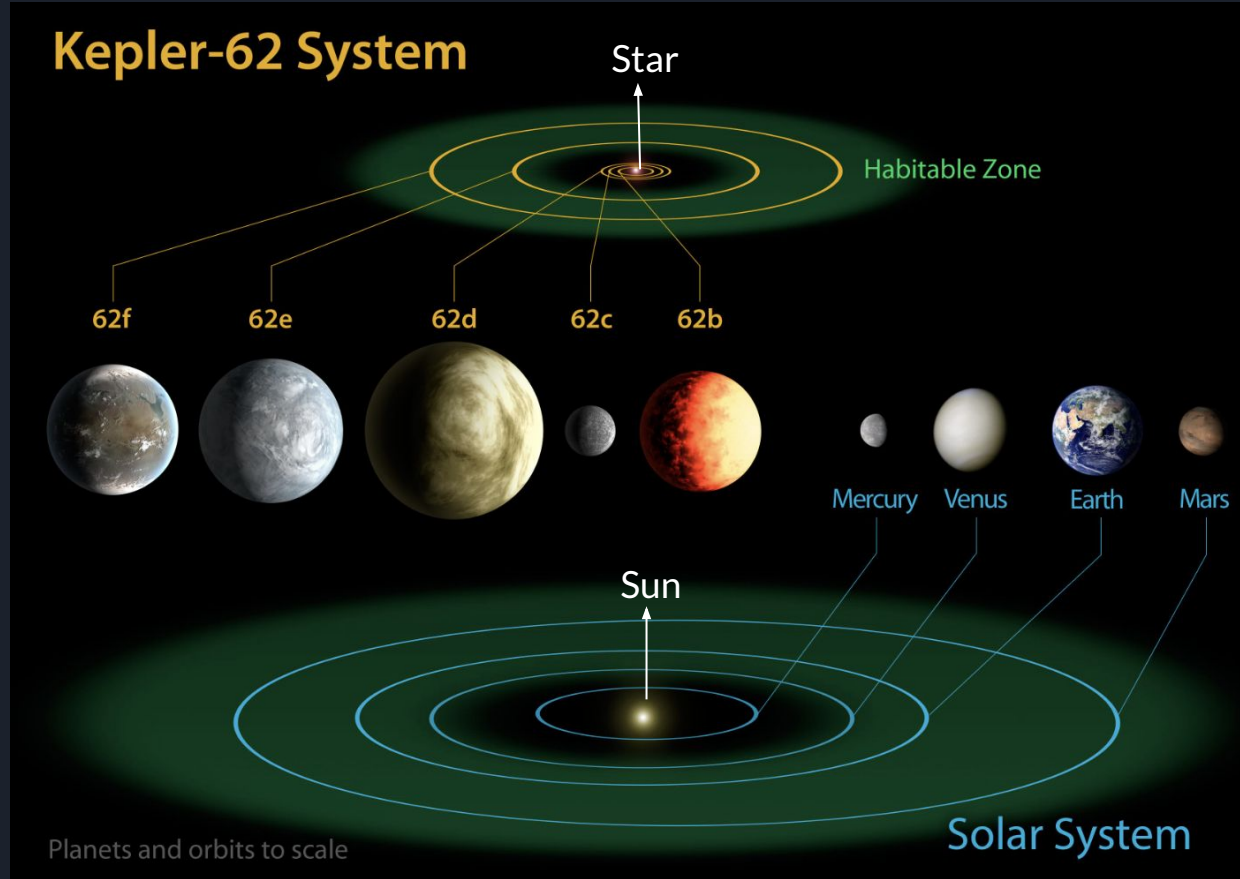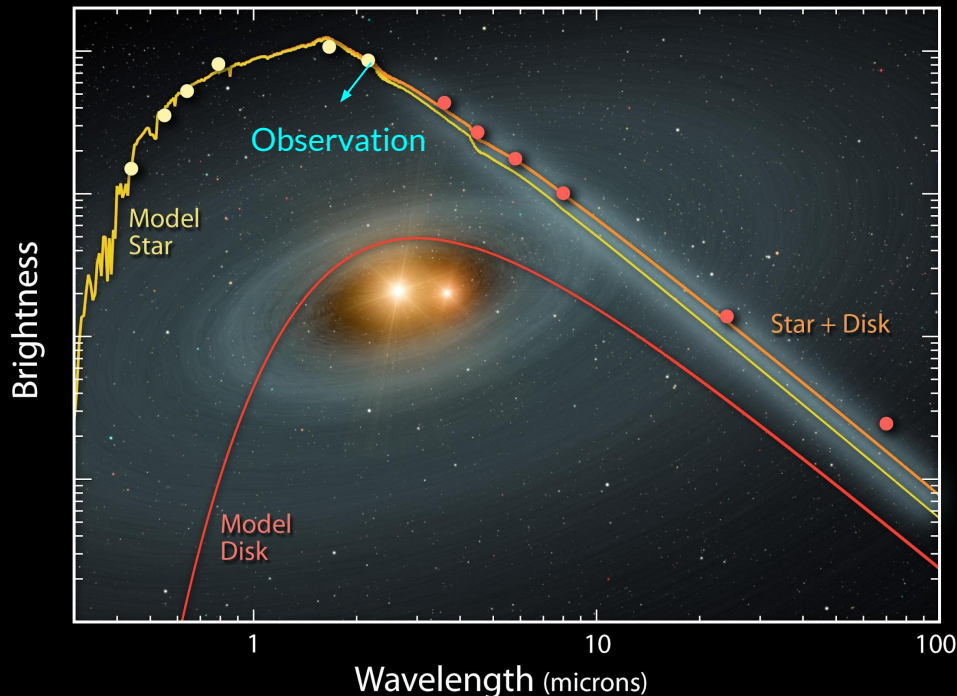
UNIVERSITY OF TORONTO

DATA SCIENCES INSTITUTE

Why do we want to know about the stars?

Alien life search:

- Exploring the Galaxy for extraterrestrial existence.
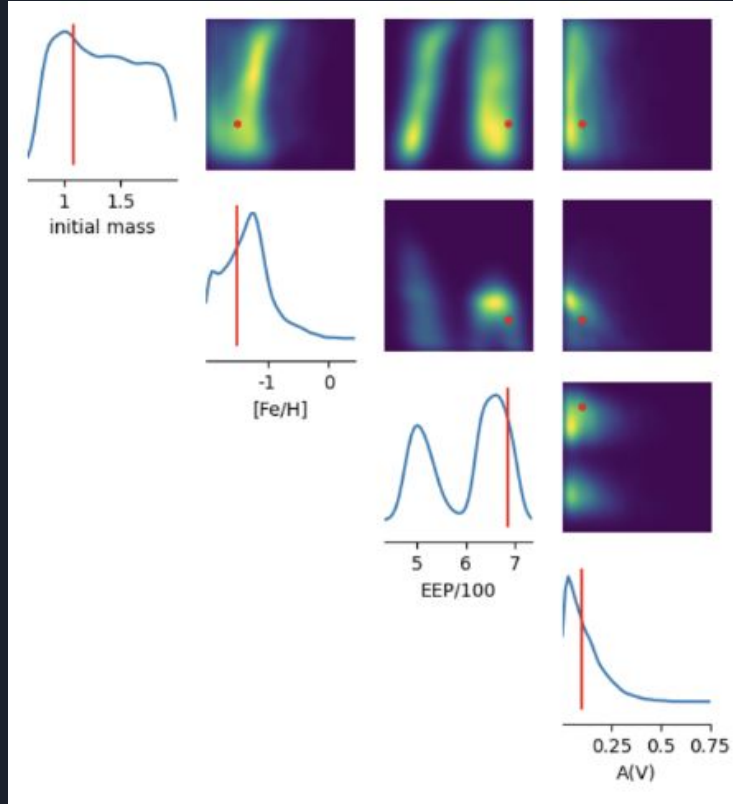
- ❖ Find a Sun-like Star

What are the observations?

Close Binaries With Infrared Excess   **Spitzer Space Telescope ● IRAC ● MIPS**

NASA / JPL-Caltech / M. Matranga (Harvard-Smithsonian CfA)   ssc2010-07a

But what do we need to know about the stars?

- Age, Mass, Distance and Chemical composition etc.

Inferring Stellar Properties

Challenges:

- Large uncertainties
- Complex degeneracies
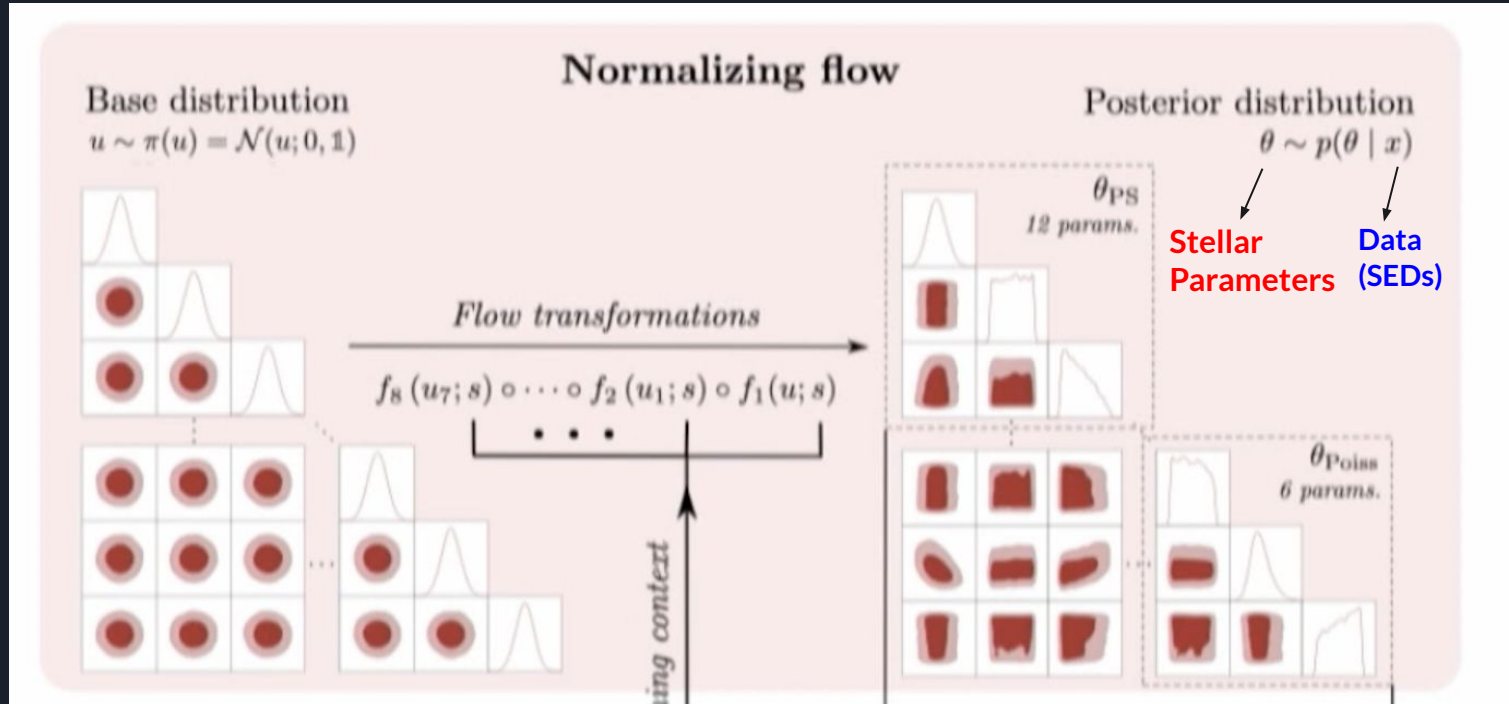- Multi-modal distributions

Traditional Methods for inference:

- Markov Chain Monte Carlo
- Nested Sampling

Slow Speed with big dataset -> 0.5 hour per star,
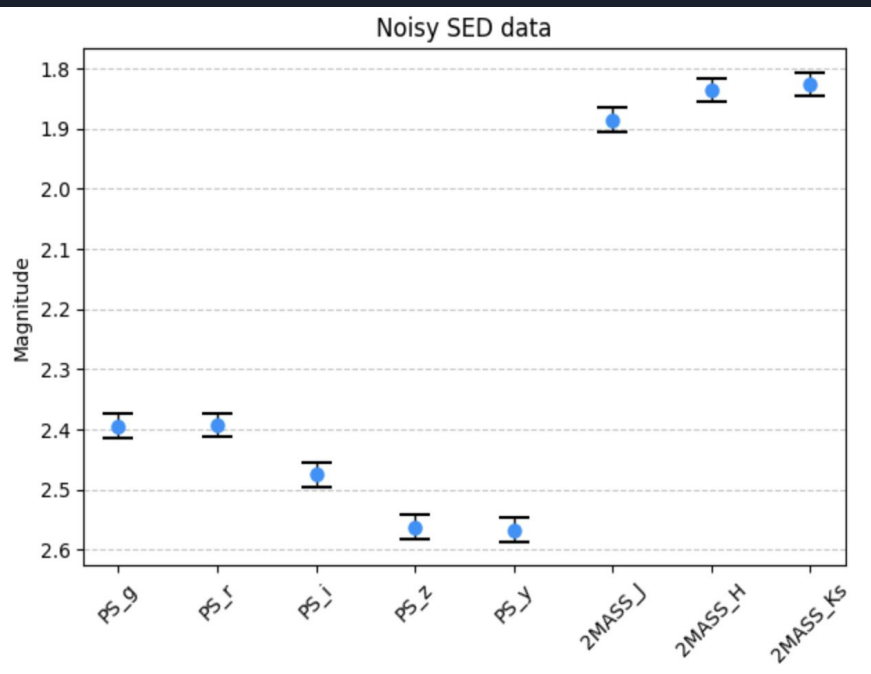**More than 60000 Years for 1 Billion stars!**

Solution: Simulation-based Inference (Normalizing Flow)
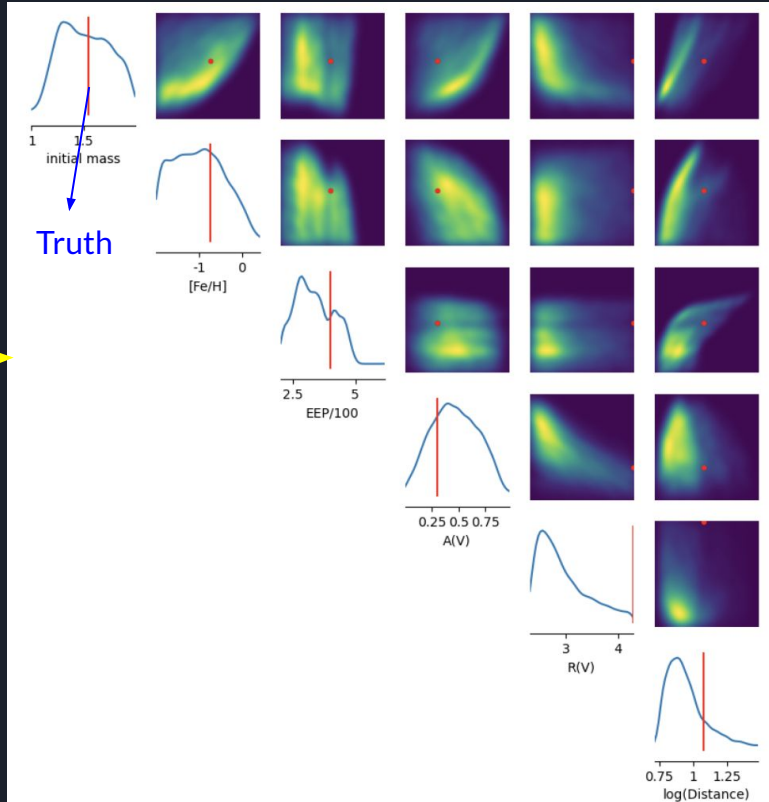
Make Inference much faster
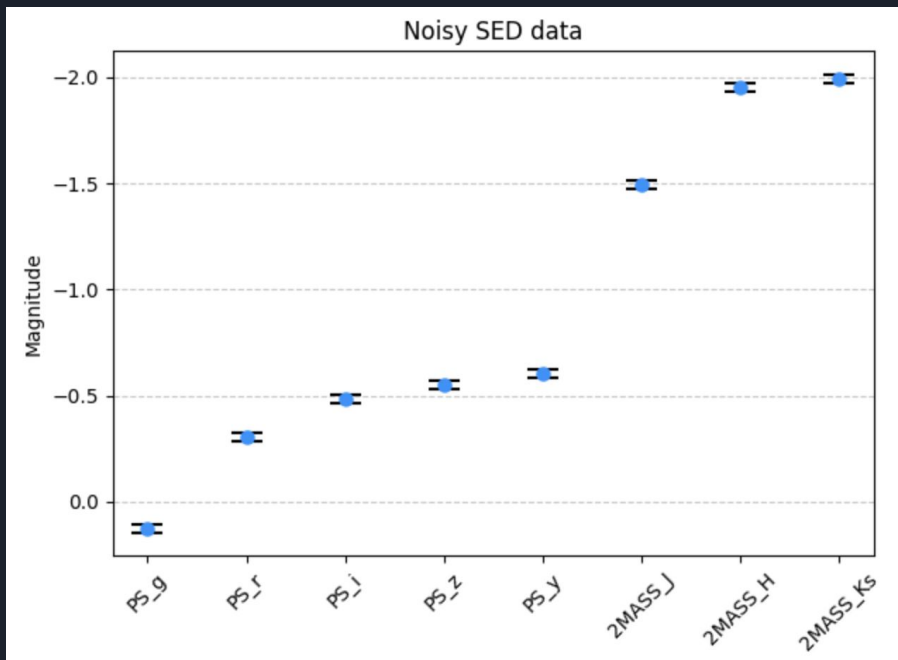
(Only seconds per star)

Result:

Estimated posteriors of the stellar parameters
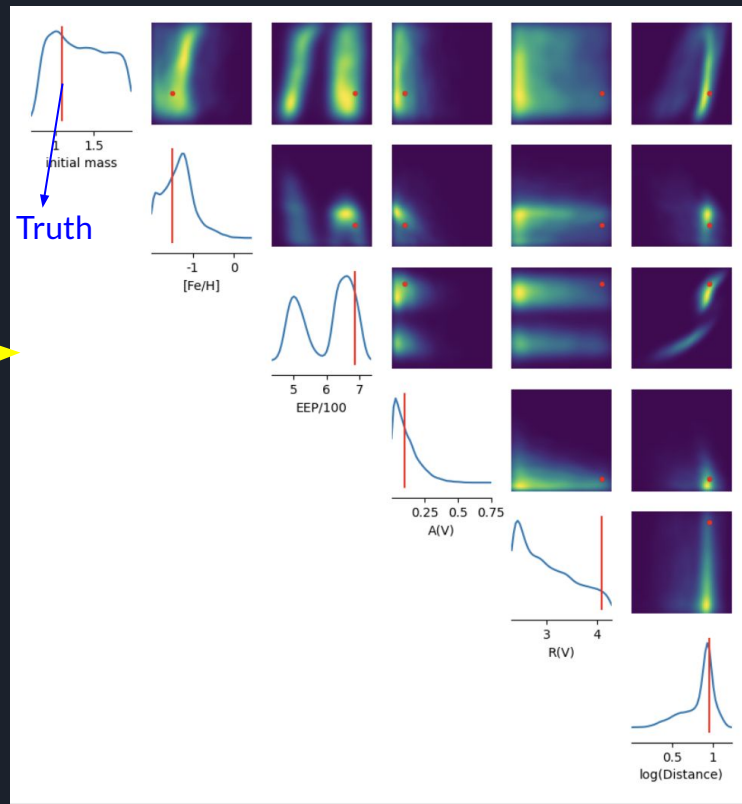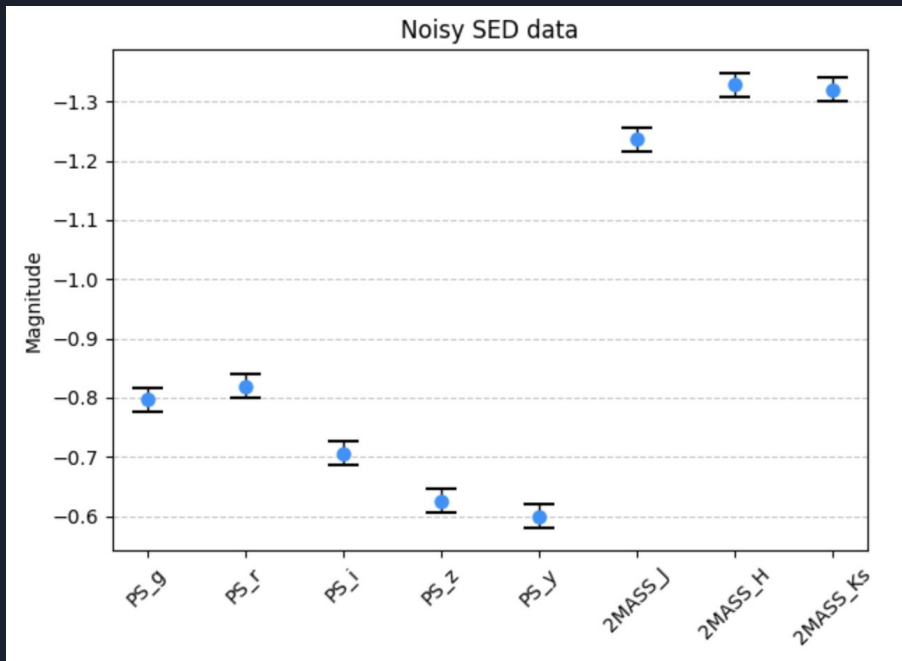
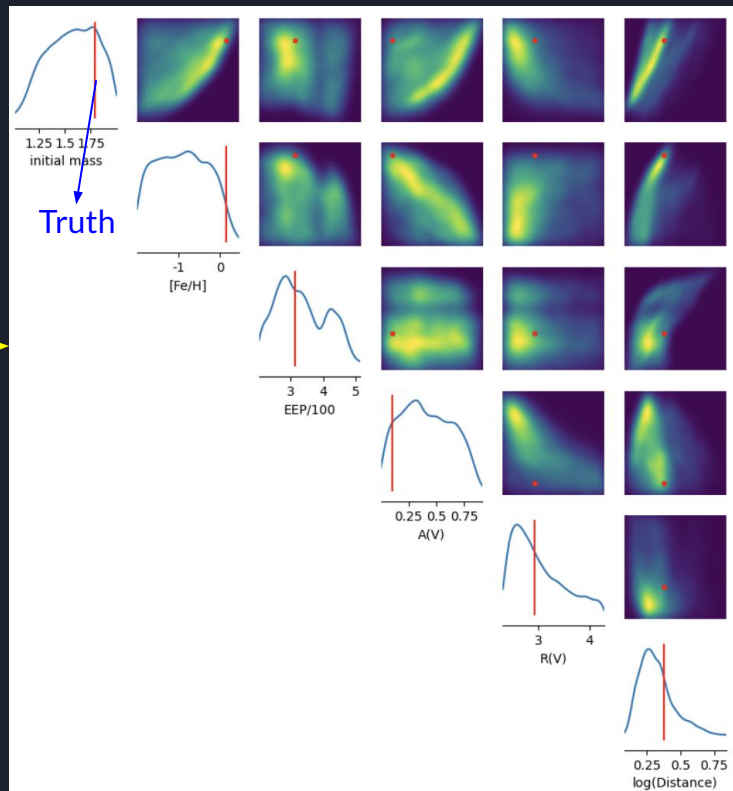SBI

Result:

Estimated posteriors of the stellar parameters

Result:

Estimated posteriors of the stellar parameters

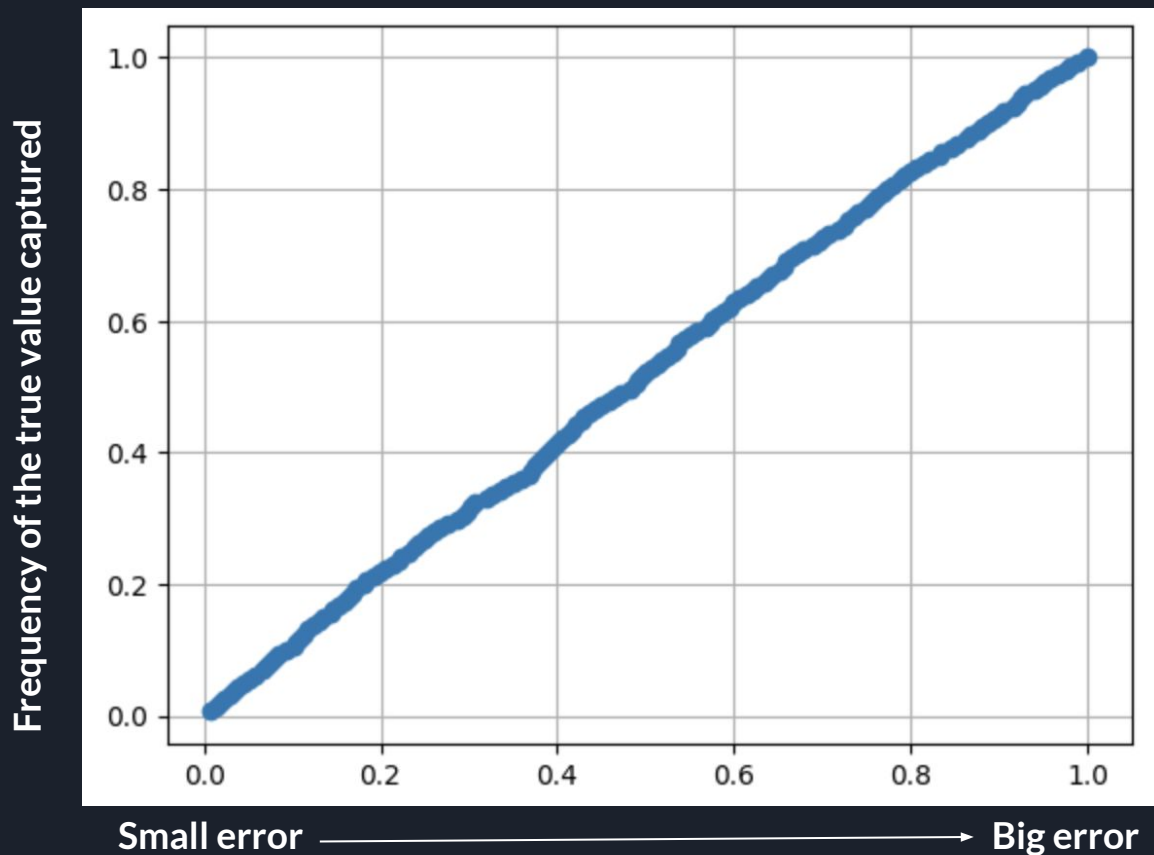# Validate the model: the performance of SBI over all stars

Frequency of the true value captured (y-axis)

Small error ⟶ **Big error**

- How often the posteriors capture the true value within the error

- Expect a straight Line

Lemos, P., Coogan, A., Hezaveh, Y., & Perreault-Levasseur, L. (2023). Sampling-Based Accuracy Testing of Posterior Estimators for General Inference

Conclusion

★ Simulation-based Inference can greatly increase the inference speed and make accurate estimation of the stellar parameters.

Next Steps

- Improve the SBI model to deal with more complicated data

- Apply to real observational data (>1 Billion stars from Gaia)

Gaia

Image credit: ESA / Gaia / DPAC