Knowledge Extraction from Machine Learning

Luisa Lucie-Smith
University College London

with H.V. Peiris (UCL, Stockholm), A. Pontzen (UCL),
M. Lochner (AIMS), B. Nord (Fermilab)

Lucie-Smith, Peiris, Pontzen (2019), arXiv:1906.06339
Lucie-Smith, Peiris, Pontzen, Lochner (2018), arXiv:1802.04271

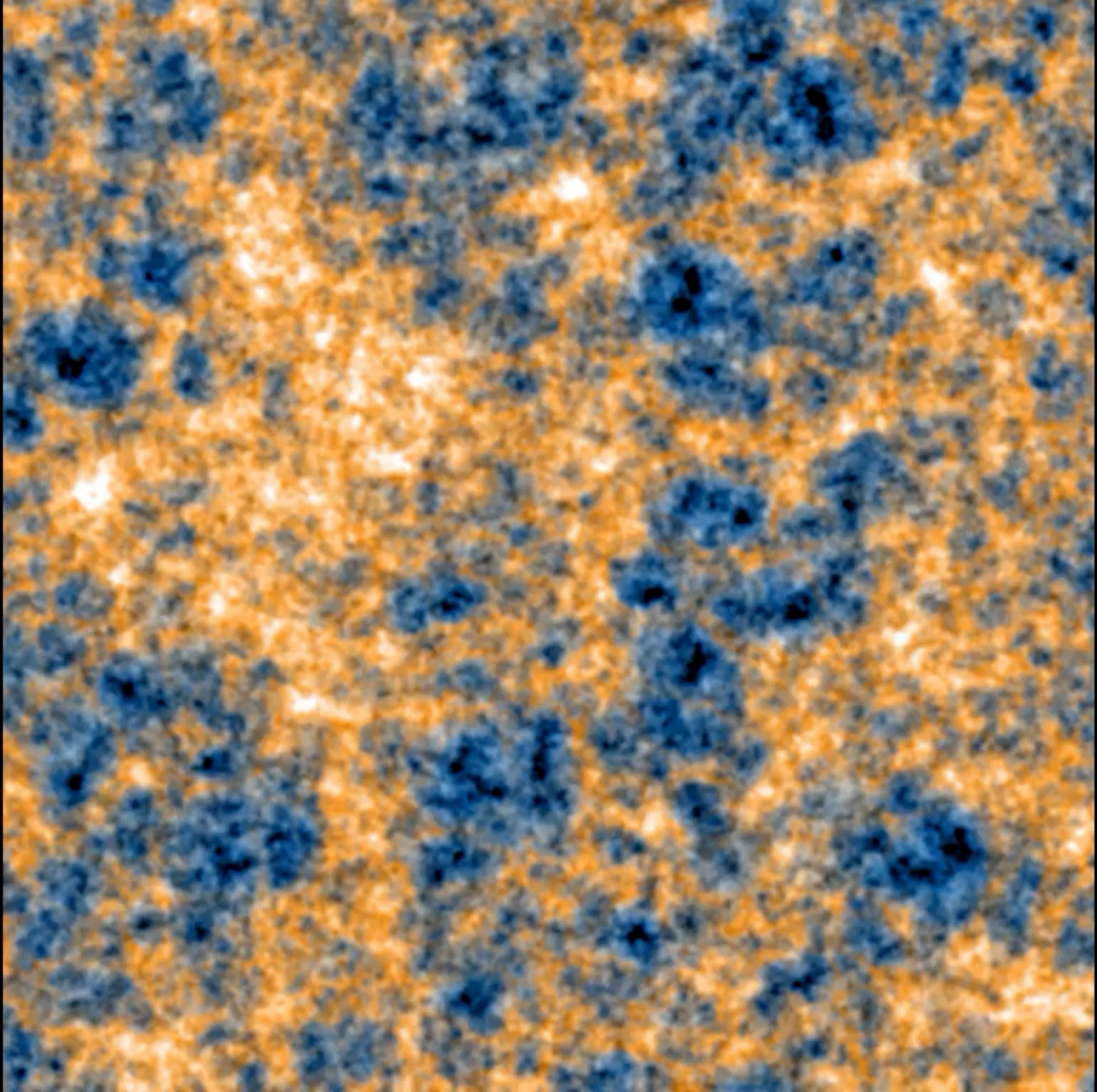
Machine learning in Astronomy

Big-data era: many traditional applications of ML.

- data mining
- classification
- data compression
- data and model emulation
- regression
- clustering analyses
- outlier detection

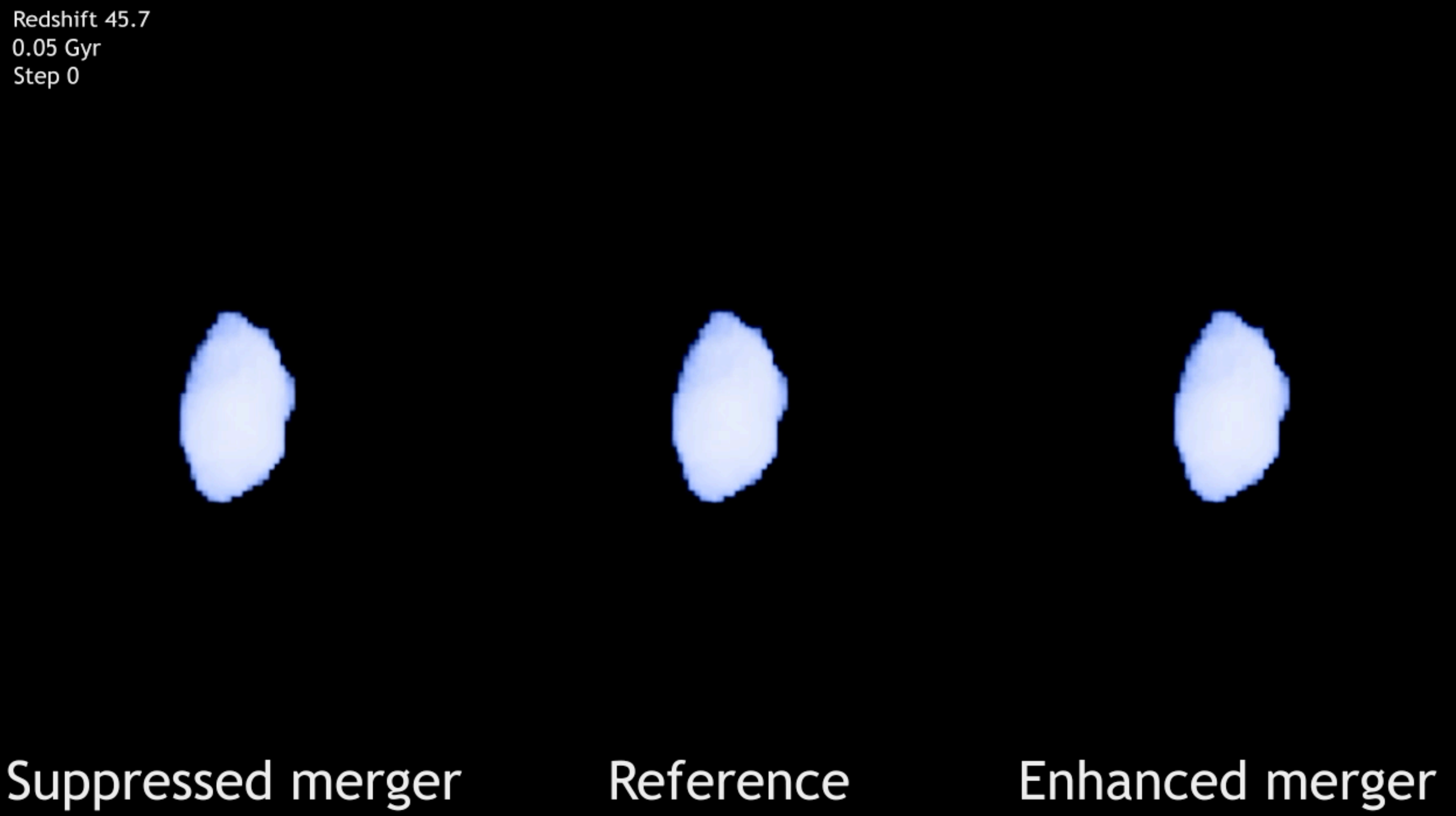
Can ML enable knowledge extraction?

- can we extract new physical insights by studying the learning of ML algorithms?



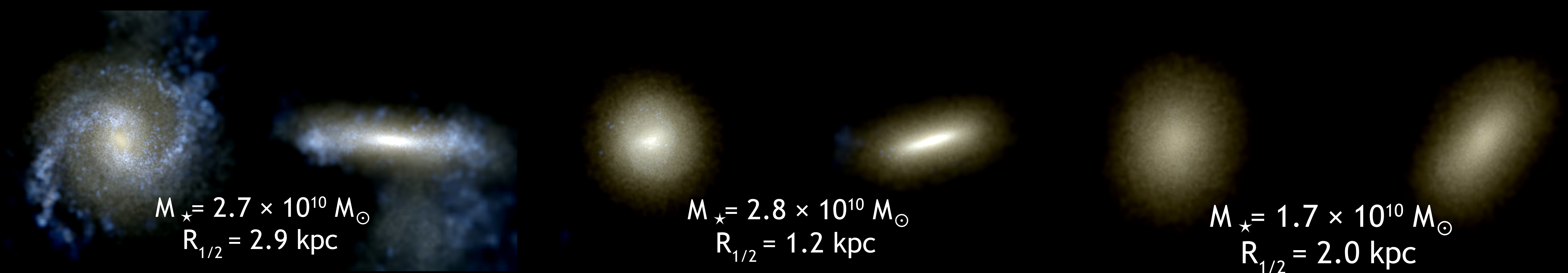
Credit: Andrew Pontzen, UCL

The genetic modification method



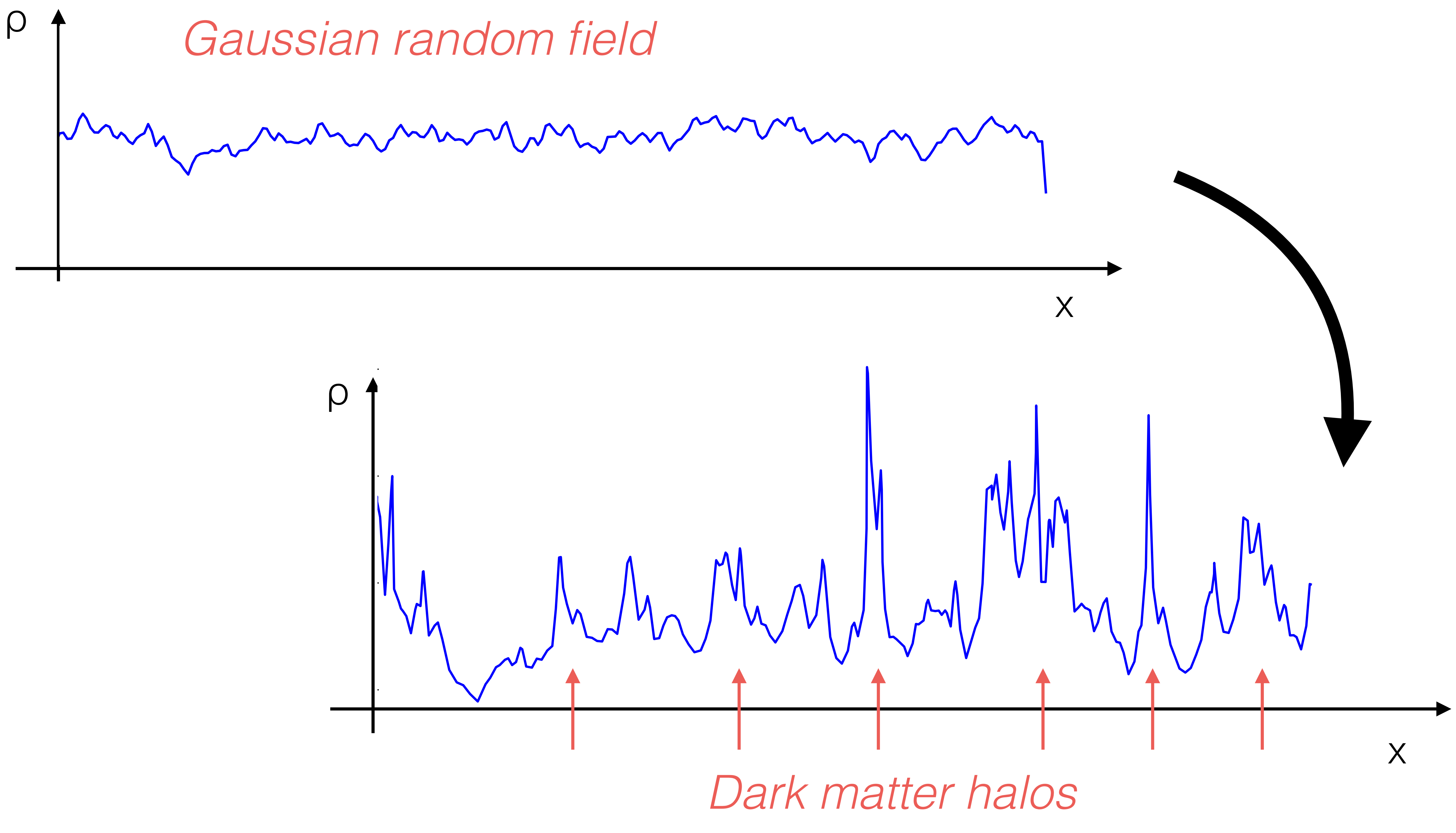
Pontzen+ 1607.02507; Tremmel+ 1607.02151

“Genetically-modified” galaxies

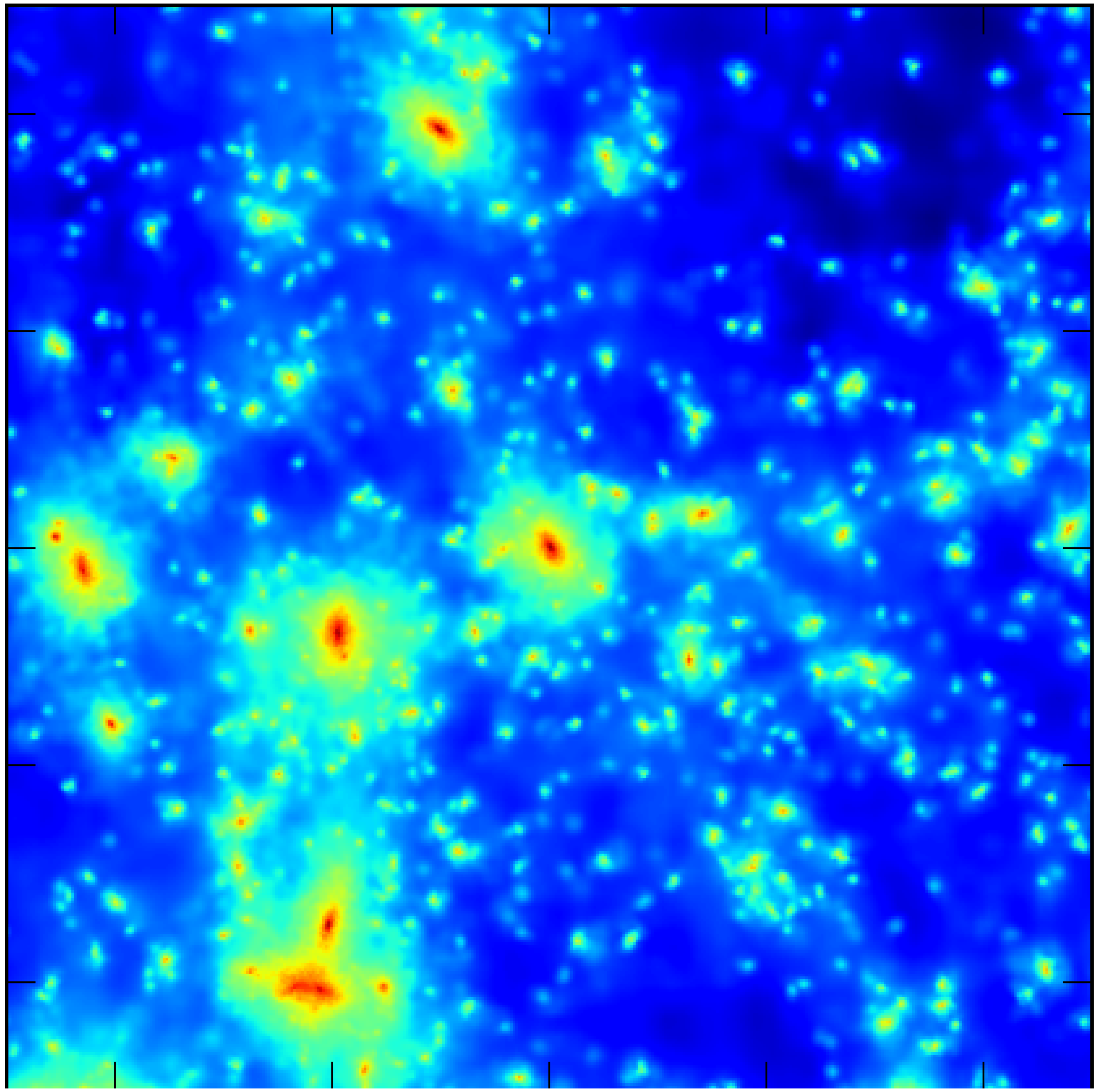


Suppressed merger　　Reference　　Enhanced merger

Non-linear dark matter halo formation



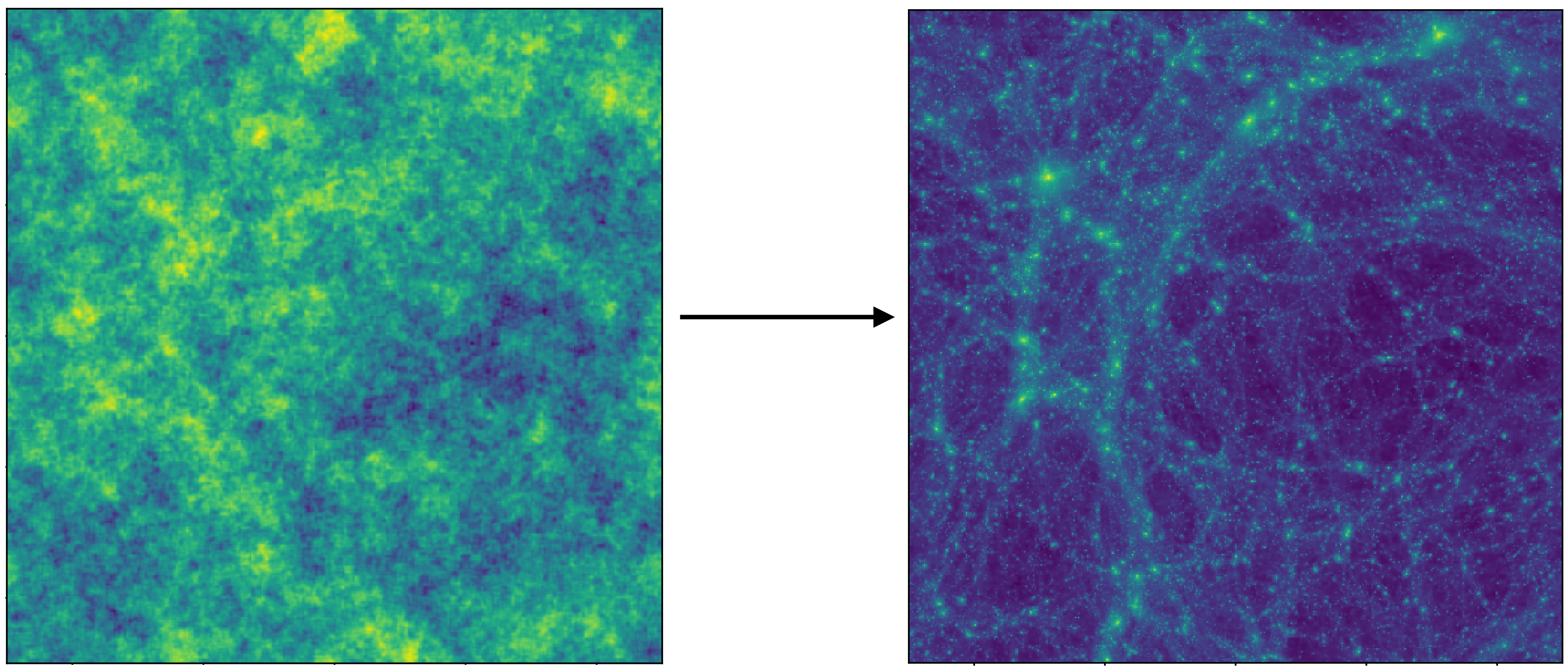
N-body simulations



Difficult ***physical*** interpretation from numerical studies alone

Insights into dark matter halo collapse from ML?

Approach*:* Train ML algorithm to learn mapping between initial conditions and dark matter halos from N-body simulations

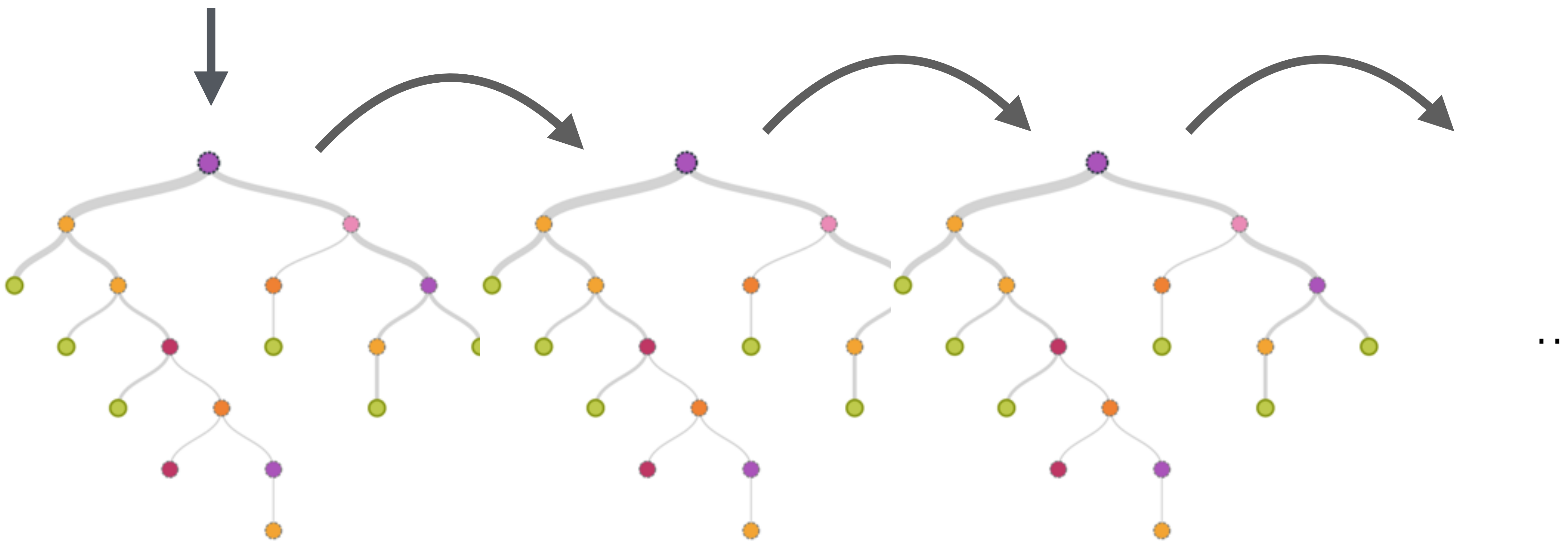


Aim*: gain new physical insights into the process of dark matter halo formation*

Lucie-Smith, Peiris, Pontzen (2018, 2019)

ML algorithm: gradient boosted trees

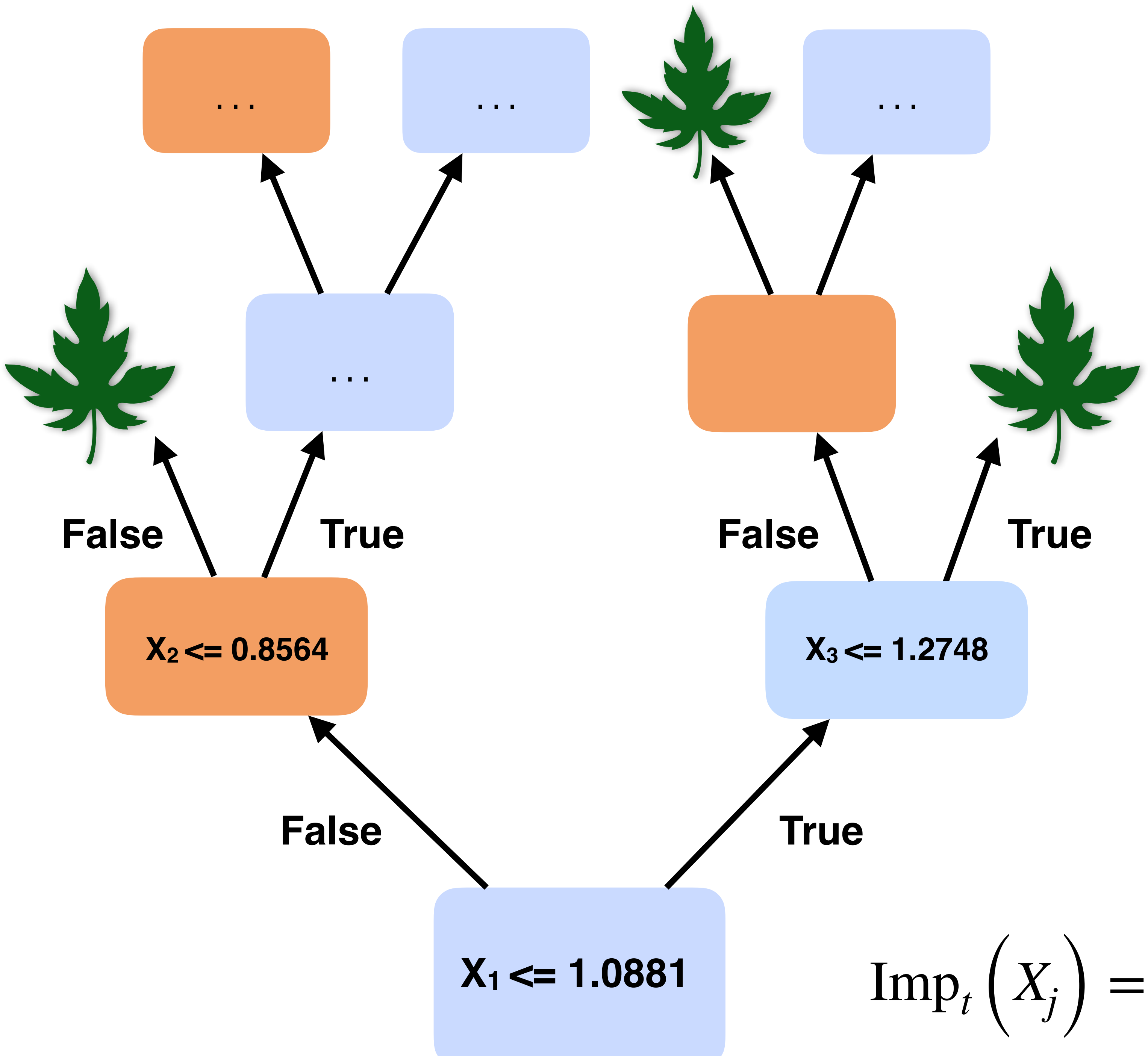
GBTs add new trees to correct the mistakes of the previous ones



$f_1(x)$ + $f_2(x)$ + $f_3(x)$ + …

Decision Tree

Feature Importance



fraction of samples

$$\mathrm{Imp}_t\left(X_j\right) = \sum_n \frac{N_n}{N_t}\left[p - \frac{N_{n_R}}{N_n}p_R - \frac{N_{n_L}}{N_n}p_L\right]$$

impurity (MSE)

ML regression model of halo formation

Initial conditions (z=99)

Features

Properties of the
local environment
around
DM particles

→

ML algorithm
(GBTs)

→

Final halos (z=0)

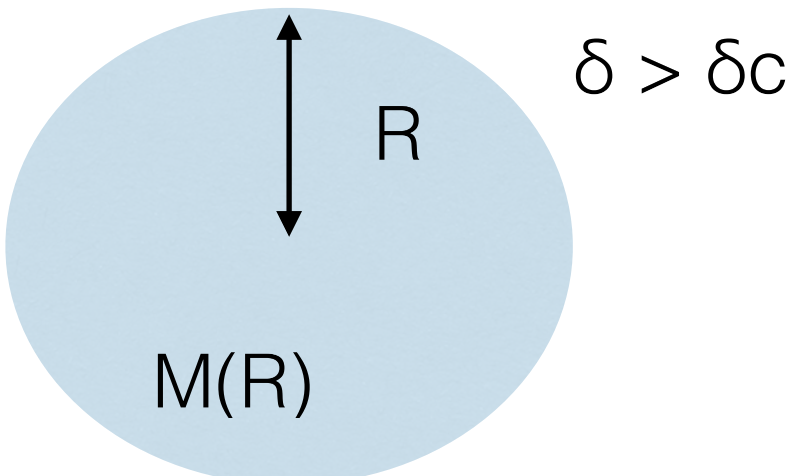
Output

Mass of the halo
to which each DM
particle will belong
at z=0

Our choice of features is motivated by existing analytic approximations of halo collapse

Features based on analytic theories of halo collapse

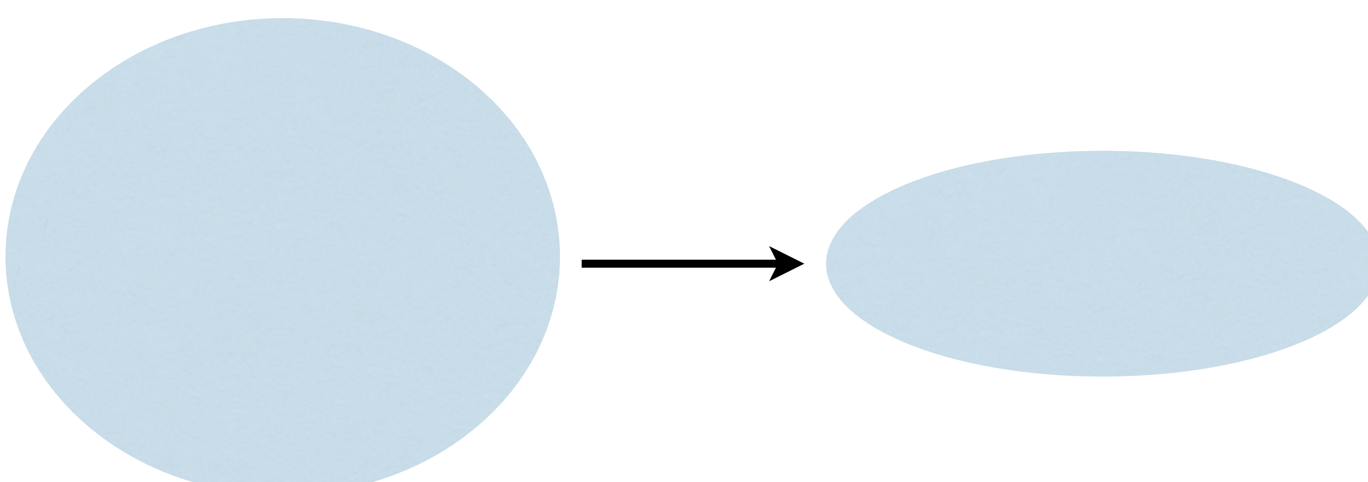
1. *Density contrast: motivated by* ***extended Press-Schechter theory***



Density contrast above threshold δ_c

Dark matter halo of mass M(R)

2. Tidal shear field (ellipticity/prolateness: motivated by ***Sheth-Tormen theory***

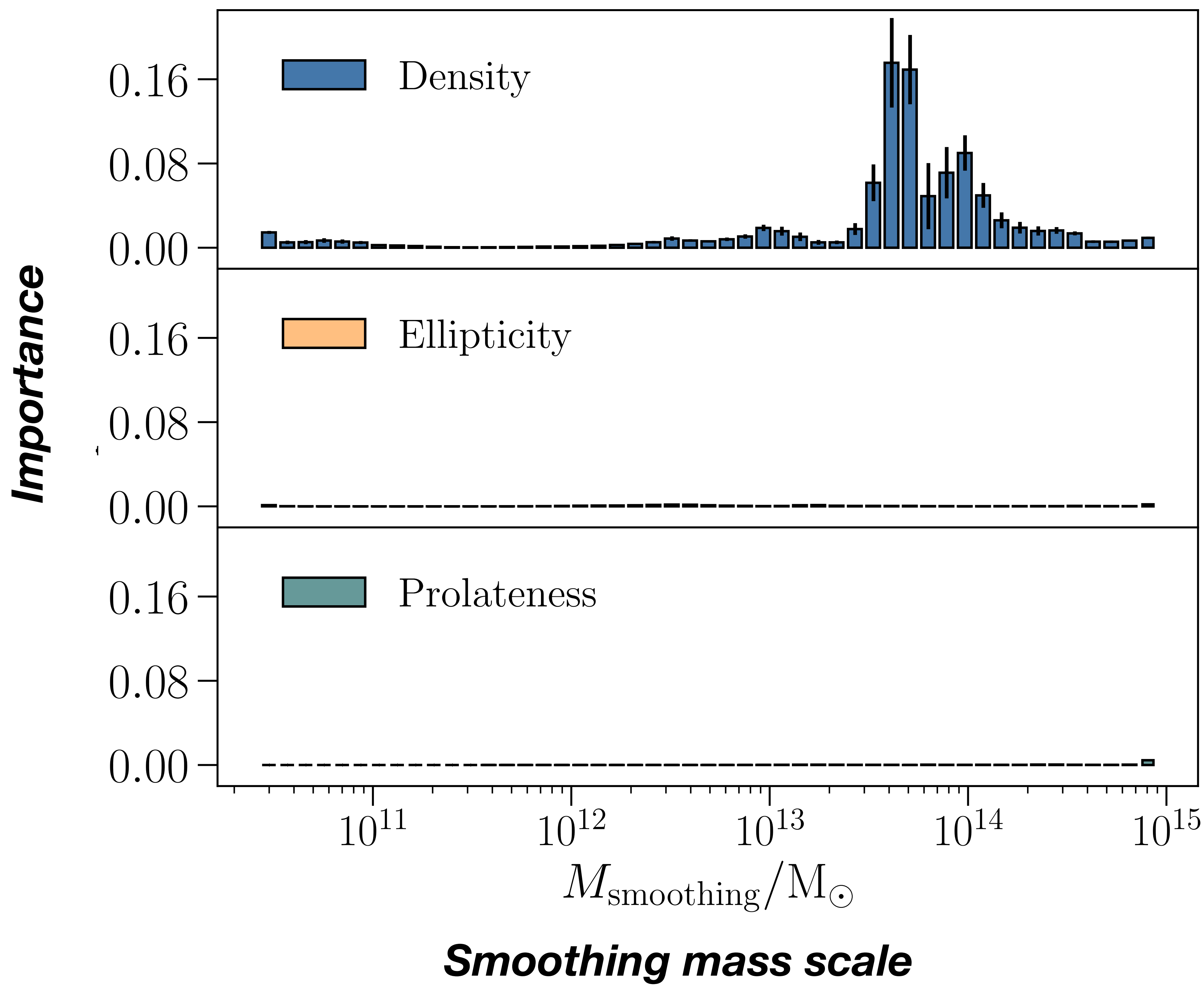


Tidal shear forces distort spheres into ellipses

Final halo mass M(R) depends on tidal shear field

Compute features in spheres of 50 different mass scales

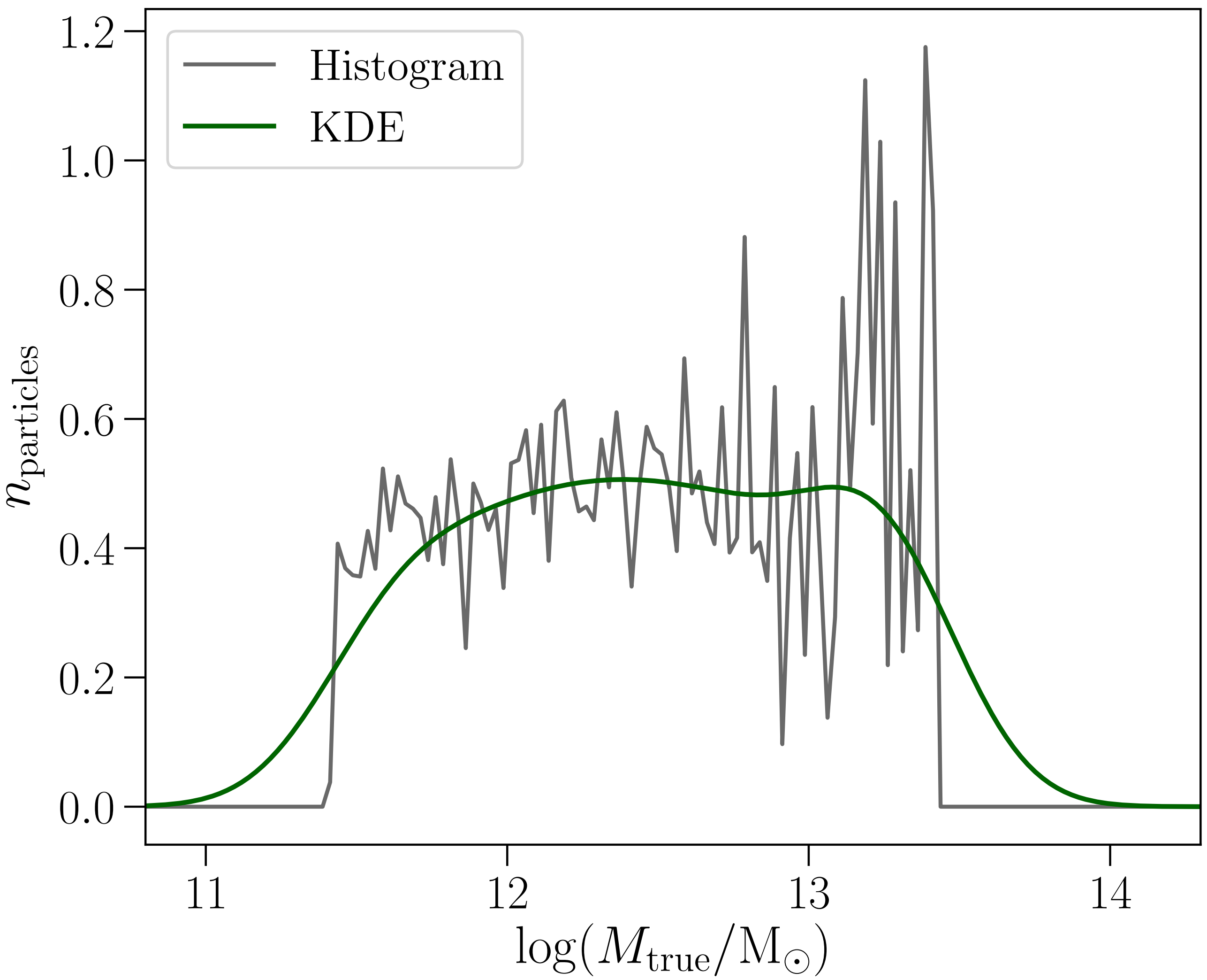
Which features were most informative?



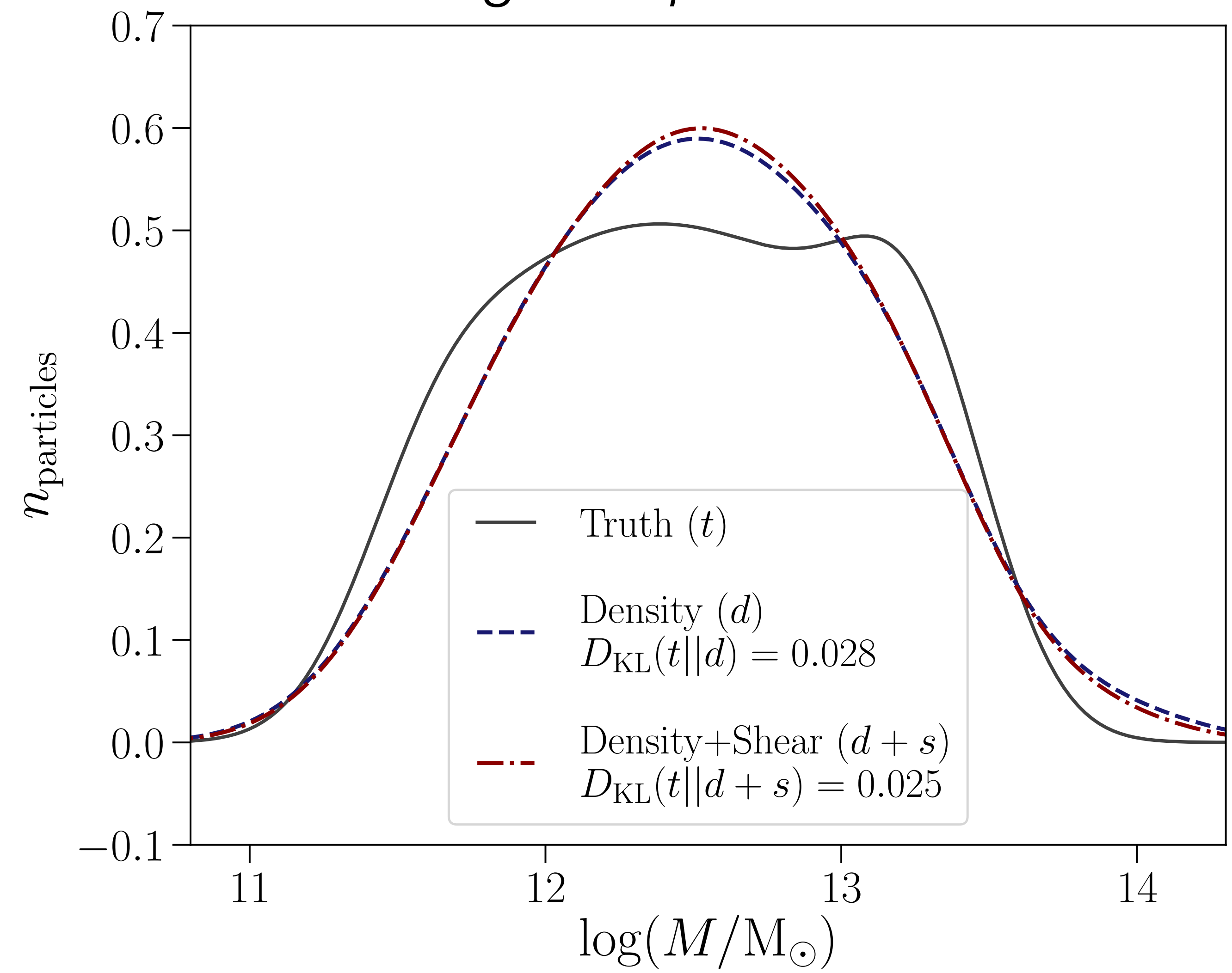
Lucie-Smith, Peiris, Pontzen (2019)

Machine learning model comparison: Kullback-Leibler (KL) divergence

1. Smooth distributions with KDE



2. KL divergence prediction vs truth

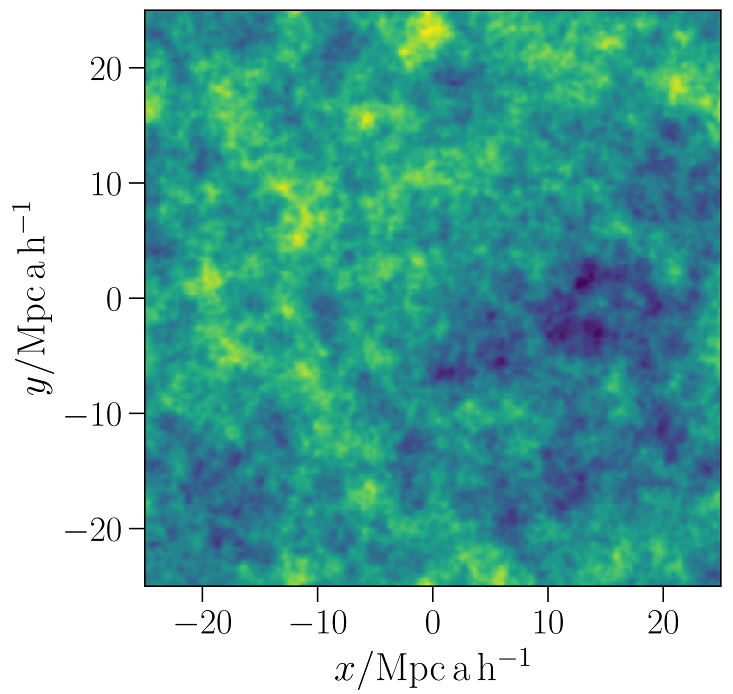


Addition of tidal shear information does not yield an improved halo collapse model in contrast to standard interpretations of Sheth-Tormen theory

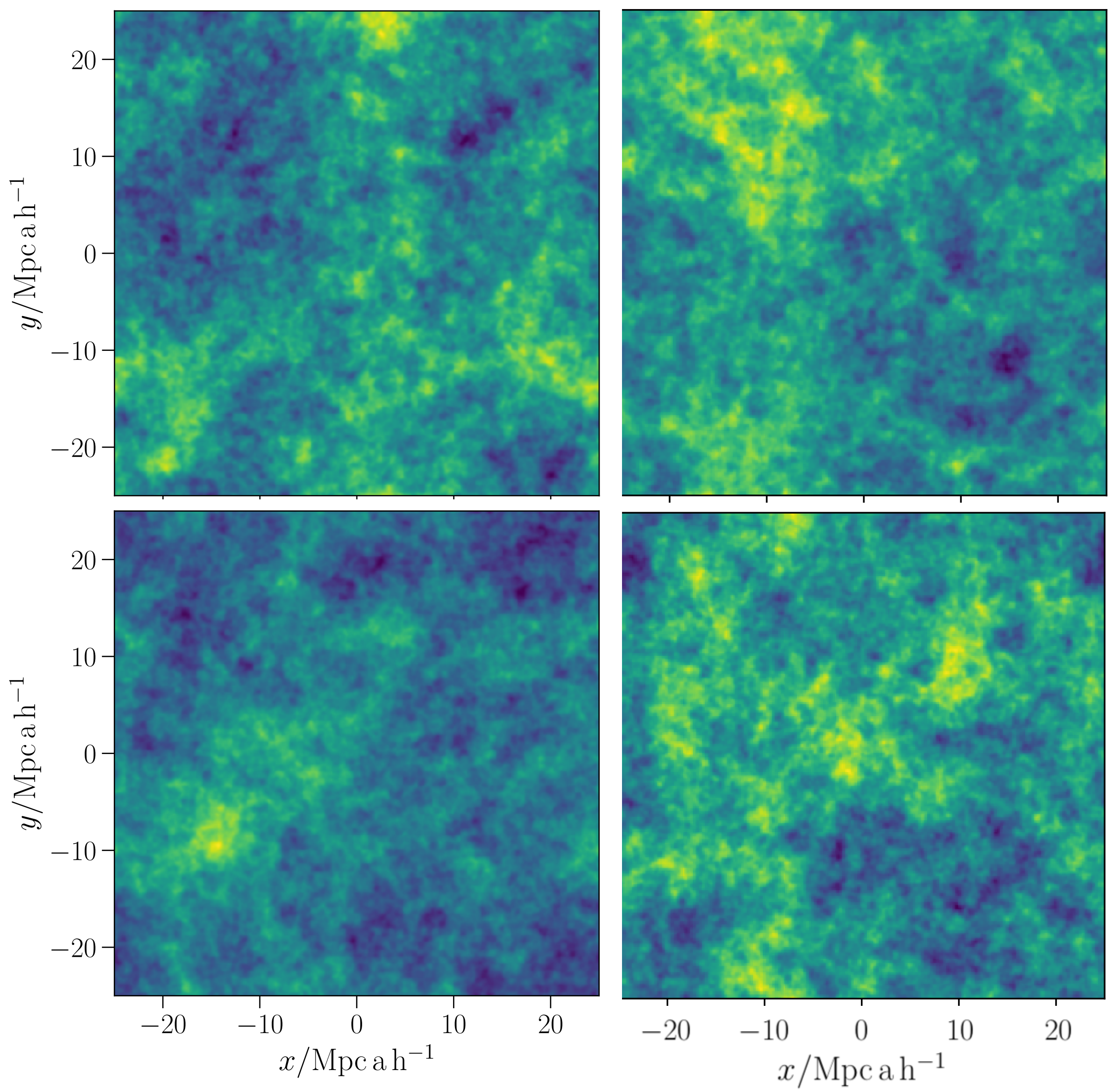
Lucie-Smith, Peiris, Pontzen (2019)

Do the results generalise to independent simulations?

One training simulation

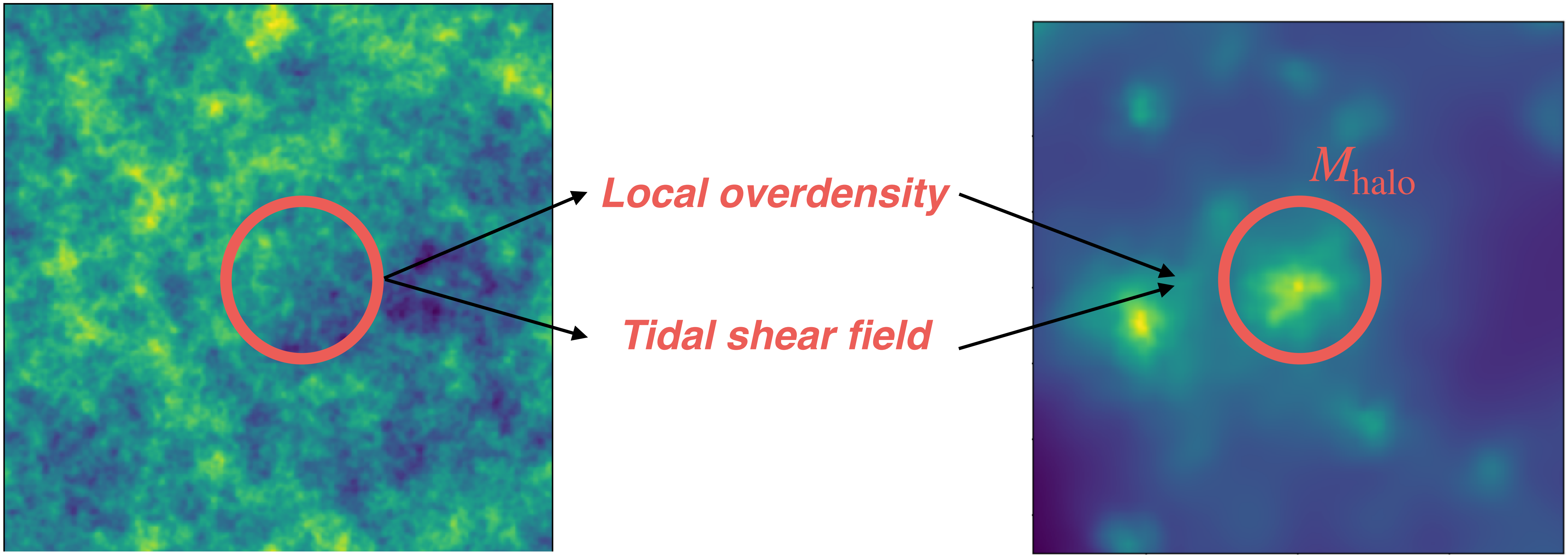


Four independent test simulations



*ML algorithm learnt **physical connection** between initial conditions and halo masses*

What we have learnt so far…



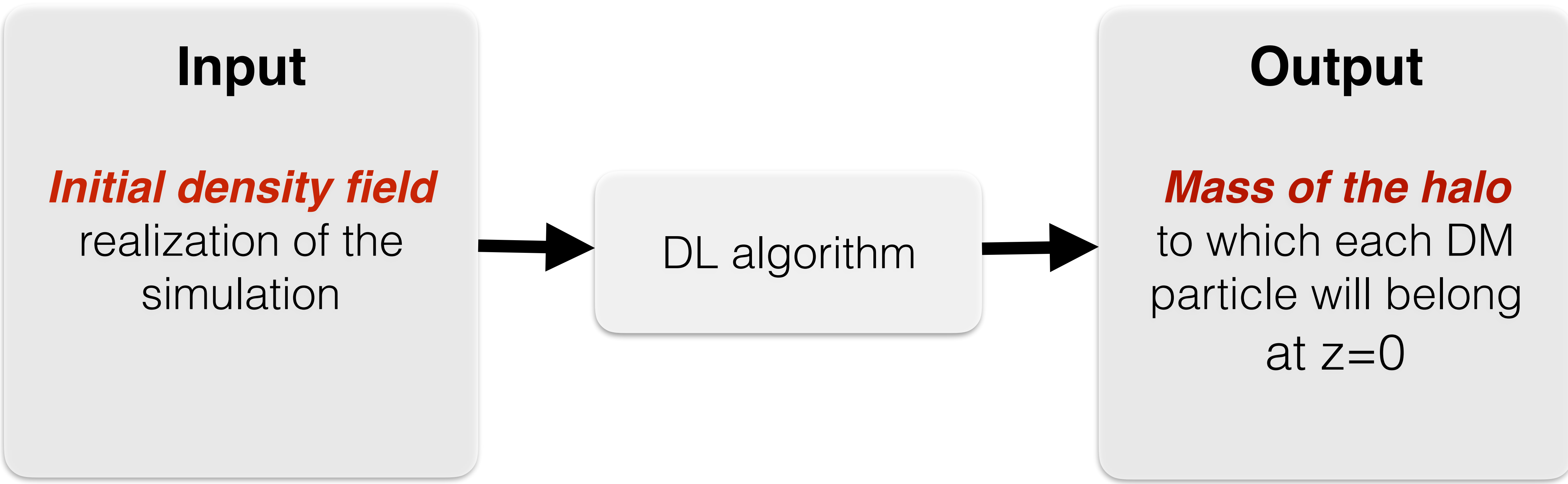
Addition of tidal shear information does not improve halo collapse model

How can we go beyond testing current interpretations of halo collapse?

A deep learning approach to halo formation

Advantages:

- *do not require featurization!*
- *provide as input the “raw data”, i.e. the initial density field*

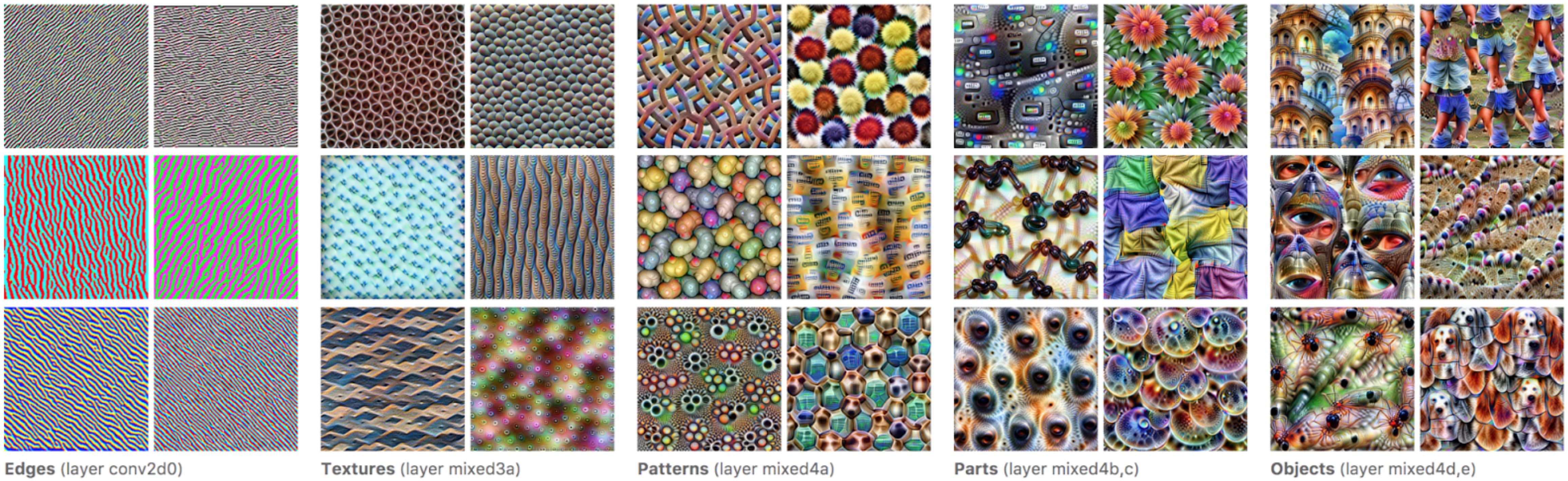


Disadvantages:

- *how do we extract physical knowledge from the DL algorithm?*

Requirements for knowledge extraction from DL

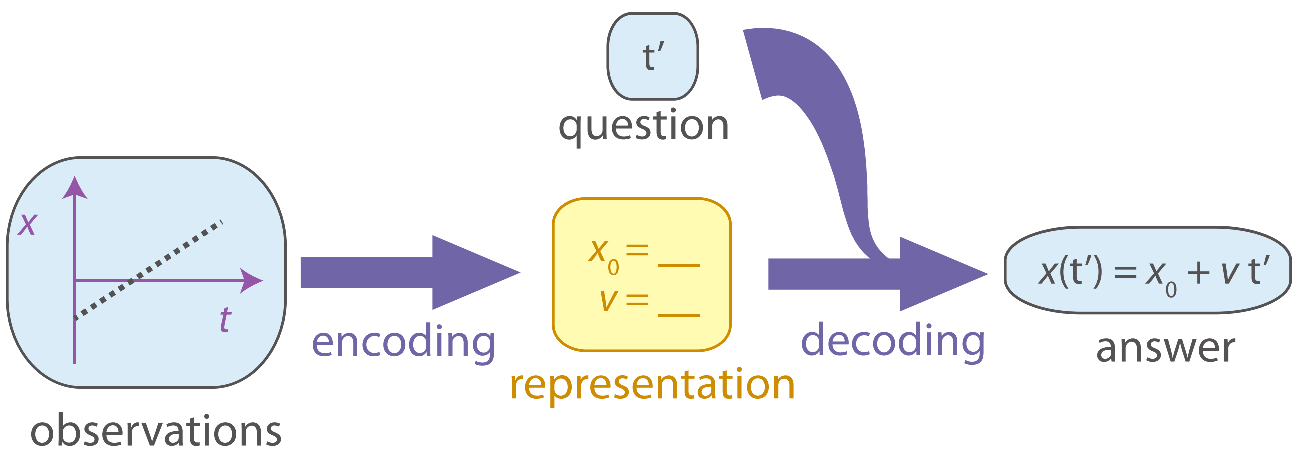
- ***interpretability***: *How did the DL model reach its predictions? Produce outputs that help us understand inner workings DL model.*



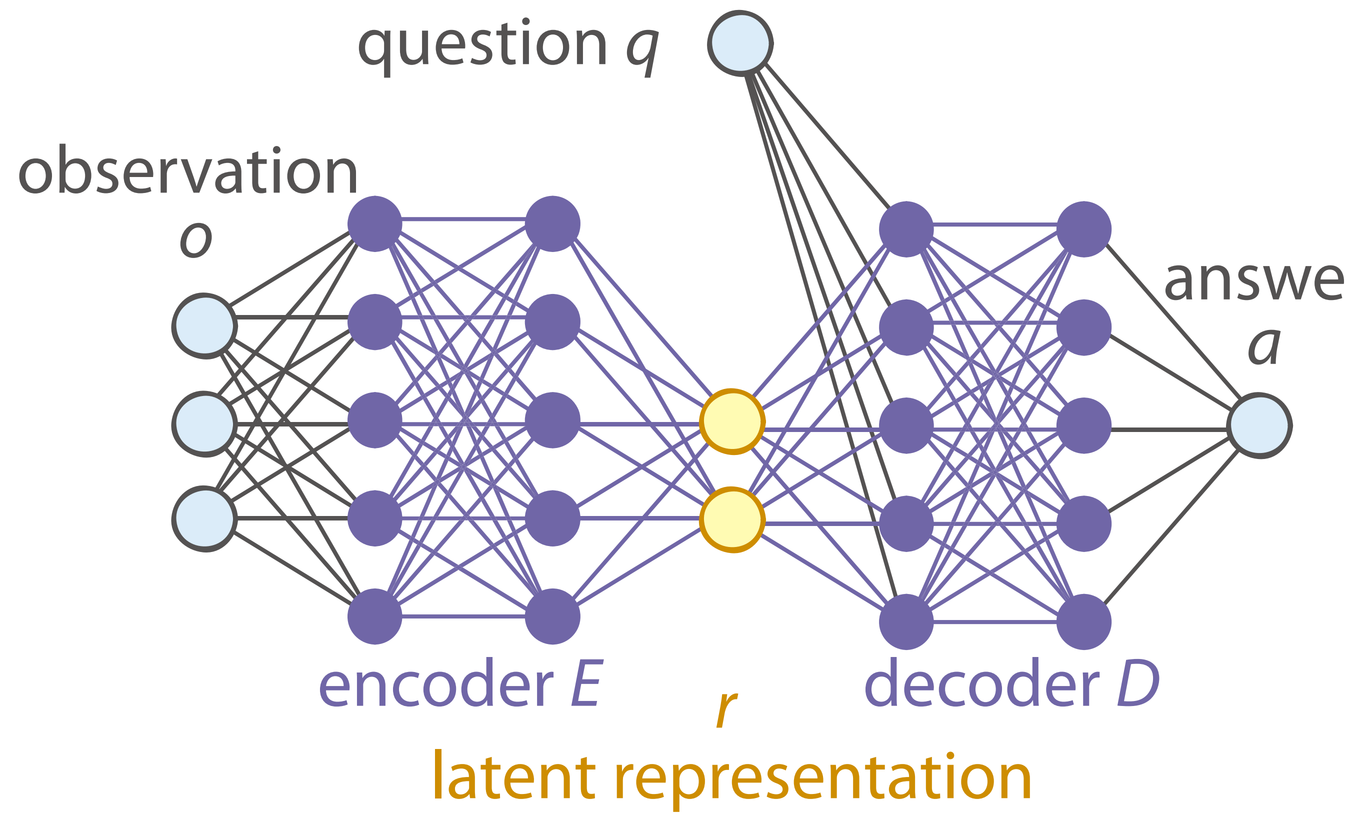
- ***explainability***: *mapping interpretability onto existing knowledge in the relevant science domain.*

Learning physical representations

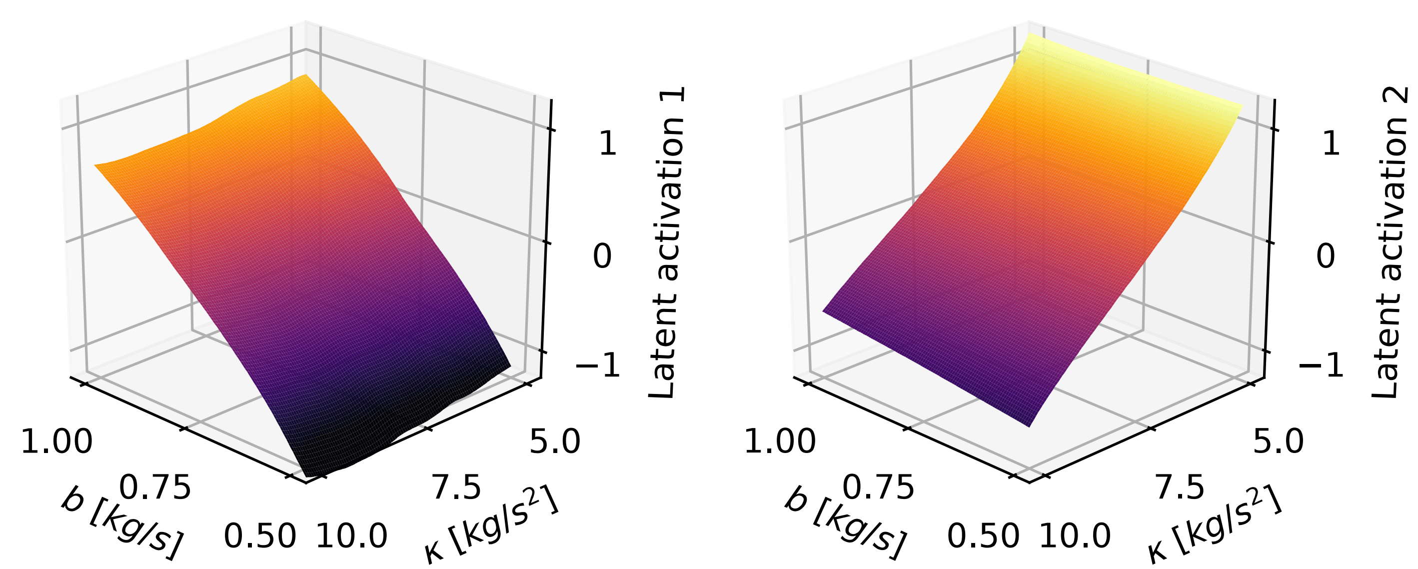
Human learning



SciNet model

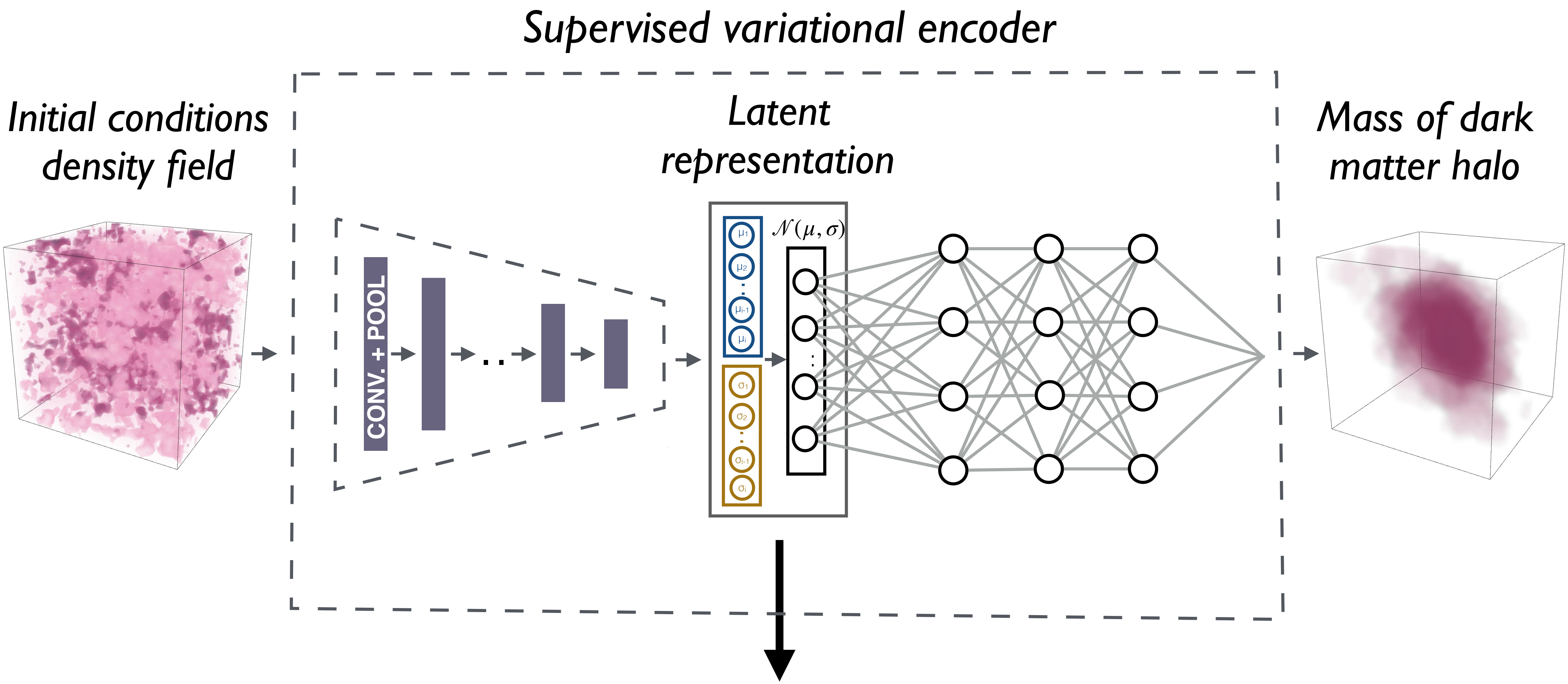


SciNet learns two relevant physical parameters of damped pendulum problem



Iten et al. (2018; arXiv:1807.10300)

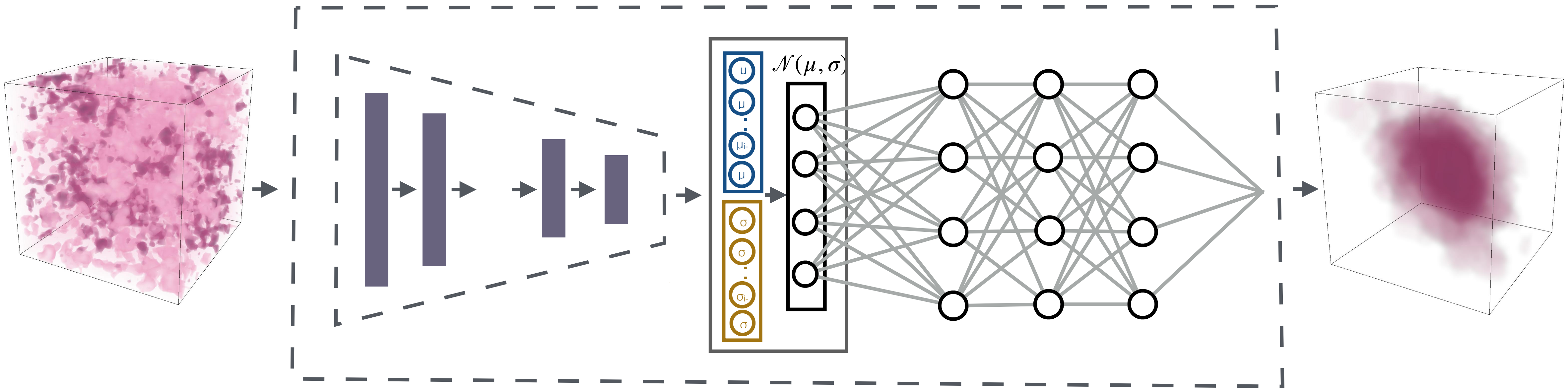
Deep learning for knowledge extraction



Latent variables encode most relevant aspects of initial conditions about final halo masses

Lucie-Smith, Peiris, Pontzen, Nord (in prep)

Deep learning for knowledge extraction



- ***Explainability***: *What physical information is compressed by neural network learning? Correlated with overdensities?*

- *Can provide different fields (e.g. density field and tidal shear field) as different 'channels' (like RGB channels for images)*

Work in progress…

Conclusions

- ML enabled new, surprising and generalisable insights into halo collapse
- **Work in progress:** interpretable deep learning networks (no featurization) to extract new physical knowledge about cosmological structure formation