Estimating Star Formation Histories with deep learning and cosmological simulations

M. Huertas-Company, F. Lanusse, A. Boucaud, H. Bretonnière

Hackathon. Ecole Astroinfo 2021

Science Case

TNG SIMULATION
(https://www.tng-project.org/)

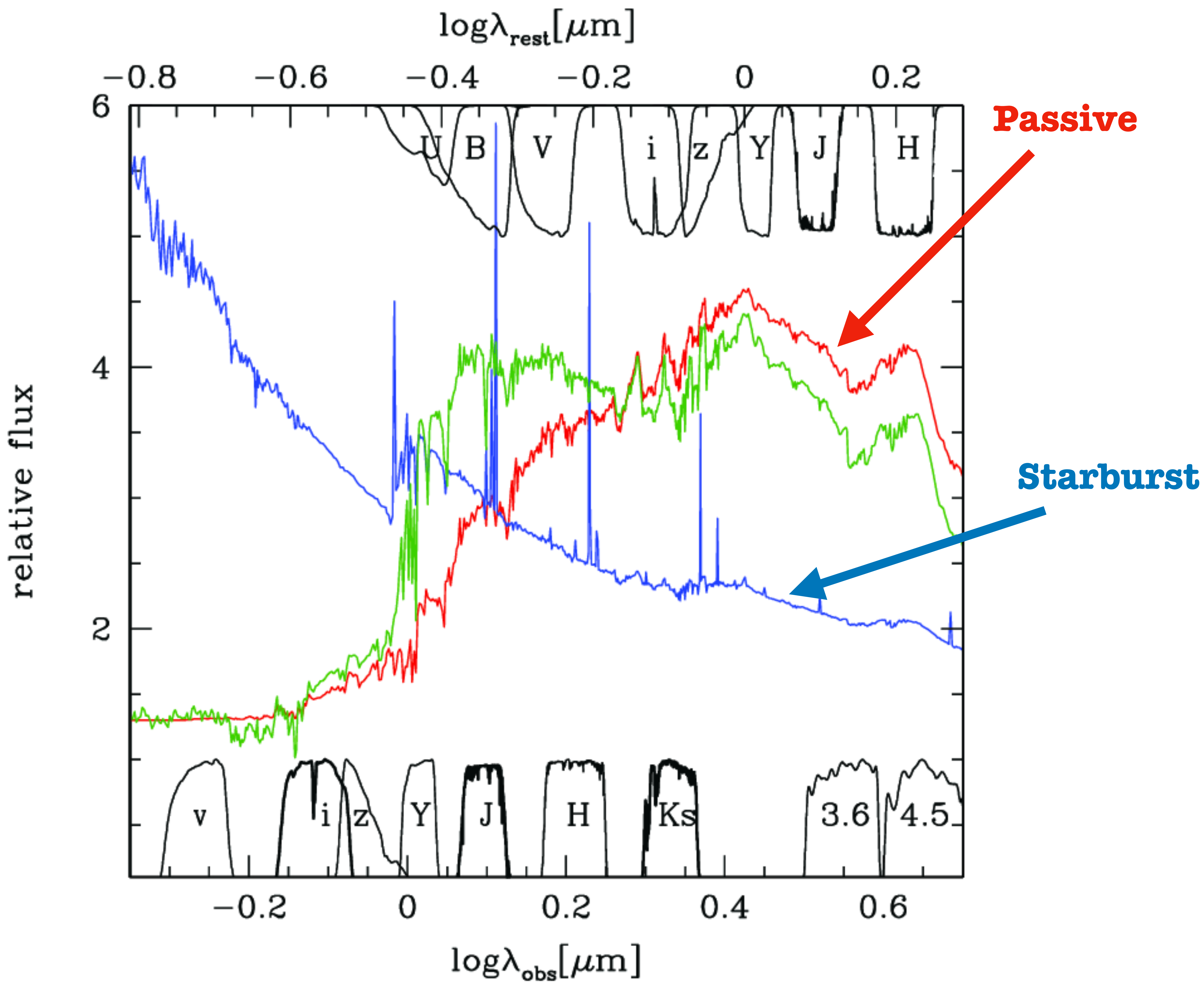
- Galaxies grow via in-situ star formation and stellar accretion (mergers)
- During its evolution, a galaxy can experience periods of more or less intense star formation. Some physical process can enhance star formation (e.g. merger, inflow…), others can suppress star formation (SN explosions, BH feedback…)
- We call Star Formation History (SFH), the Star Formation Rate (SFR) as a function of cosmic time - it summarises the growth history of a galaxy
- Estimating SFHs of large samples of galaxies, gives us clues on how galaxies assembled their mass.

How do we estimate SFHs?

- Problem: observations provide only a snapshot of a galaxy at a given time.

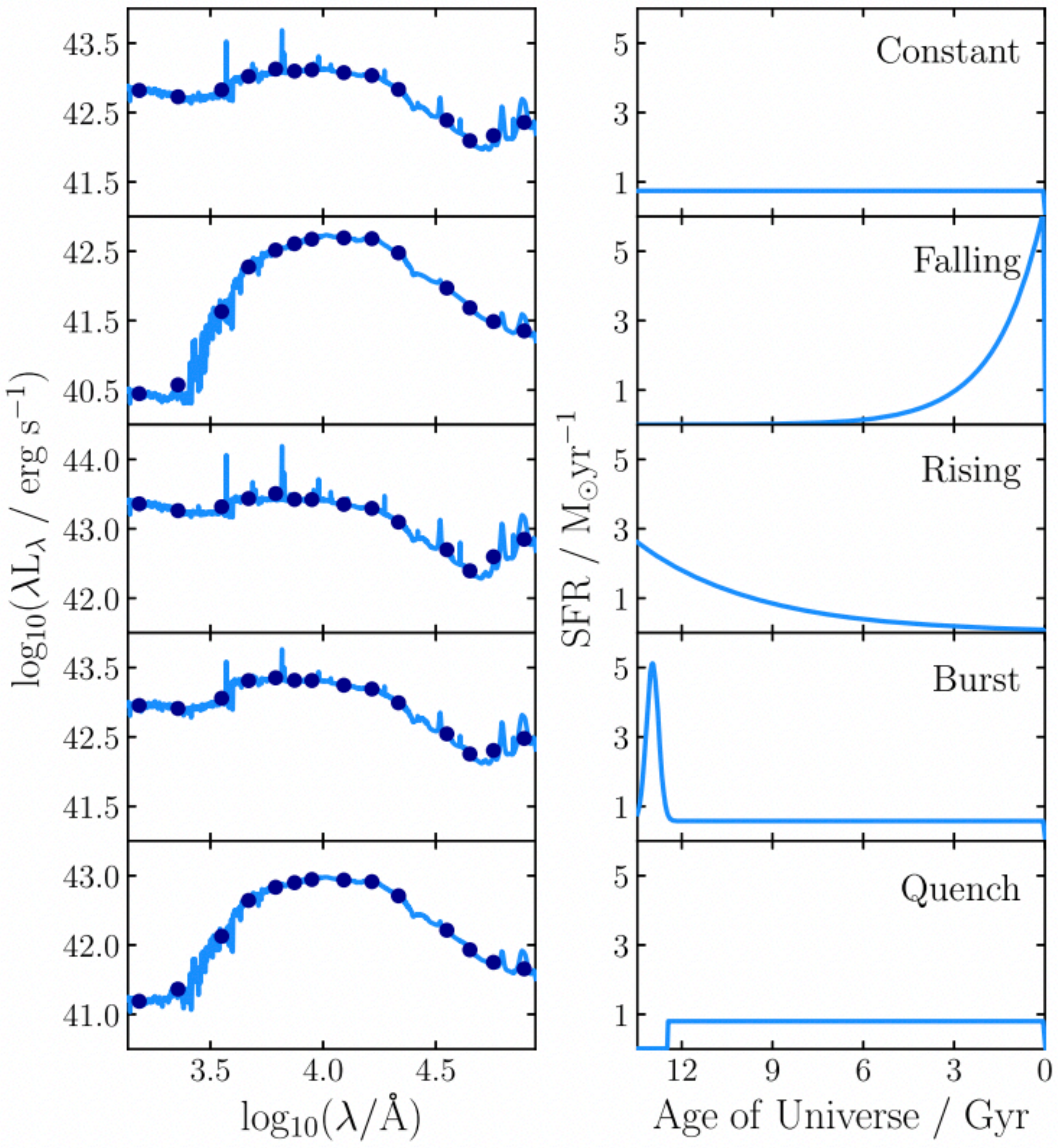
- However, the emission of stars in the galaxy - **Spectral Energy Distribution (SED)** - provides clues about the past star formation

- For example: Recent Star Formation would result in young stars with enhanced blue emission. A galaxy with no on-going start formation would have older, redder stars.



State-of-the-art approaches

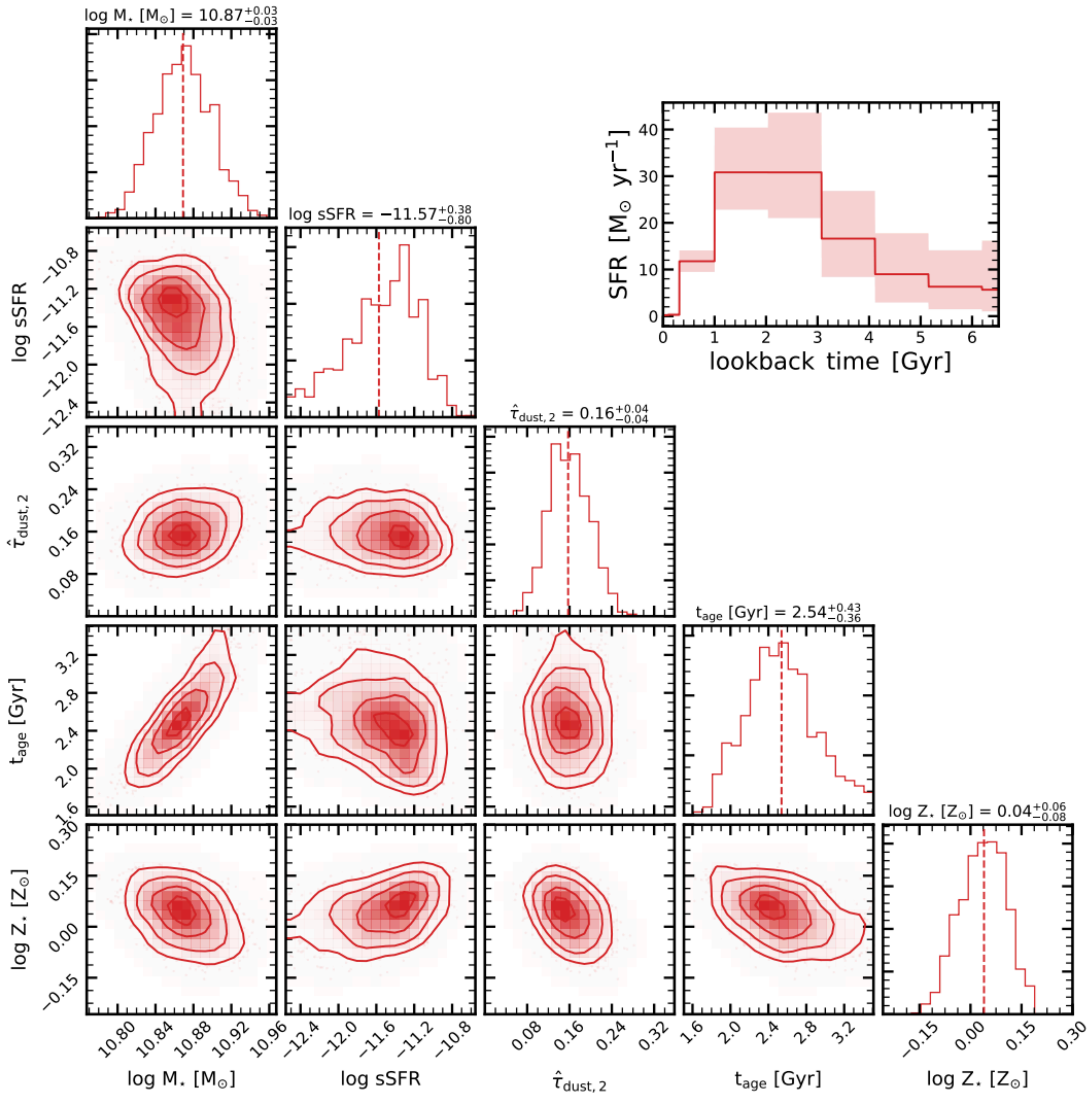
- Forward modelling with stellar population synthesis models
  - Assume a parametric SFH (tau models, delayed tau models)



Carnall+19

State-of-the-art approaches

- Forward modelling with stellar population synthesis models
  - Assume a parametric SFH (tau models, delayed tau models)
    - Problem: too simplistic SFHs, little can be said about specific events
  - More recently, non parametric SEDs
    - Depends on a number of assumptions, bins, also very slow



Tachhella+20

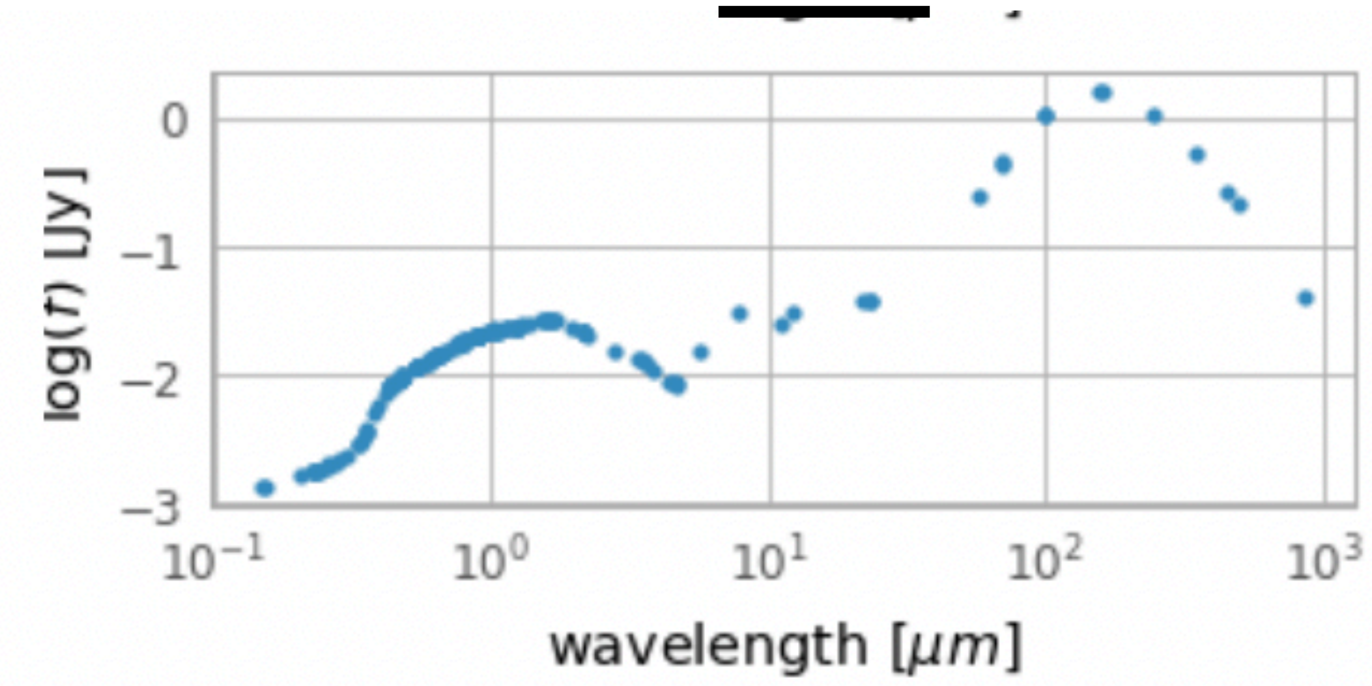
The inference problem

IDEA:

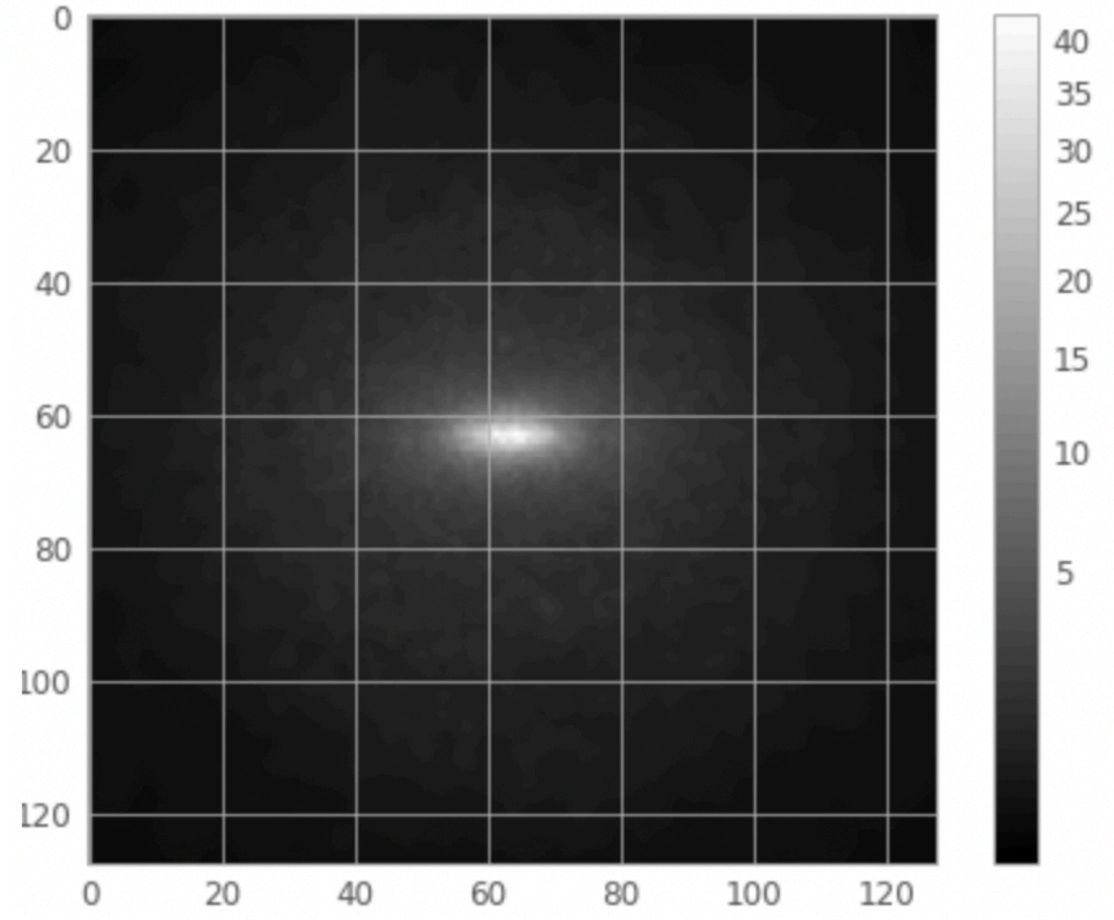
Use information from cosmological simulations to infer SFHs from the SEDs (+ images) using Neural Networks.

Input

SED

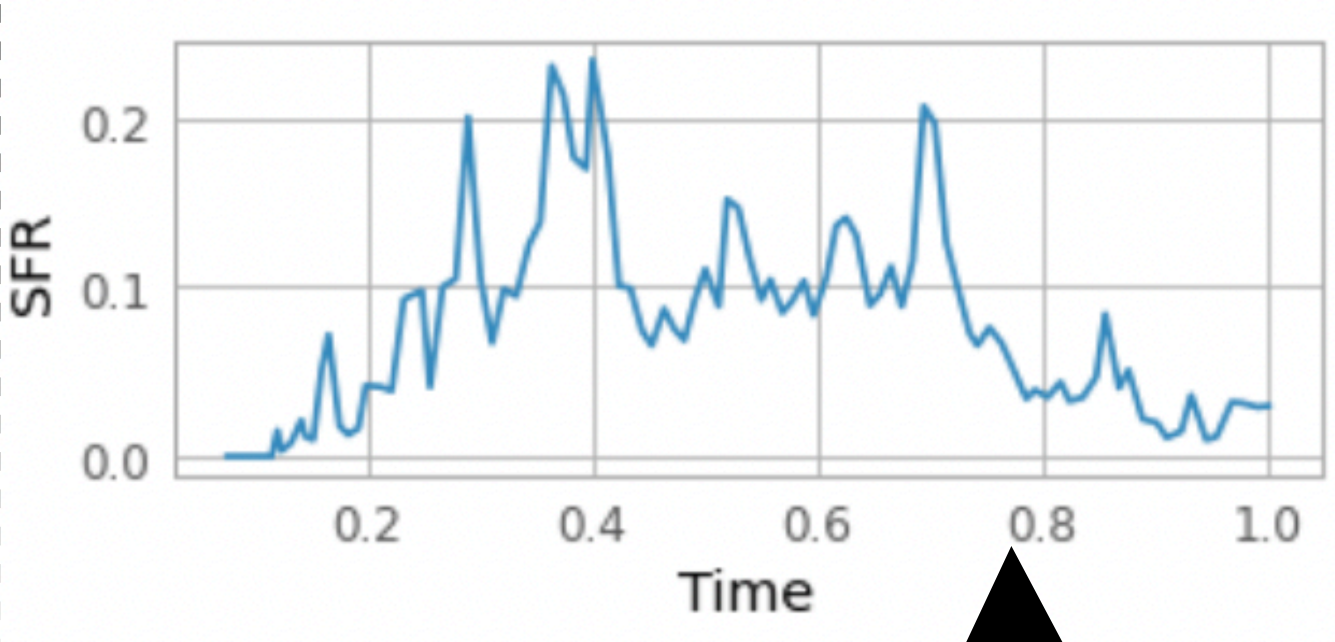


+

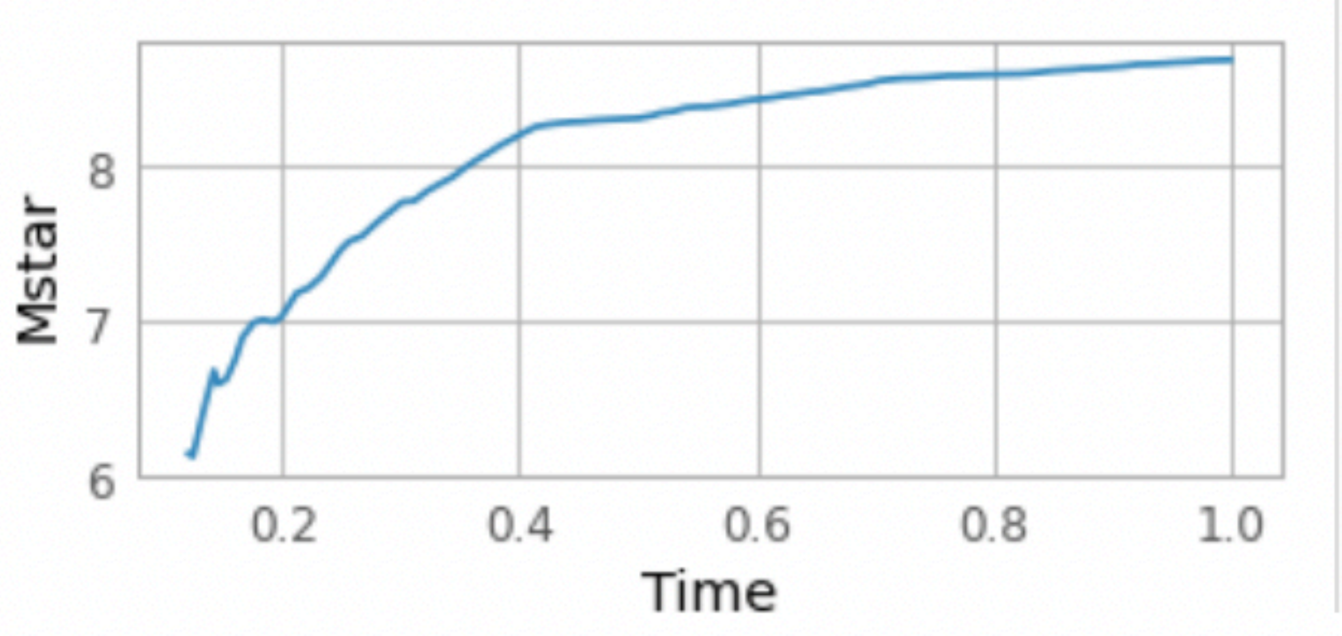


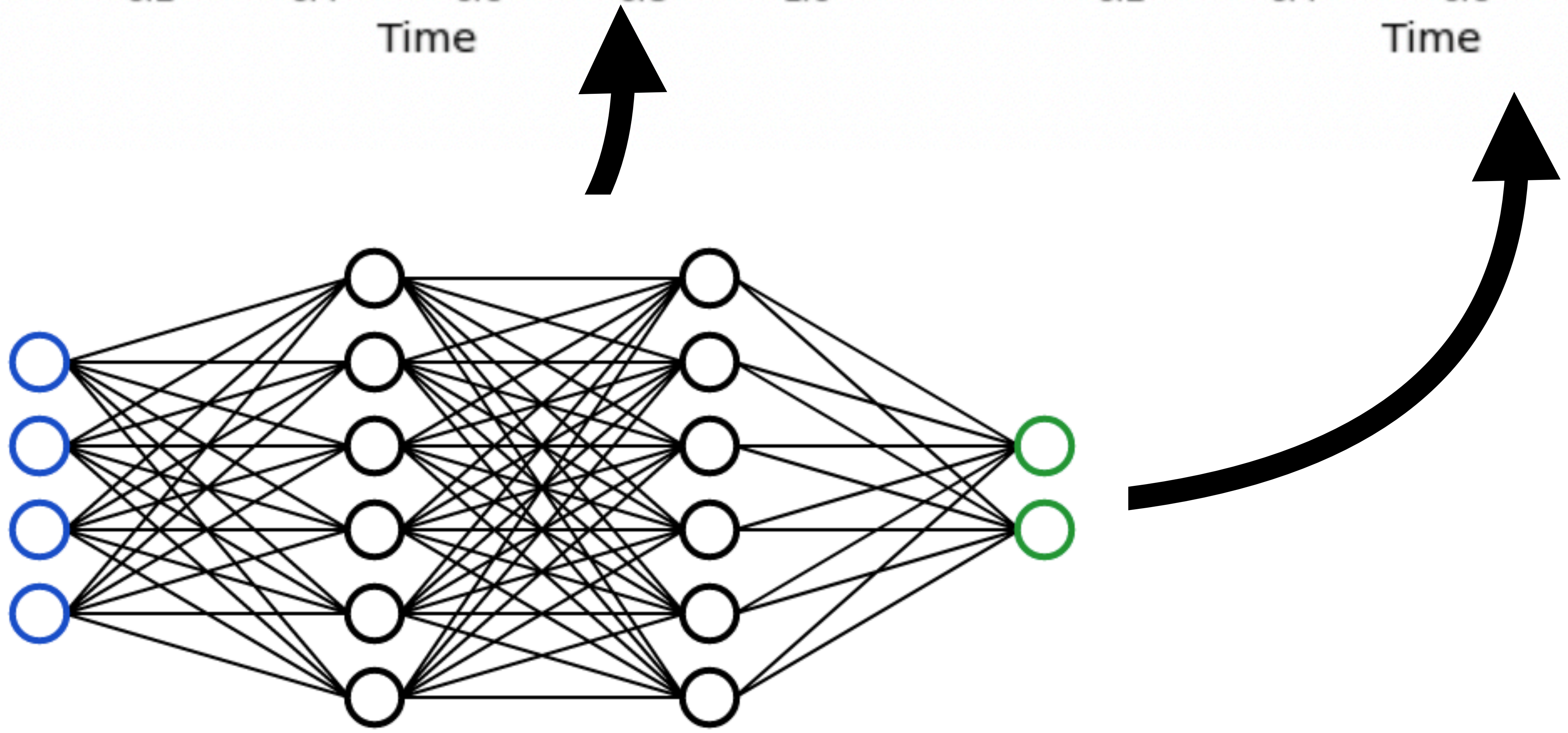
Output

SFH



Mass Growth





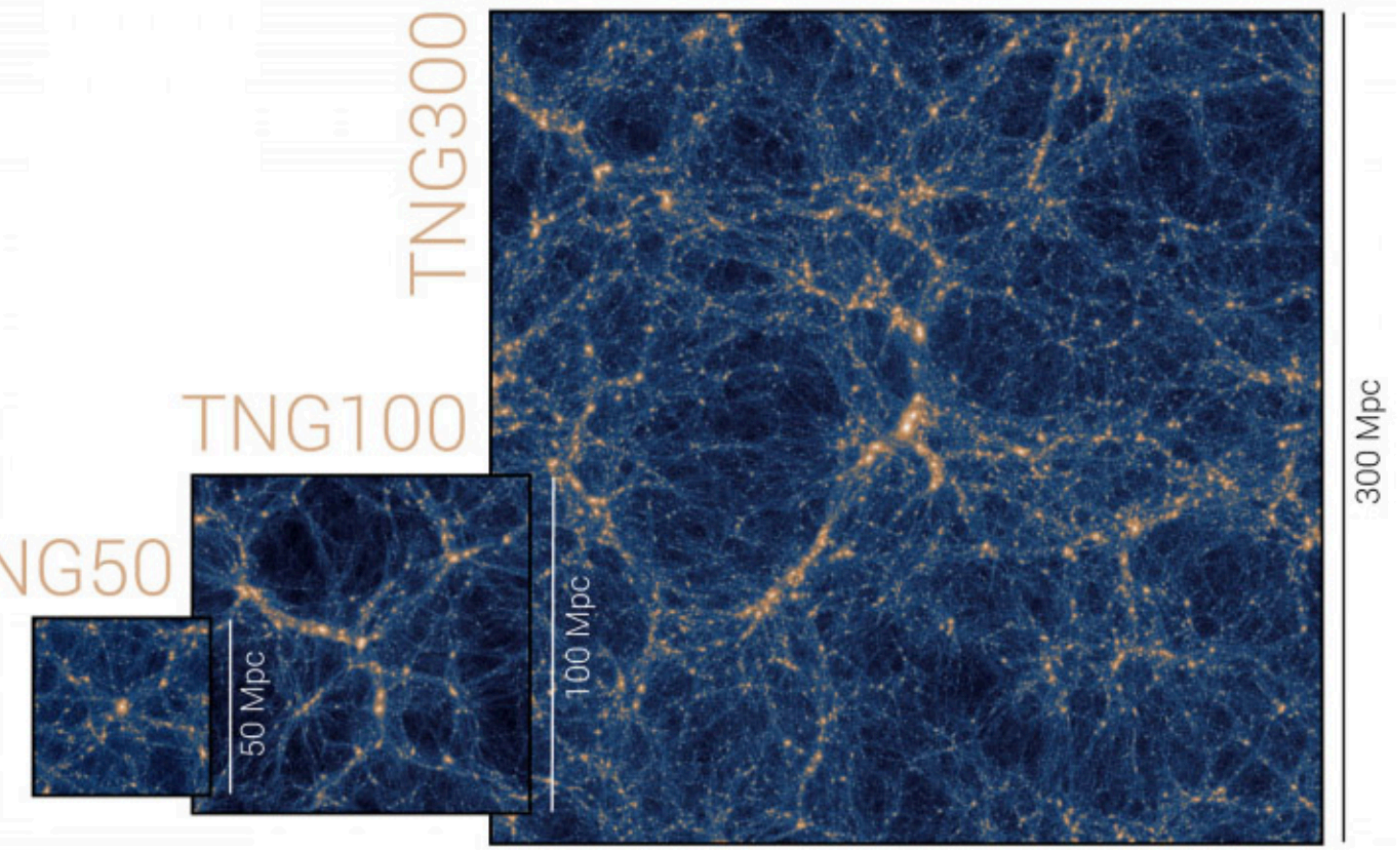
Why simulations ?

- We can follow individual objects back in time
- We get “realistic” SFHs since simulated galaxies match several properties of the observed ones (e.g.Iyer+20)

Why machine learning ?

- Speed. Future surveys will contain millions of objects.
- Complex, non-linear mapping between SED and SFH…

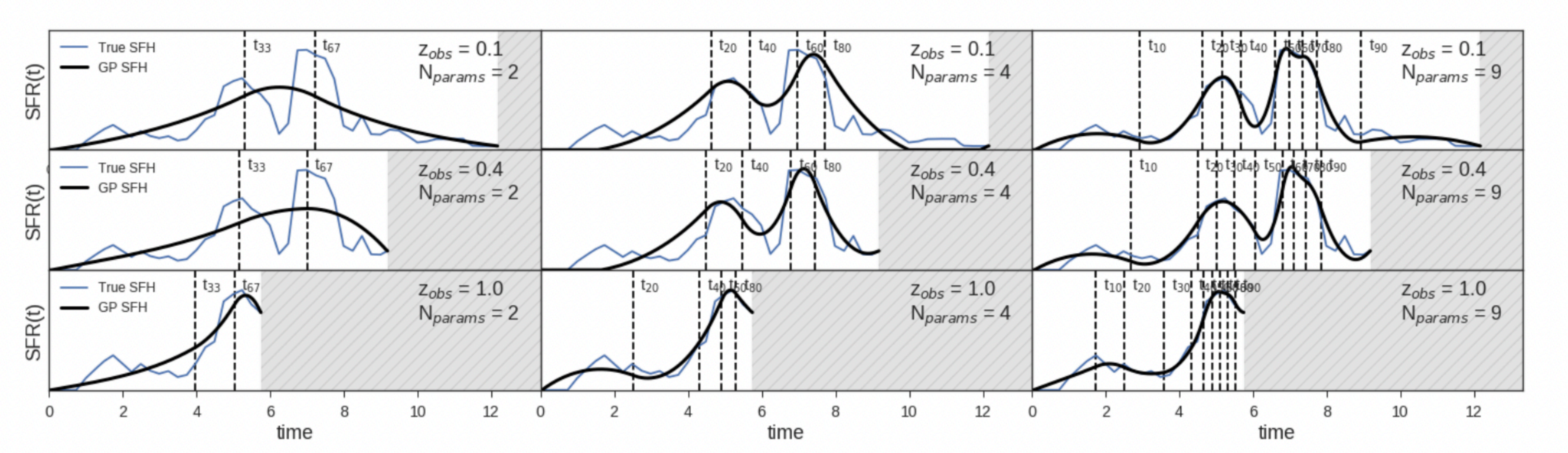
Data for the week



- SEDs for ~20.000 z=0 galaxies with log(Mstar)>9 from the IllustrisTNG100 simulations + SDDS like images with and without noise. *These SEDs contain only stellar light (no dust)
- Full radiative transfer modelled SEDs (inc. dust) for ~2500 IllustrisTNG50 galaxies.
- For each of the datasets, we have extracted from the simulation the SFHs and stellar mass growth curve of the main progenitor

Possible approach 1

- Describe the SFH with a set of percentiles (Iyer+20)
- Build a NN from SED+images to set of scalars
- Uncertainties are important here, BNNs, others?



Iyer+19

Possible Approach 2
(Advanced ML)

- Generative Modelling + Flow for full reconstruction of SFHs