Machine Learning in Cosmological Simulations to Efficiently Model the Epoch of Reionization

Gaurav Pundir

What is the Epoch of Reionization (EoR)?

- First stars and galaxies ionized neutral hydrogen
- Universe 150 million years old to 1.2 billion years old
- Ionization ‘bubbles’ grew and merged

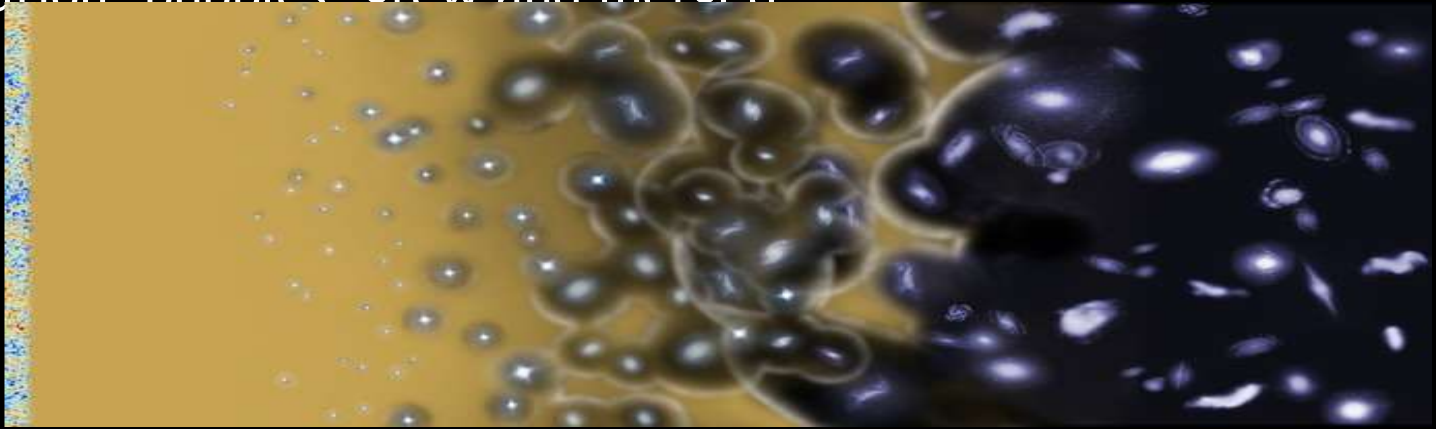


Image credits: US Planck Data Center[1]

What is the Epoch of Reionization (EoR)?

- Observational signatures of this era
  - *Lyman-alpha lines*
  - *21 cm signal*
  - *CMB secondary anisotropies*

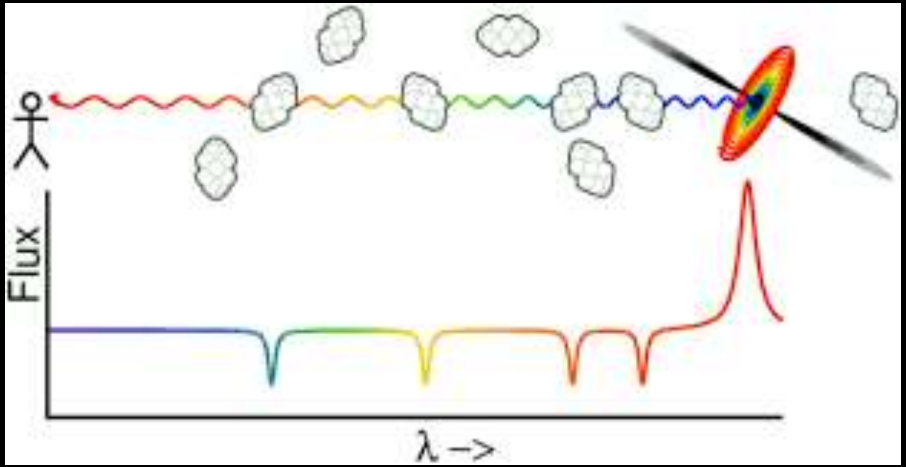


Image credits - https://astro.ucla.edu/~wright/Lyman-alpha-forest.html

- We have models for the EoR

- Need fast and accurate predictions for these observables!

Modeling the EoR

- First galaxies form in massive enough ‘dark matter halos’

→ 1st ingredient: Spatial distribution of dark matter halos

Image credits - CERN Courier

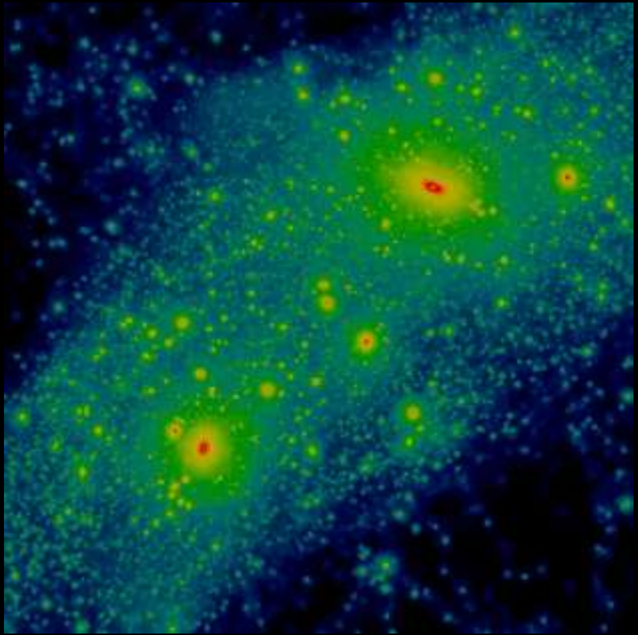
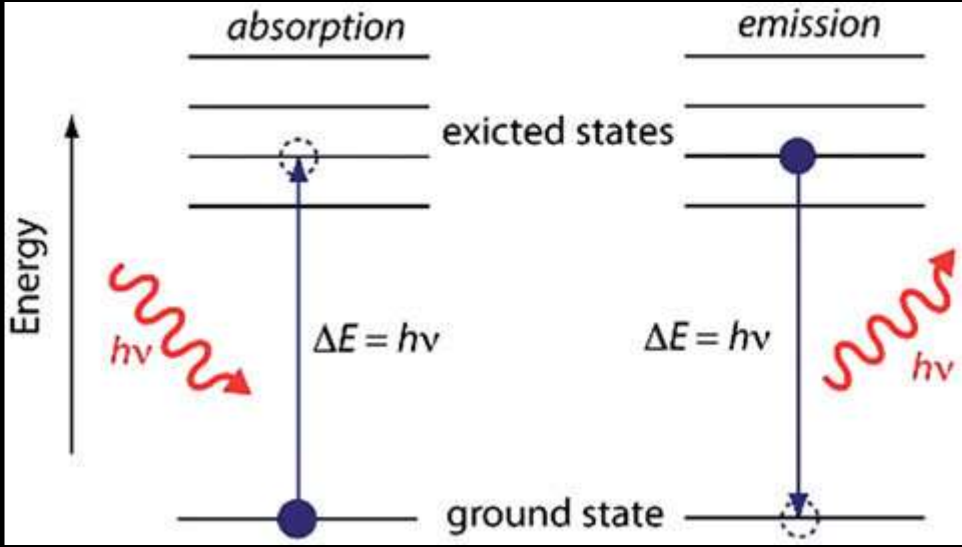


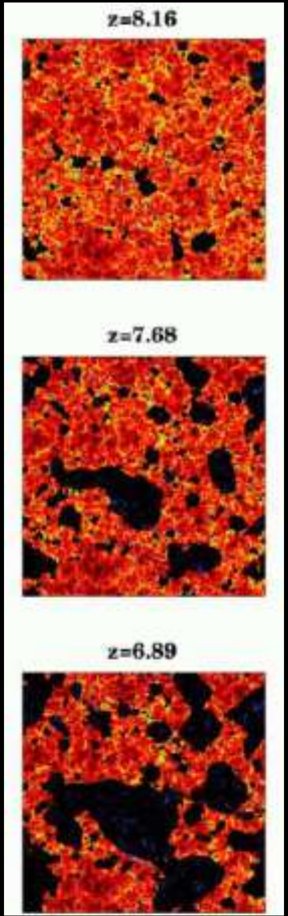
Image credits - Astronomical Society of the Pacific

Modeling the EoR

- Stars within galaxies produce high energy photons that can ionize hydrogen

➡ 2nd ingredient: Understanding interaction between photons and matter





What *exactly* are we modeling?

- The distribution of 'ionized bubbles'

- Most accurate and slowest way:
  N-body simulations + Radiative transfer simulations

- Least accurate and fastest way:
  Semi-numerical approximations + Semi-numerical approximations

- A Middle ground:
  N-body simulations + Semi-numerical approximations

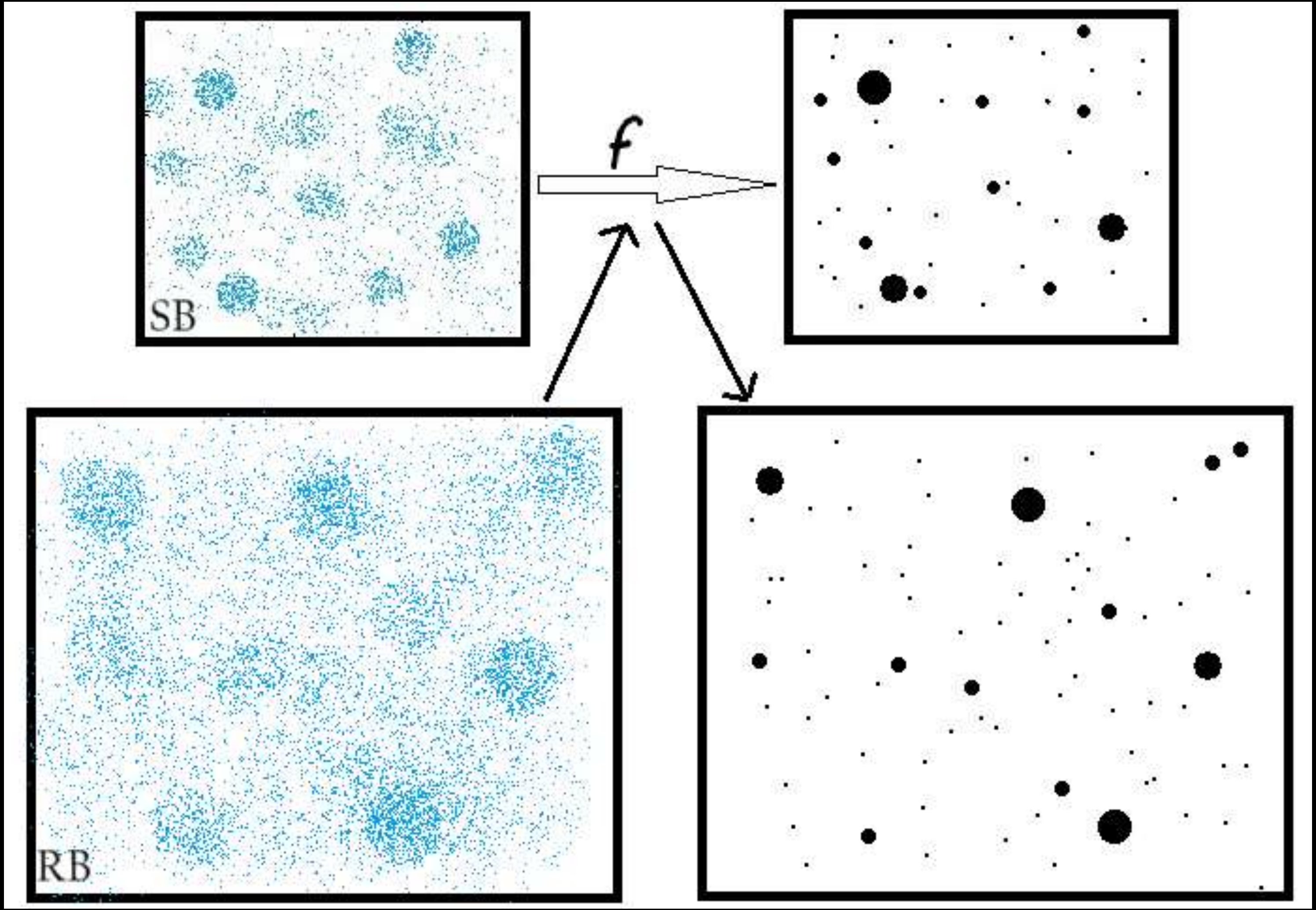
A Middle ground:

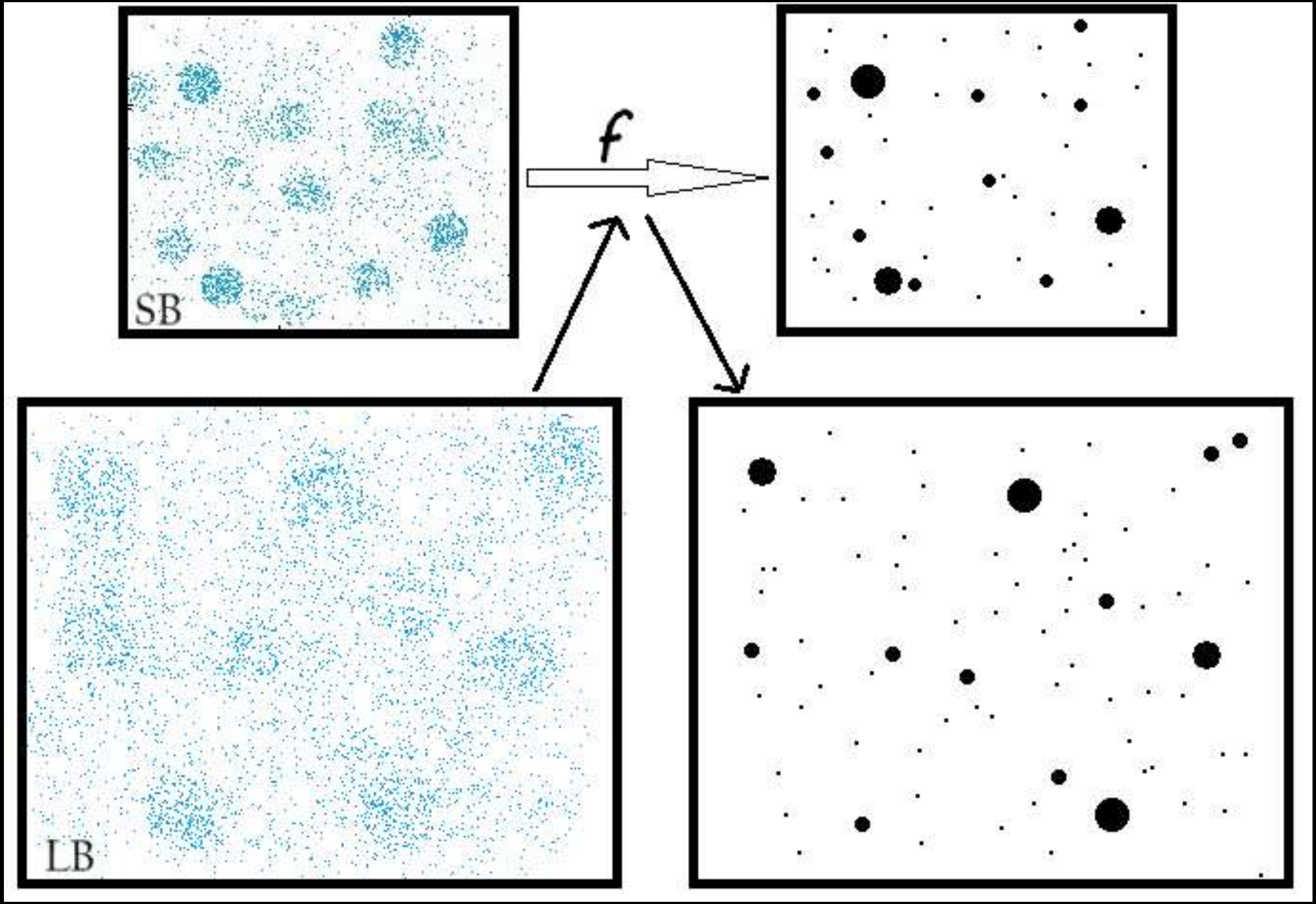
N-body simulations + Semi-numerical approximations

- N-body simulations are computationally expensive
  - High-resolution and large volume

- Bypass using a combination of simulations that are
  - High-resolution and low volume
  - Low-resolution and large volume

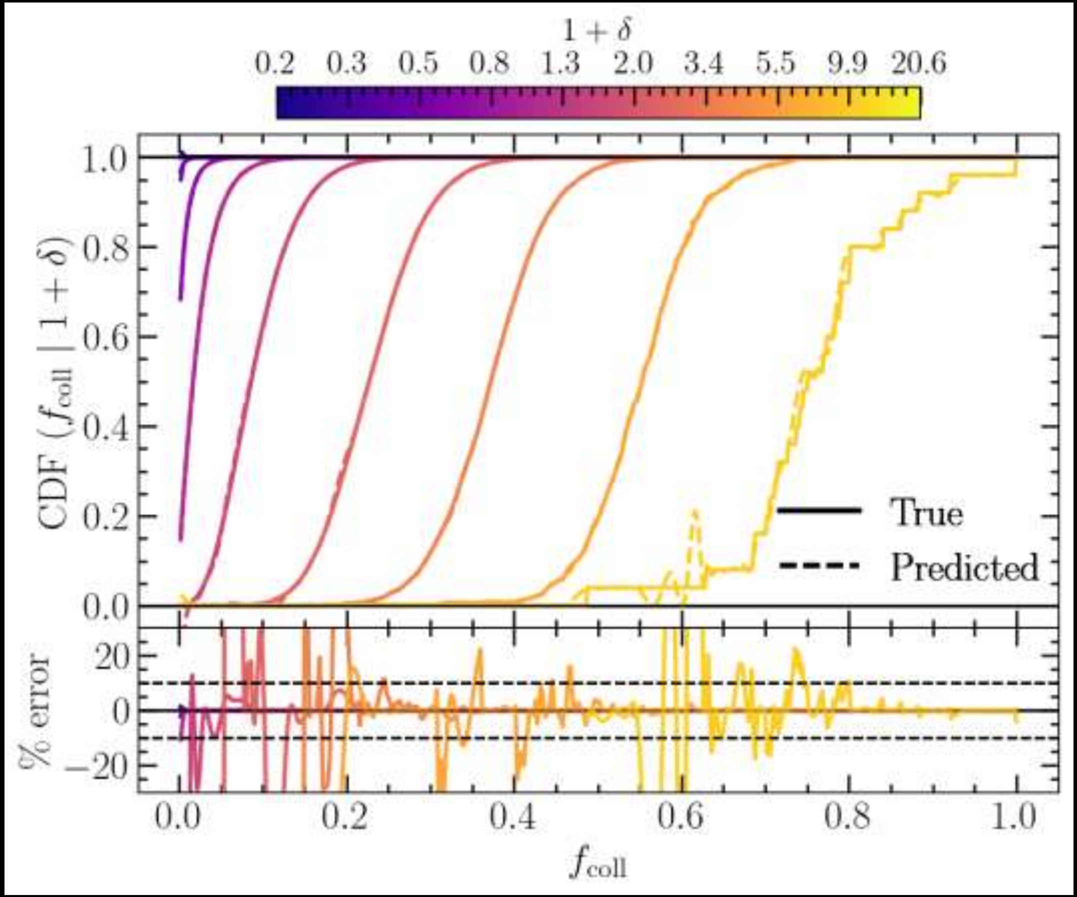
- But what is machine learning?


SB
f
RB


SB
f
LB

Technique

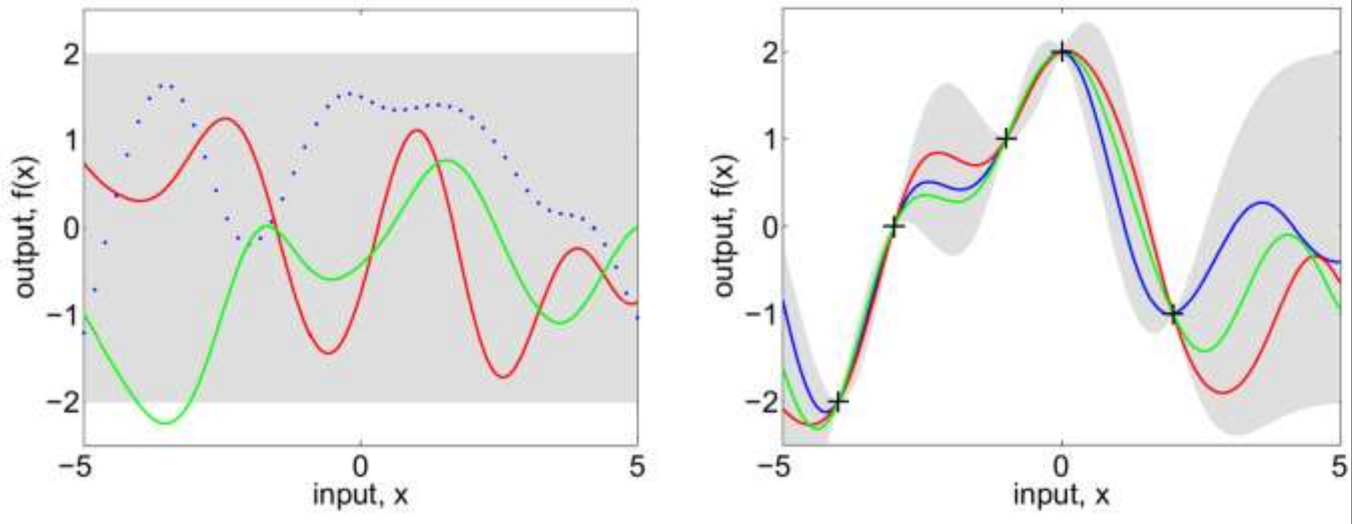
- **What to train the ML model on?**
- CDFs of f_{coll} conditioned on $1+\delta$
- Obtained from SB



Technique

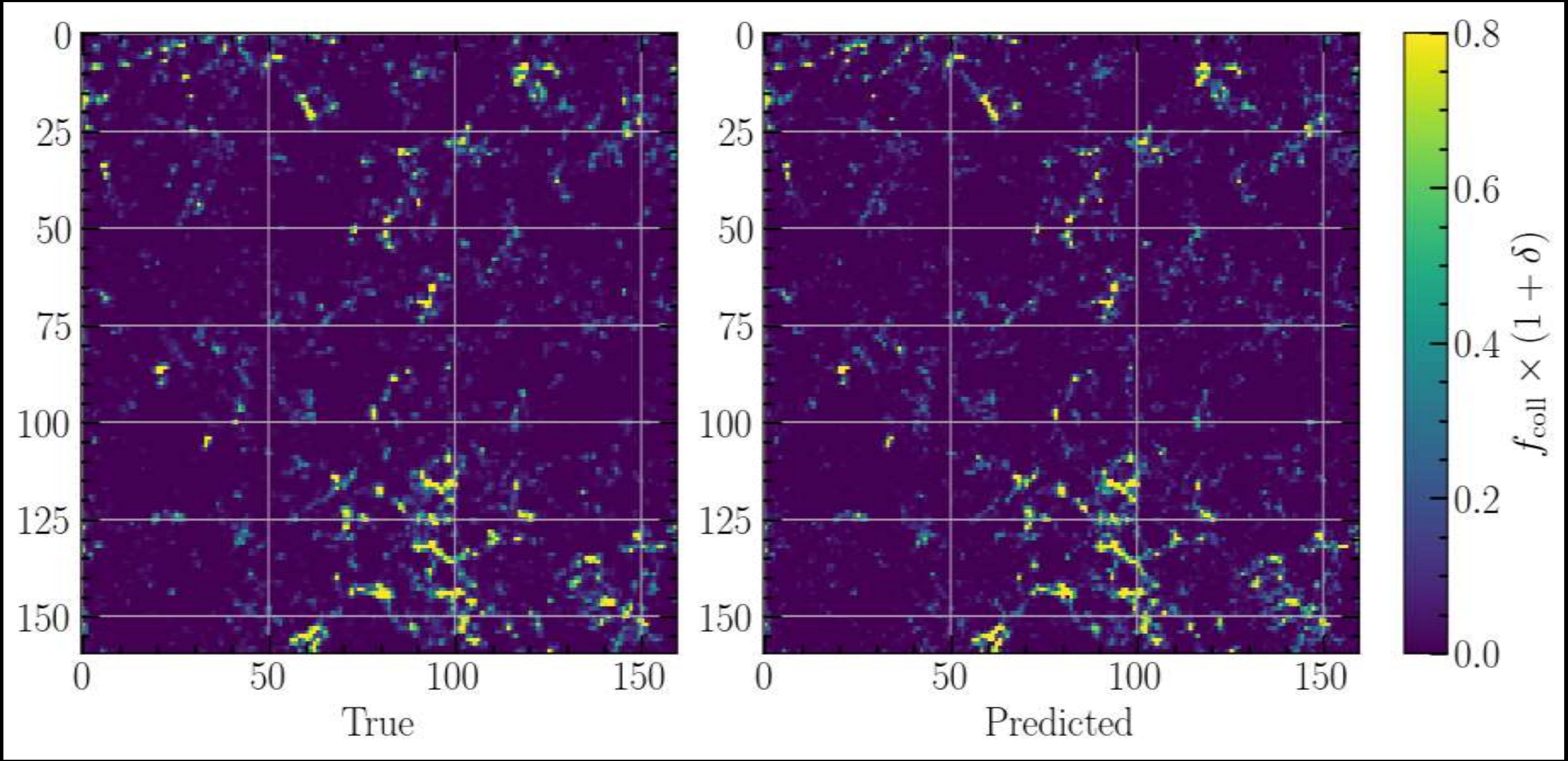
- **What is our ML model?**

- Gaussian Process Regression - “Trains” in ~10 minutes

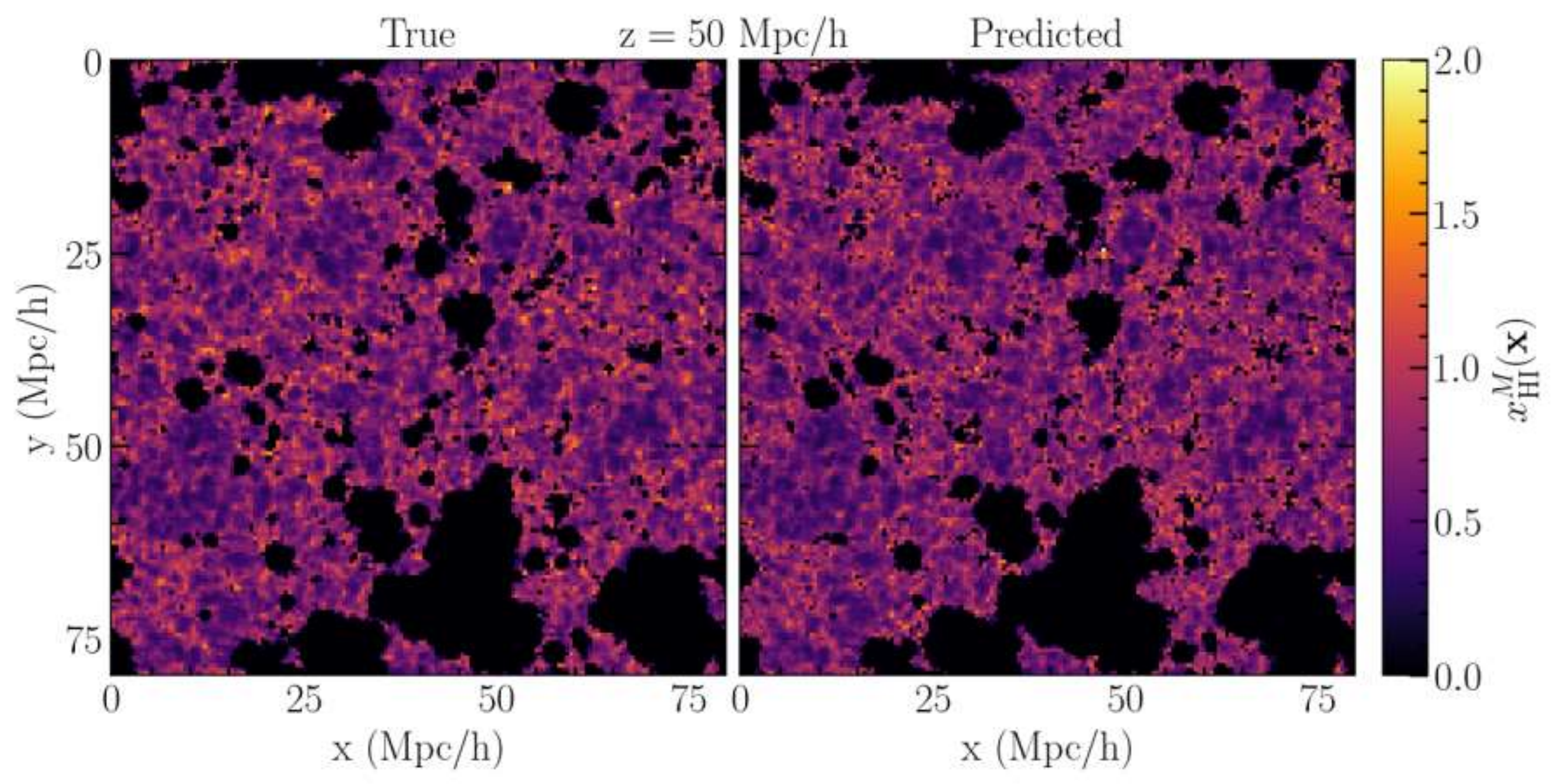


(a), prior (b), posterior

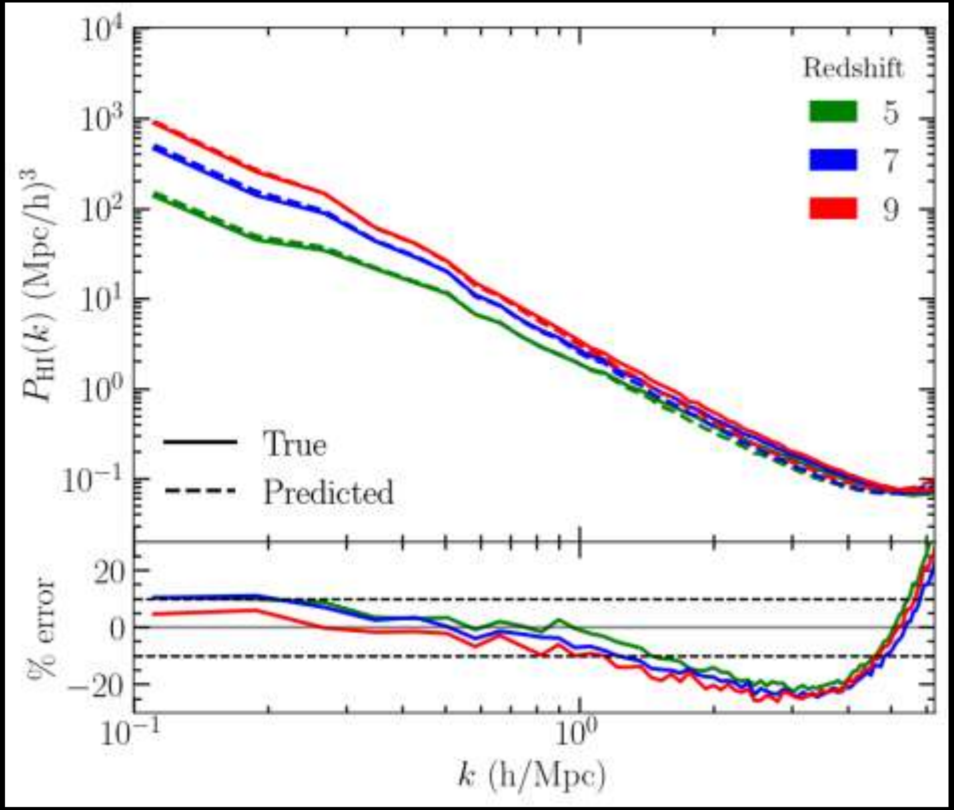
Results - Mass-averaged f_{coll} slice



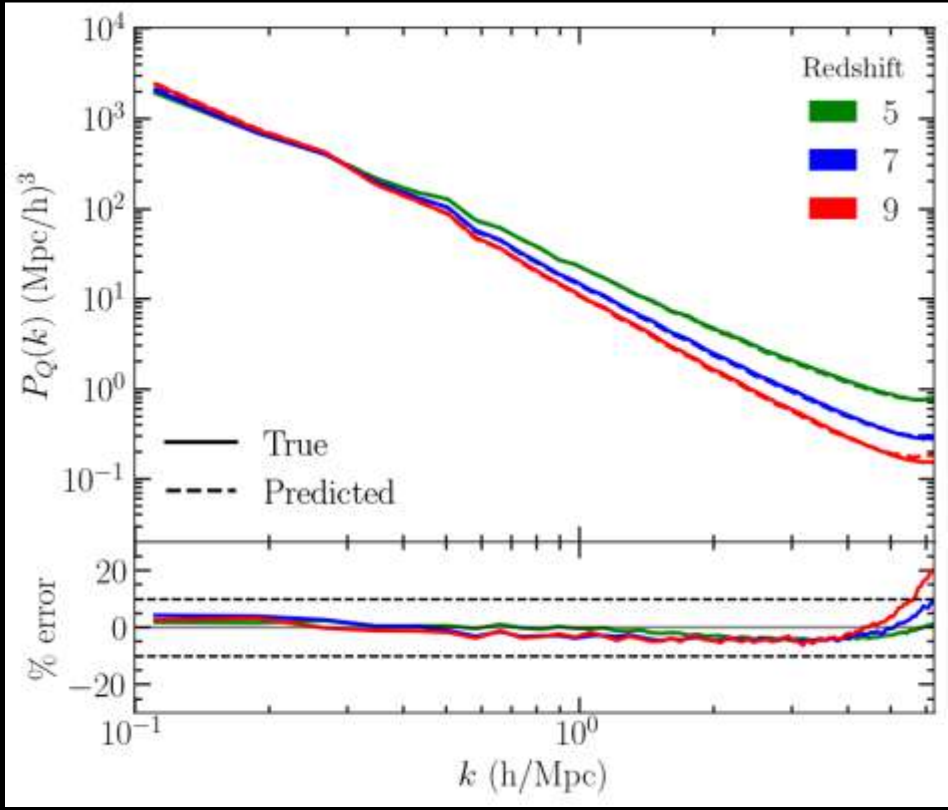
Results - HI map and salient features



Results -



Redshift
5
7
9
True
Predicted
$P_{HI}(k)$ (Mpc/h)^3
% error
k (h/Mpc)



Redshift
5
7
9
True
Predicted
$P_Q(k)$ (Mpc/h)^3
% error
k (h/Mpc)

Next Steps

- **Building a redshift history:**
  - Figure out an optimal Δz
  - Build emulator, interpolate for intermediate z
- **Pushing efficiency further:**
  - Reduce SB and LB size by 8
  - Repeat the whole process
- **Better environment variables:**
  - Use λ_1, λ_2, λ_3 instead of δ
  - Use a neural network


Thank you