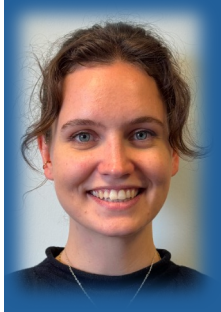


Cross-simulator simulation-based inference of astrophysics during the epoch of reionization

Carina Norregaard[1]
c.norregaard22@ic.ac.uk
Jonathan Pritchard[1,2], Romain Meriot[1]

[1]Imperial College London
[2]MPI for Radio Astronomy

CONTEXT

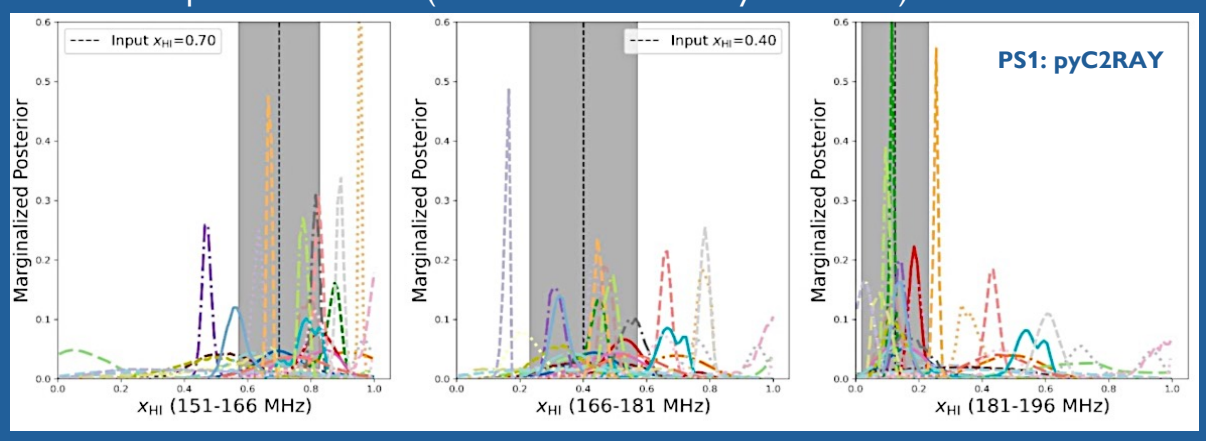
SKA Science Data Challenge 3b (SDC3b): Ionization history inference [1]

Goal: infer x_{HII} at 3 redshifts given realistic mock 2DPS, generated with unknown simulation codes

- 25+ international teams
- Challenges: cosmic variance, residual foregrounds, instrumental effects
- Emulators/simulations, MCMC/SBI

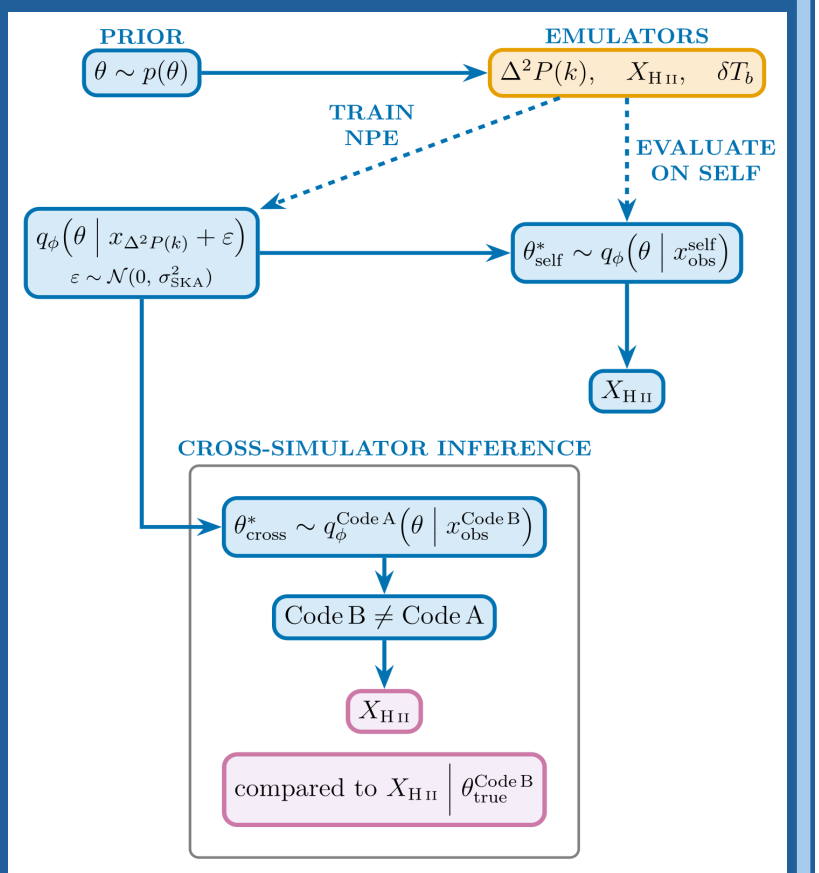
Results:

- Large discrepancy between posteriors in PS1 (pyC2Ray [2] a less commonly used code) vs between posteriors in PS2 (21cmFAST a commonly used code)



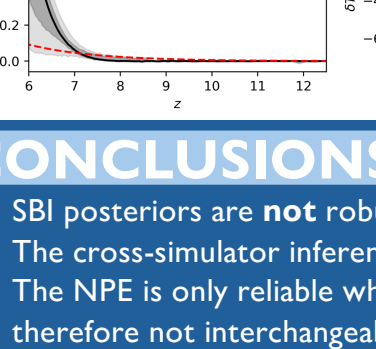
How robust is simulation-based inference of reionization observables to mismatches in the underlying forward model?

METHOD

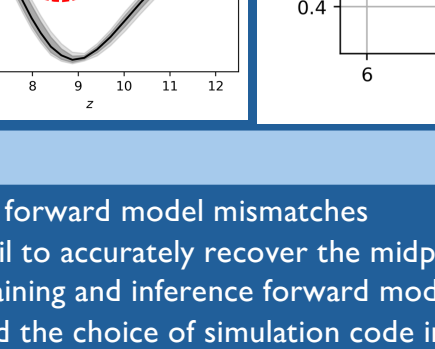


SIMULATIONS

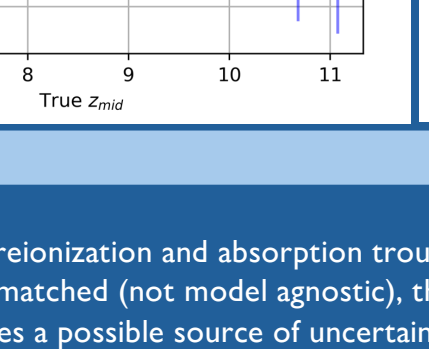
Image from [3]



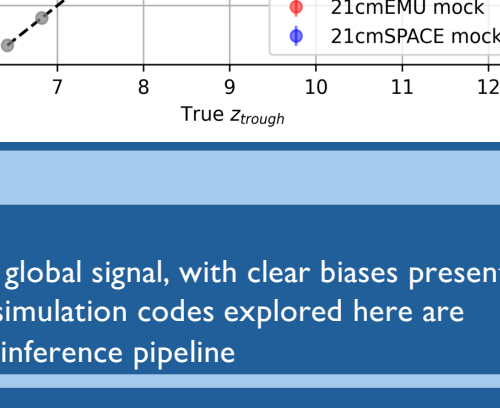
21cmSPACE: semi-numerical simulation code (f_{starII}, $f_{starIII}$, v_c, f_X, α, v_0, ζ, τ, f_{radio}, pop_{delay}) [4]



LoReLi II (generated with Licorice): hydro-radiative code (f_X, τ (gas to star), M_{min}, f_{esc}, X-ray ratio H/S) [5]



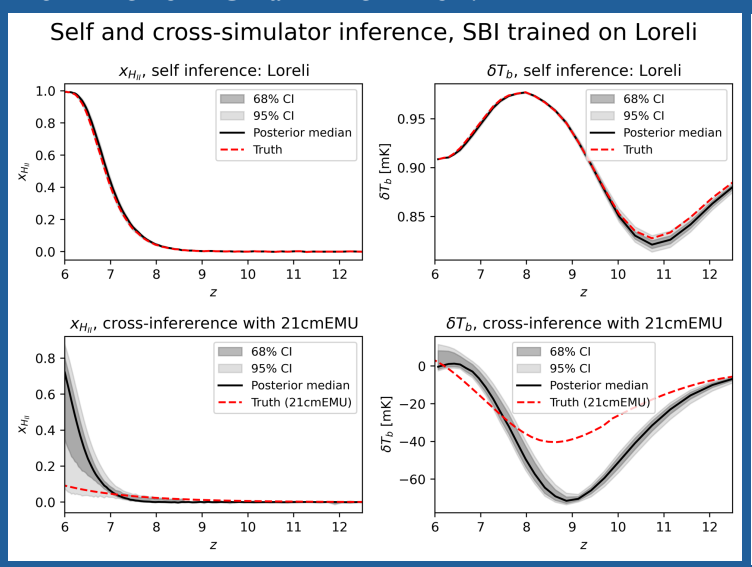
21cmFAST: semi-numerical simulation code (f_{star10}, $α_{star}$, f_{esc10}, $α_{esc}$, M_{turn}, t_{star}, L_X, $v_{xthresh}$, $x\text{-}ray_{spec\ index}$), (using 21cmEMU) [6, 7, 8]

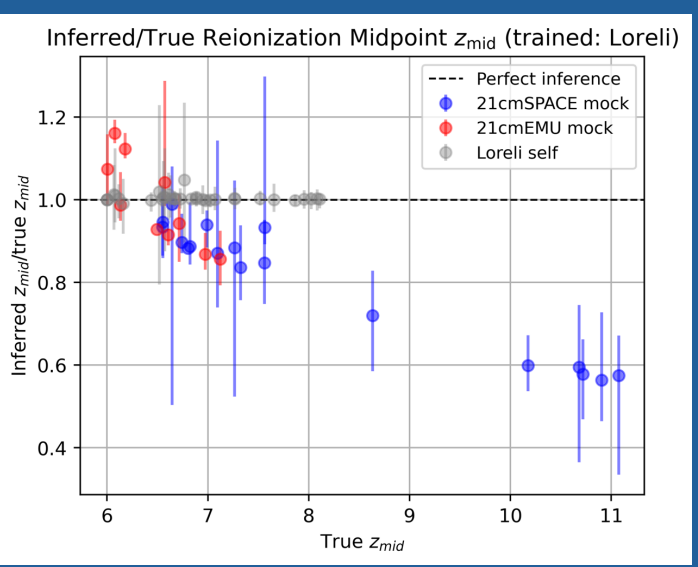


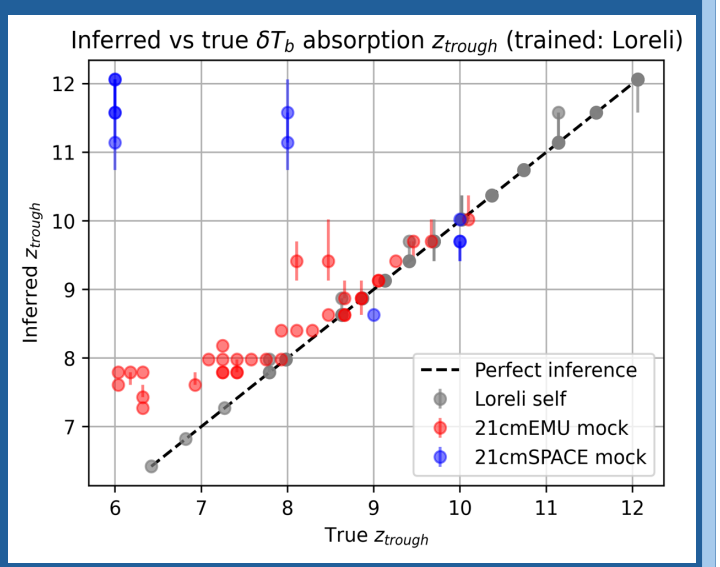
Example of a noisy mock observation used for inference, generated from the emulator trained on LoReLi II 1DPS data

RESULTS

Results from the pipeline trained on LoReLi II data, and effects of cross-simulator inference for the z_{min} of x_{HII} and z_{trough} of δT_b when using mock observations from 21cmSPACE and 21cmFAST:







CONCLUSIONS

- SBI posteriors are **not** robust to forward model mismatches
- The cross-simulator inferences fail to accurately recover the midpoint of reionization and absorption trough of the global signal, with clear biases present
- The NPE is only reliable when training and inference forward models are matched (not model agnostic), the 21cm simulation codes explored here are therefore not interchangeable and the choice of simulation code introduces a possible source of uncertainty in any inference pipeline

References:

[1] SKA SDC3b, 2025, https://sdc3.skao.int/challenges/inference/results
[2] Hirling, P., et al., 2024, A & C, 48, 100861
[3] Fialkov, A., et al., 2014, MNRAS, 445, 213–224
[4]Pochinda, S., et al., 2024, MNRAS, 531, 1
[5] Meriot R., Semelin B., Cornu D., 2025, A&A, 698, A80
[6] Mesinger A., et al., 2011, MNRAS, 411, 2, 955-972
[7] Murray, S. G., et al., 2020, JOSS, 5(54), 2582
[8] Breitman, D., et al., 2023, MNRAS, 527, 9833-9852