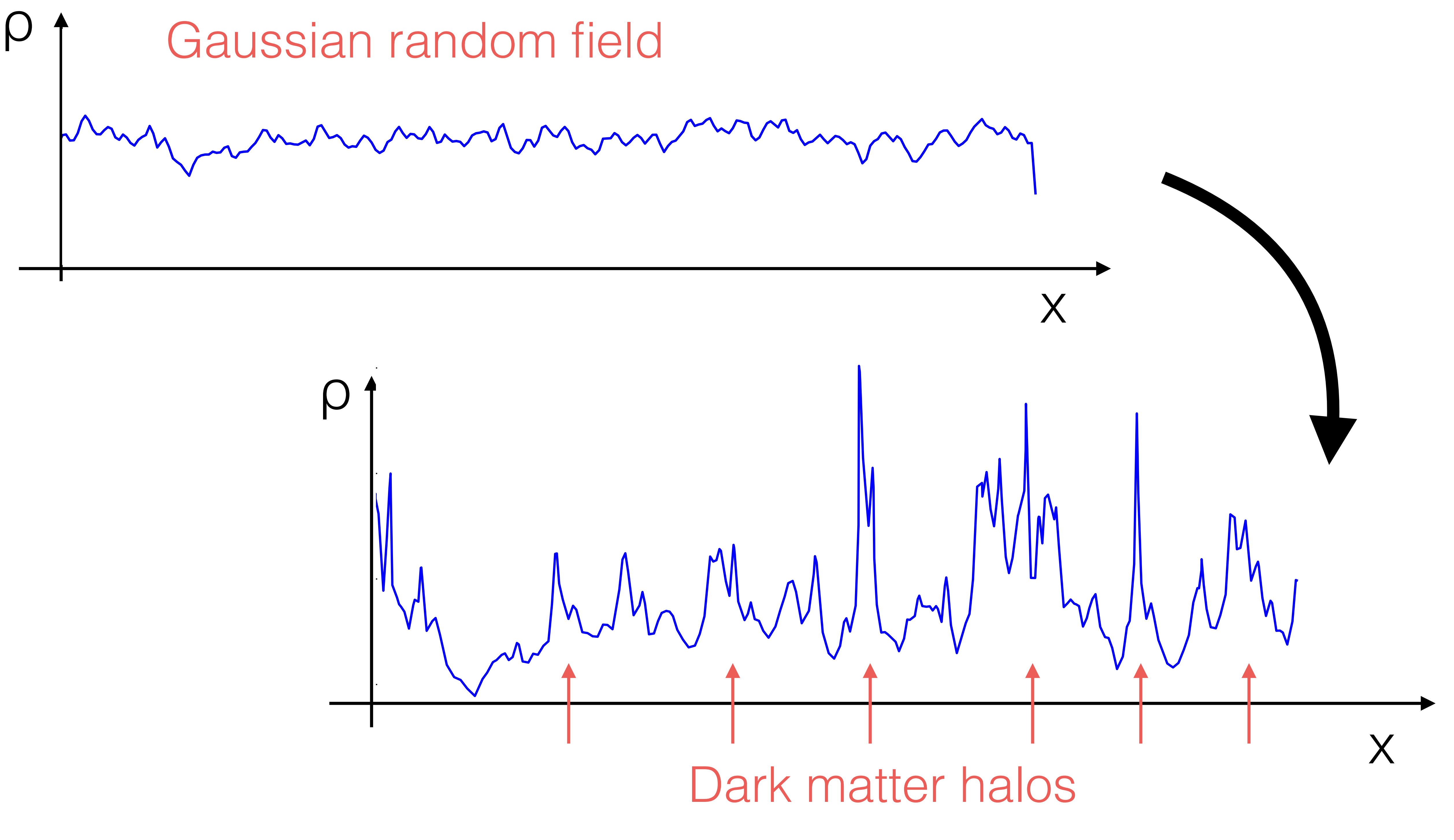
Machine Learning Dark Matter Halo Formation

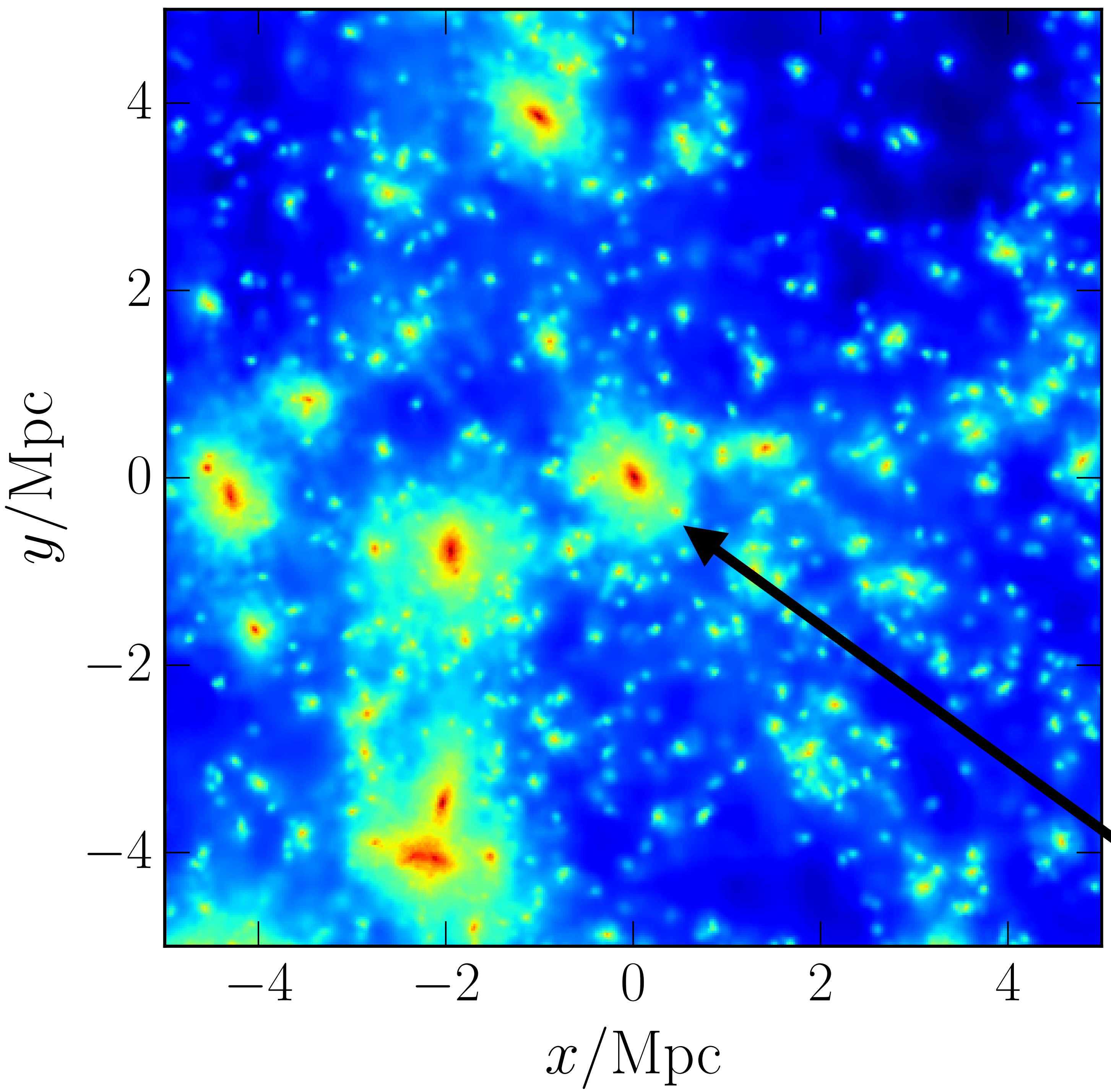
Luisa Lucie-Smith
University College London

with H.V. Peiris, A. Pontzen

Cosmological structure formation



N-body simulations



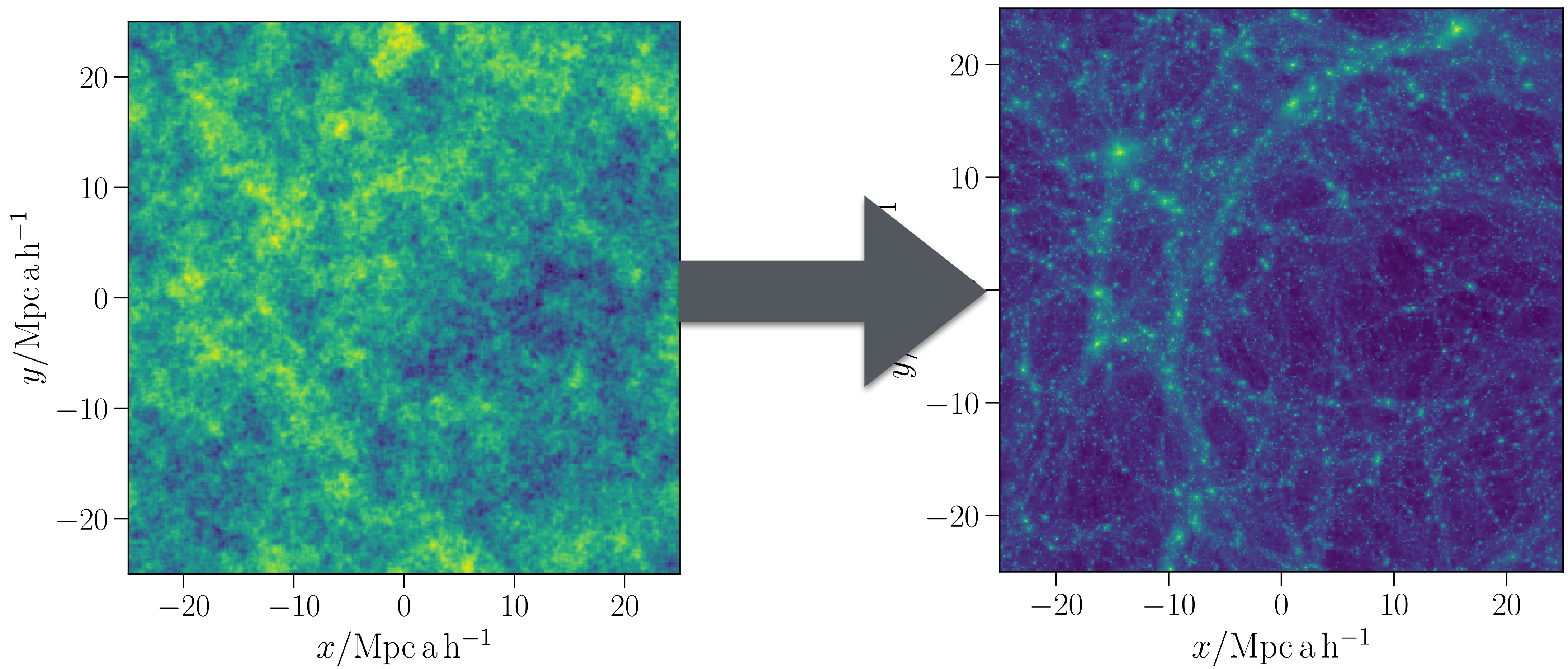
Evolve dark matter "particles" through cosmic time

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

Dark matter halo

A machine learning approach

Train a machine learning algorithm to learn the mapping between the initial conditions and the final dark matter haloes of N-body simulations.



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

Lucie-Smith et al. (2018)

Binary classification

Initial conditions (z=99)

Final halos (z=0)

FEATURES

Information about the local environment around DM particle

ML algorithm

OUTPUT

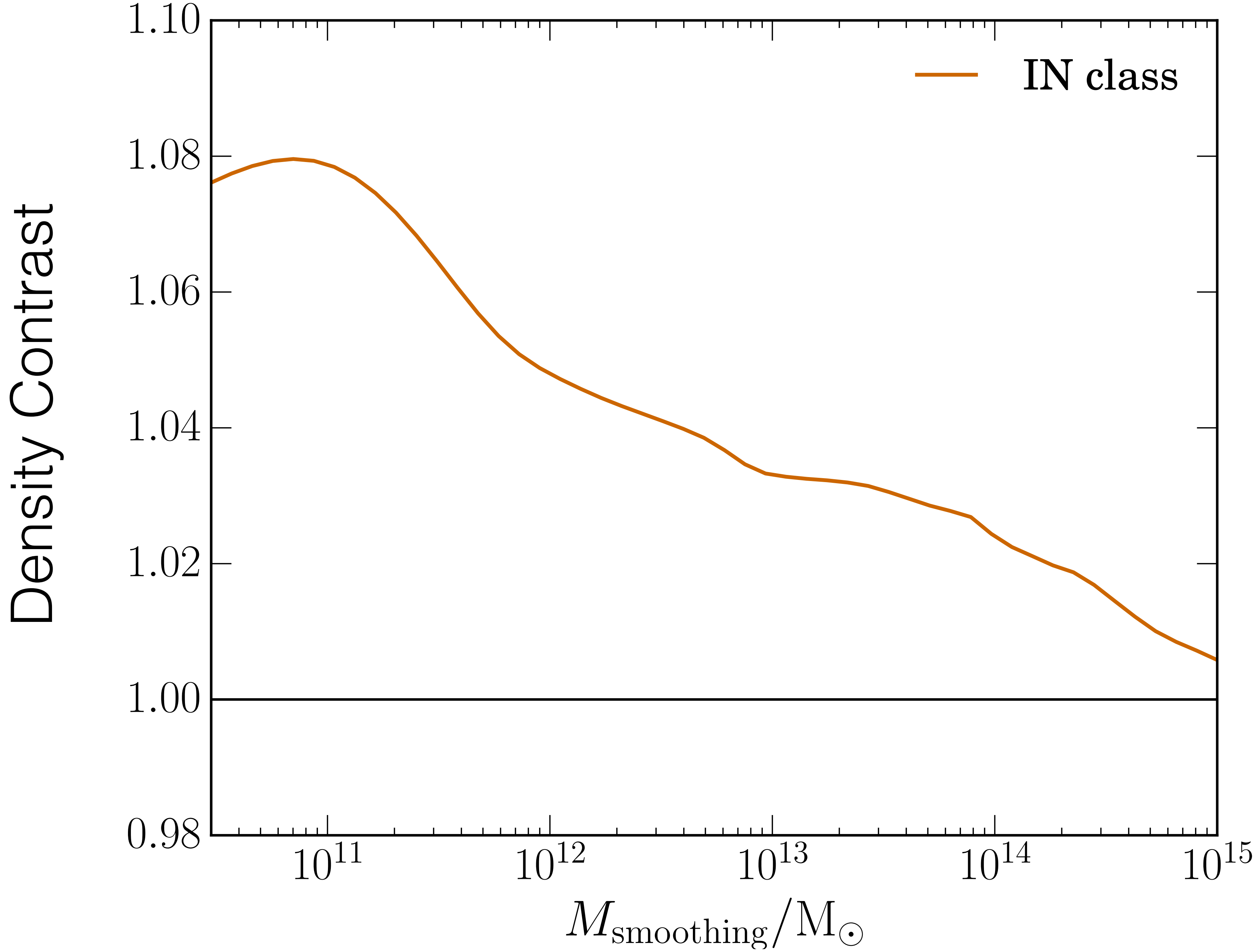
Predict whether DM particles end up:

(1) **IN** halos of mass, $10^{12} < M / M_{\odot} < 10^{14}$

(2) **OUT**, otherwise.

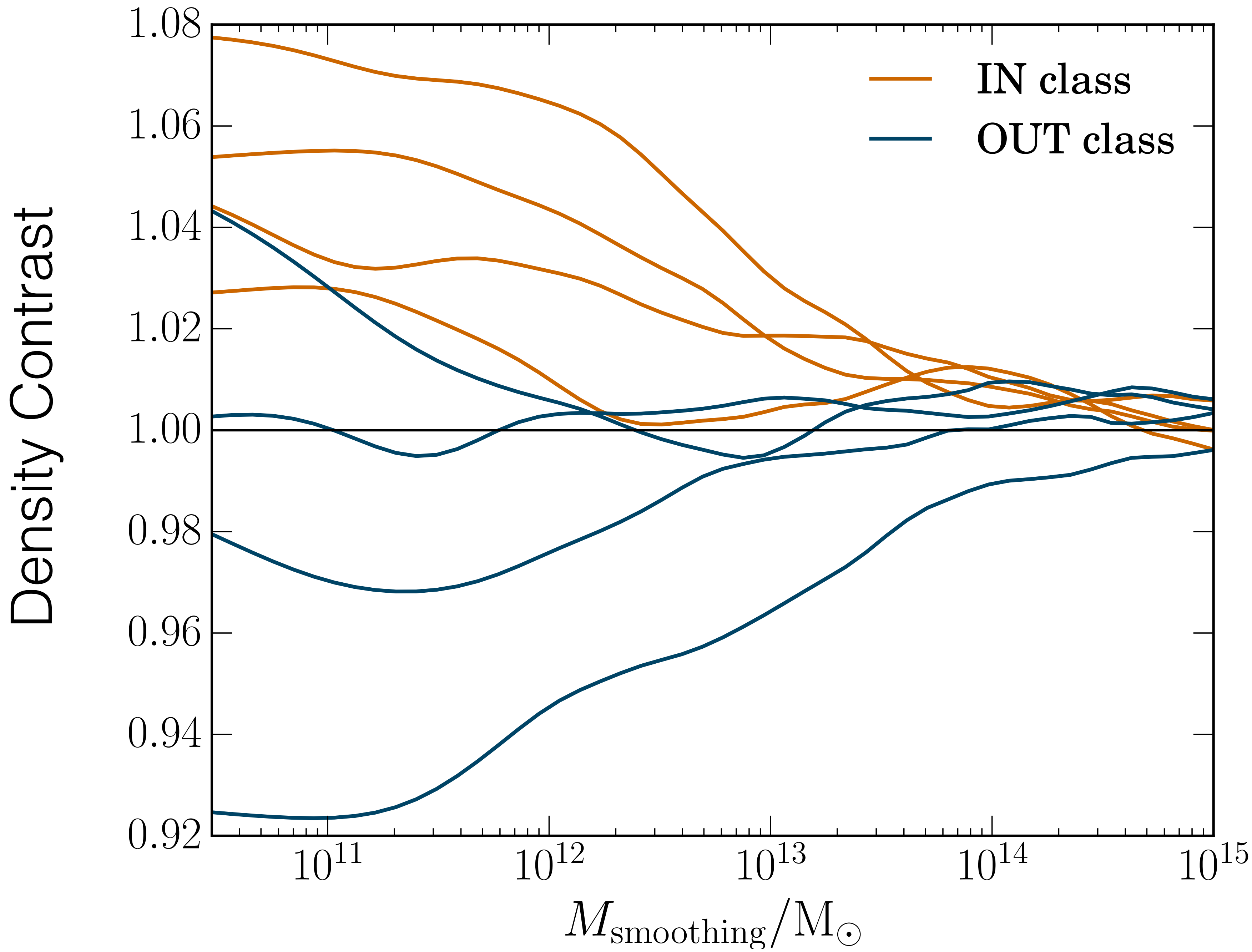
Density features

Density contrast in spheres of 50 different smoothing mass scales centred on each particle's initial position

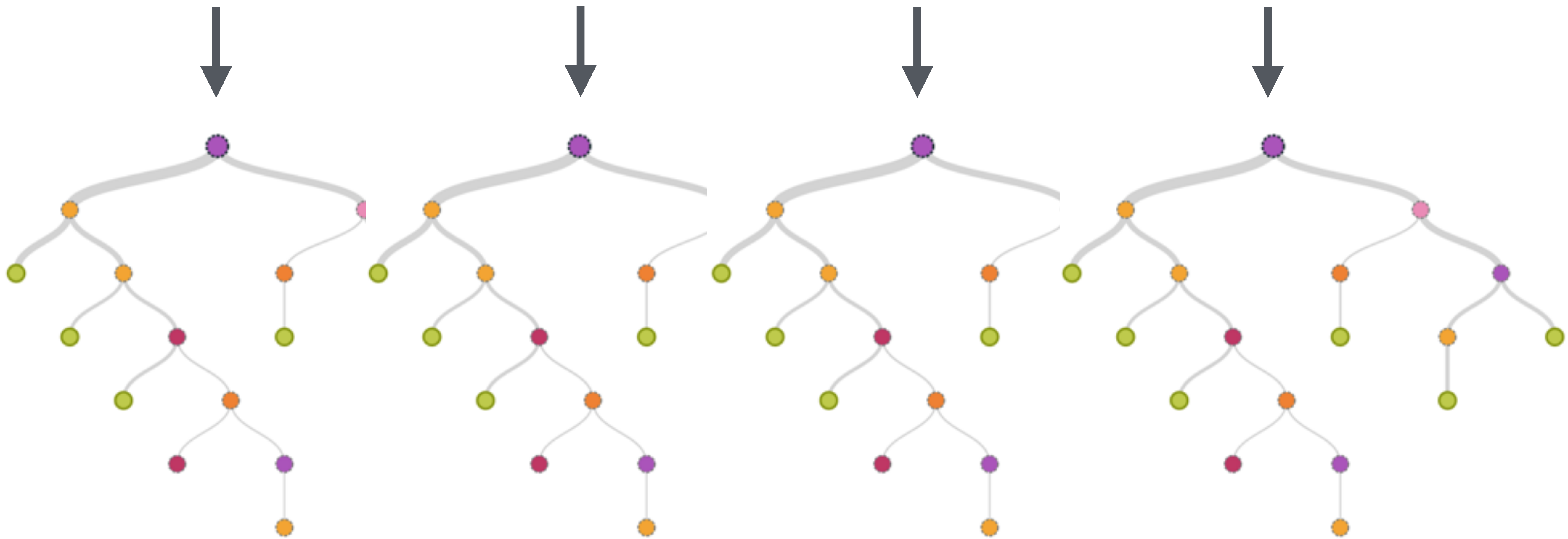


Density features

Density contrast in spheres of 50 different smoothing mass scales centred on each particle's initial position



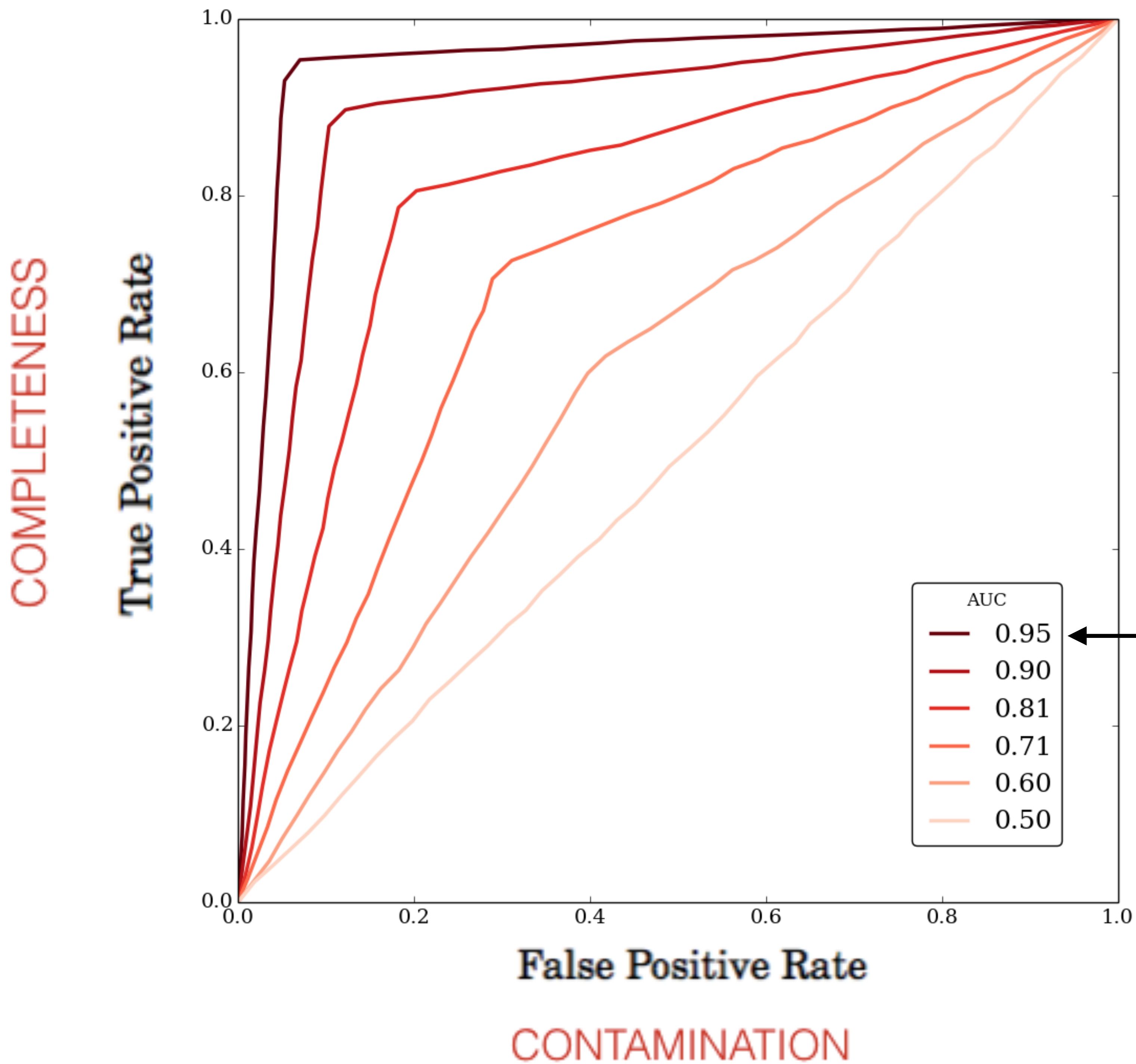
Random forest



Decision Tree

Final prediction =
average probabilistic predictions

Receiver Operating Characteristic (ROC) curves

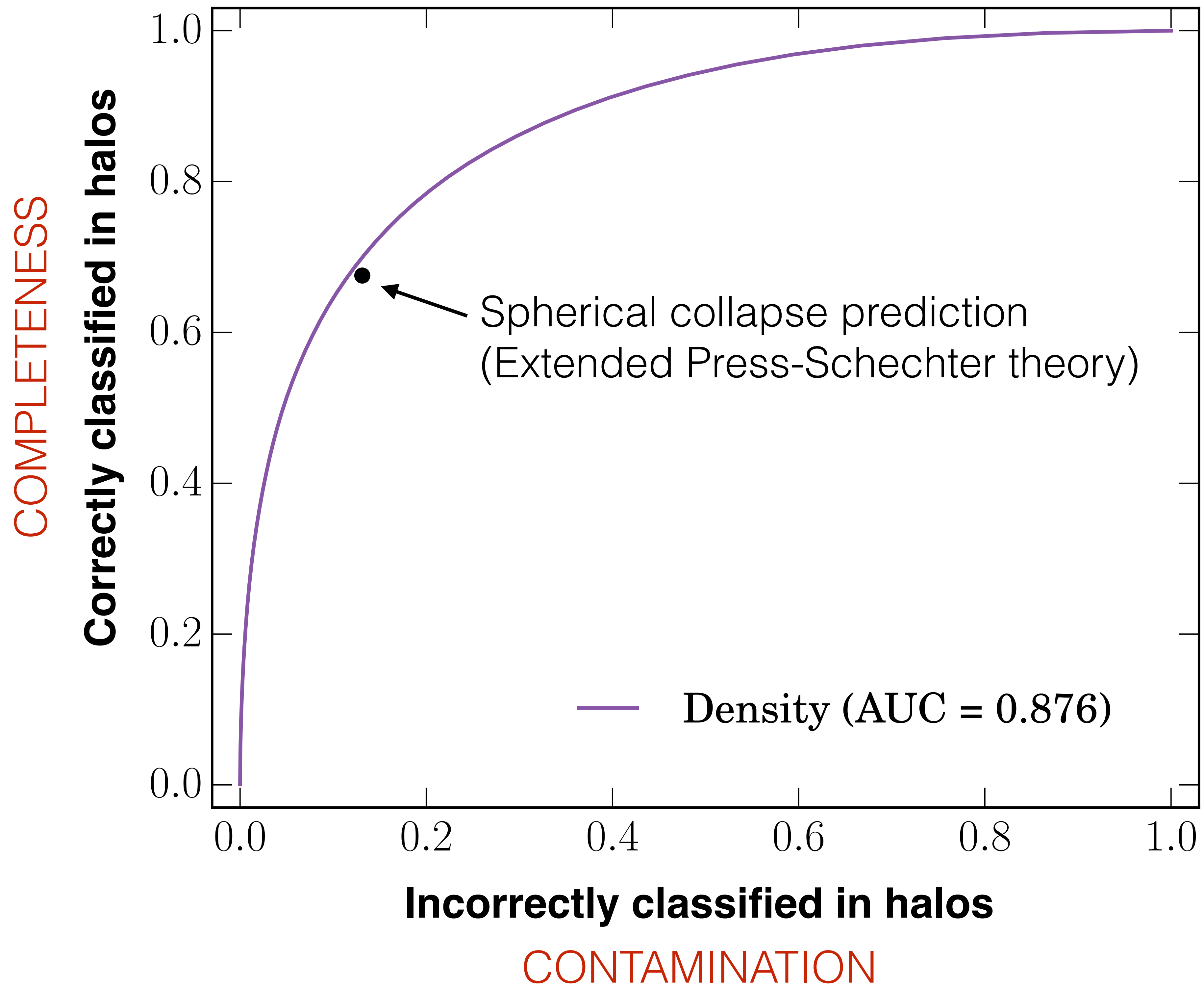


Toy examples of ROC curves

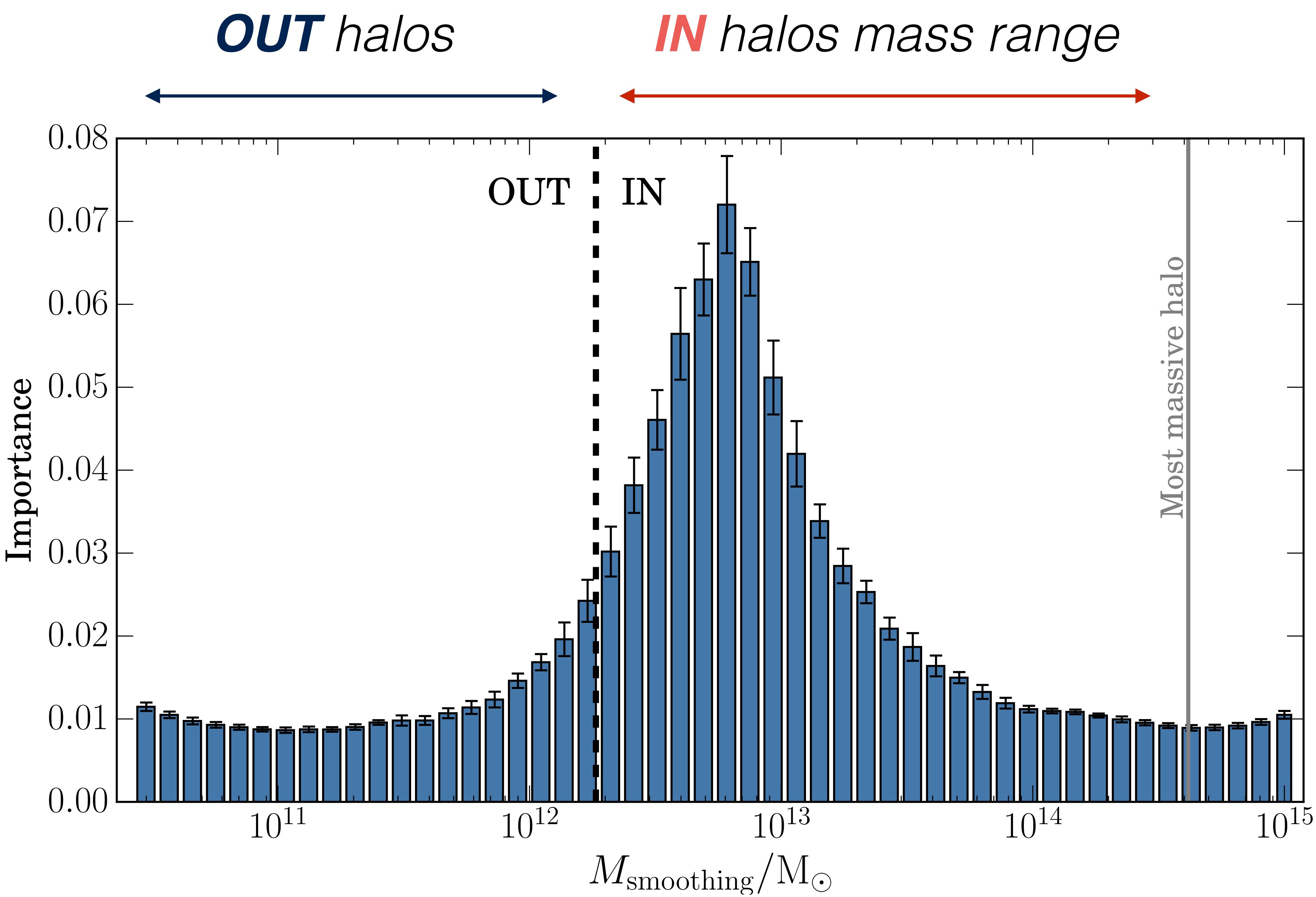
Area under the curve

Credit: Michelle Lochner

Machine learning vs extended Press-Schechter



Density Importances



Additional physics

- **Tidal shear effects** affect the formation of dark matter halos. Motivated by *Sheth-Tormen (ST) theory* on ellipsoidal collapse

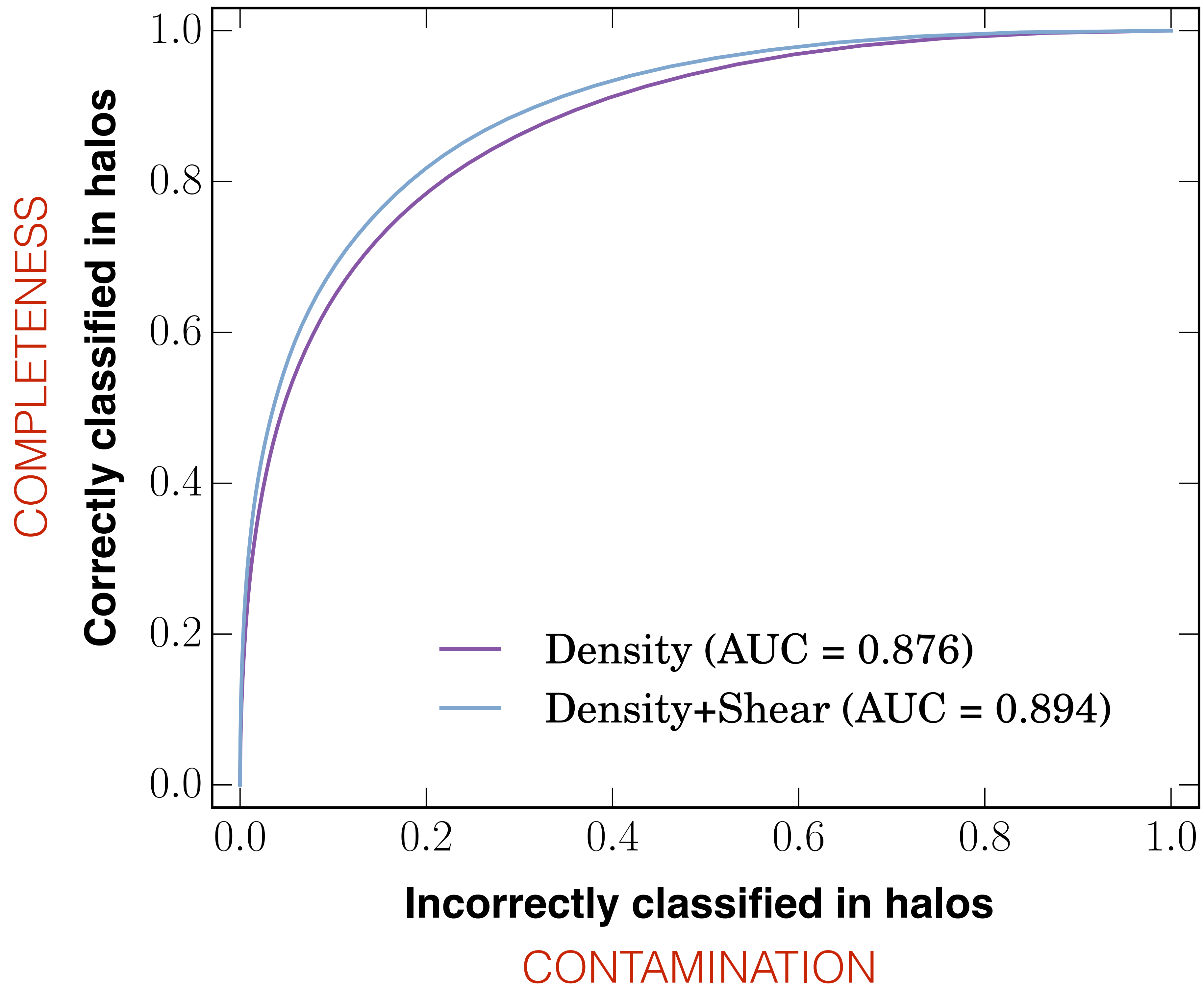
Difficult analytically ✗

Straightforward with machine learning ✔

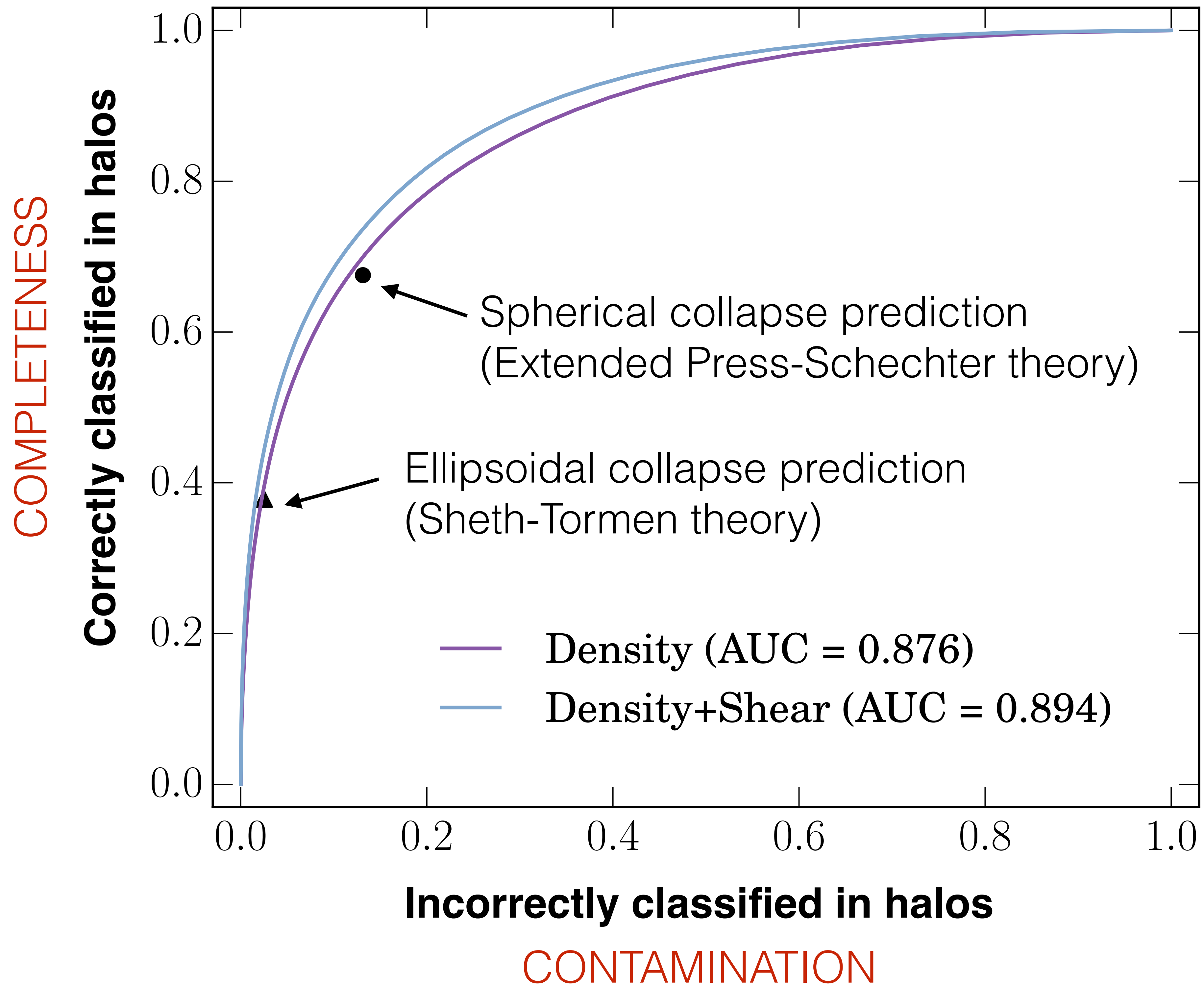
↓

Translate the shear field into new features!

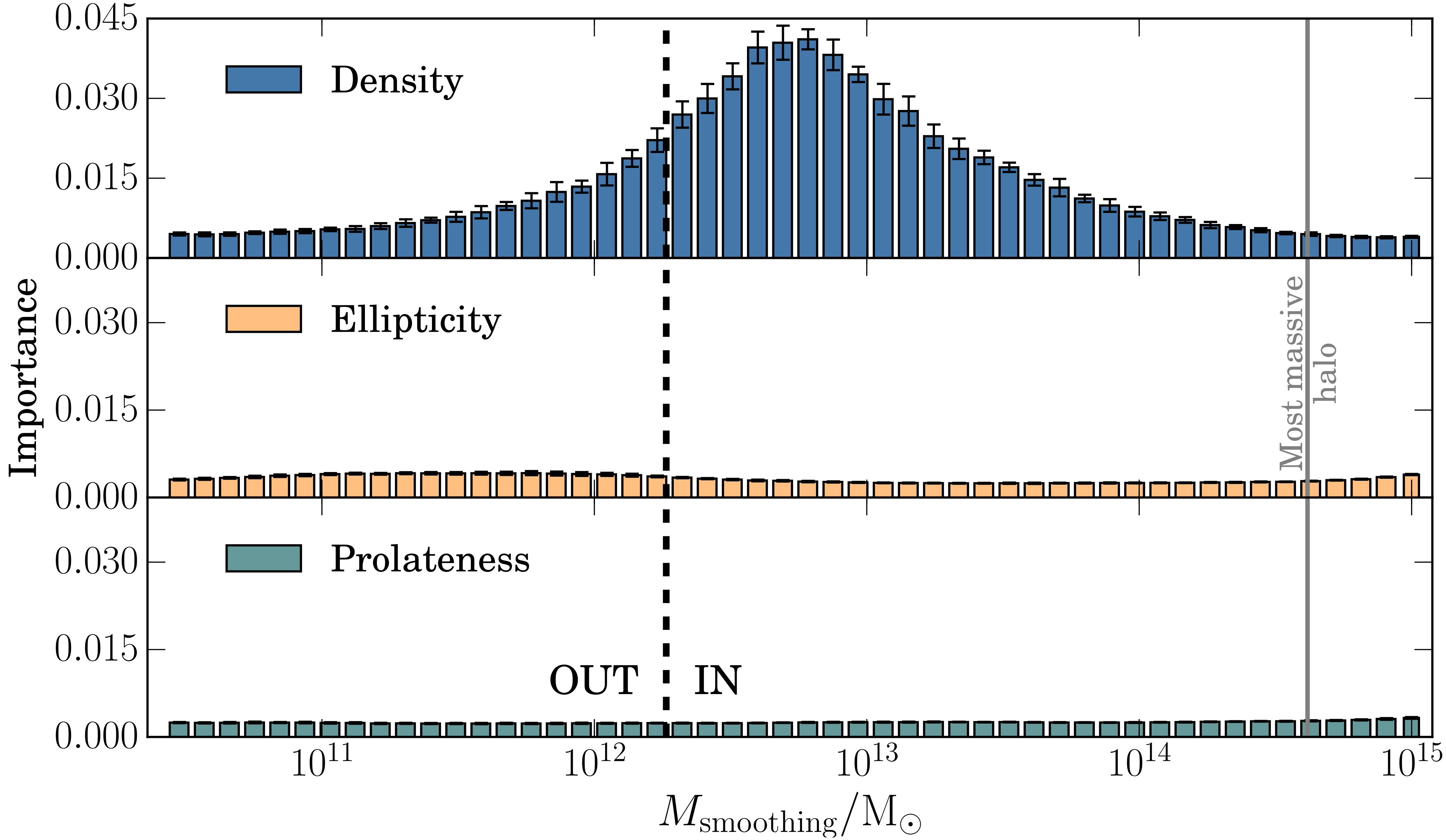
Adding the shear shows little improvement



What is the difference between ST and EPS?



Density + Shear importances



Conclusions & future work

- An interpretable machine learning framework to gain physical insights into cosmological structure formation

- The tidal shear field contains little additional information over that contained in the linear density field to describe the formation of dark matter halos

- At Fermilab, we are extending our framework to deep learning algorithms which do not require extracting "features" from the simulation's initial conditions

Extra Slides

Feature importance

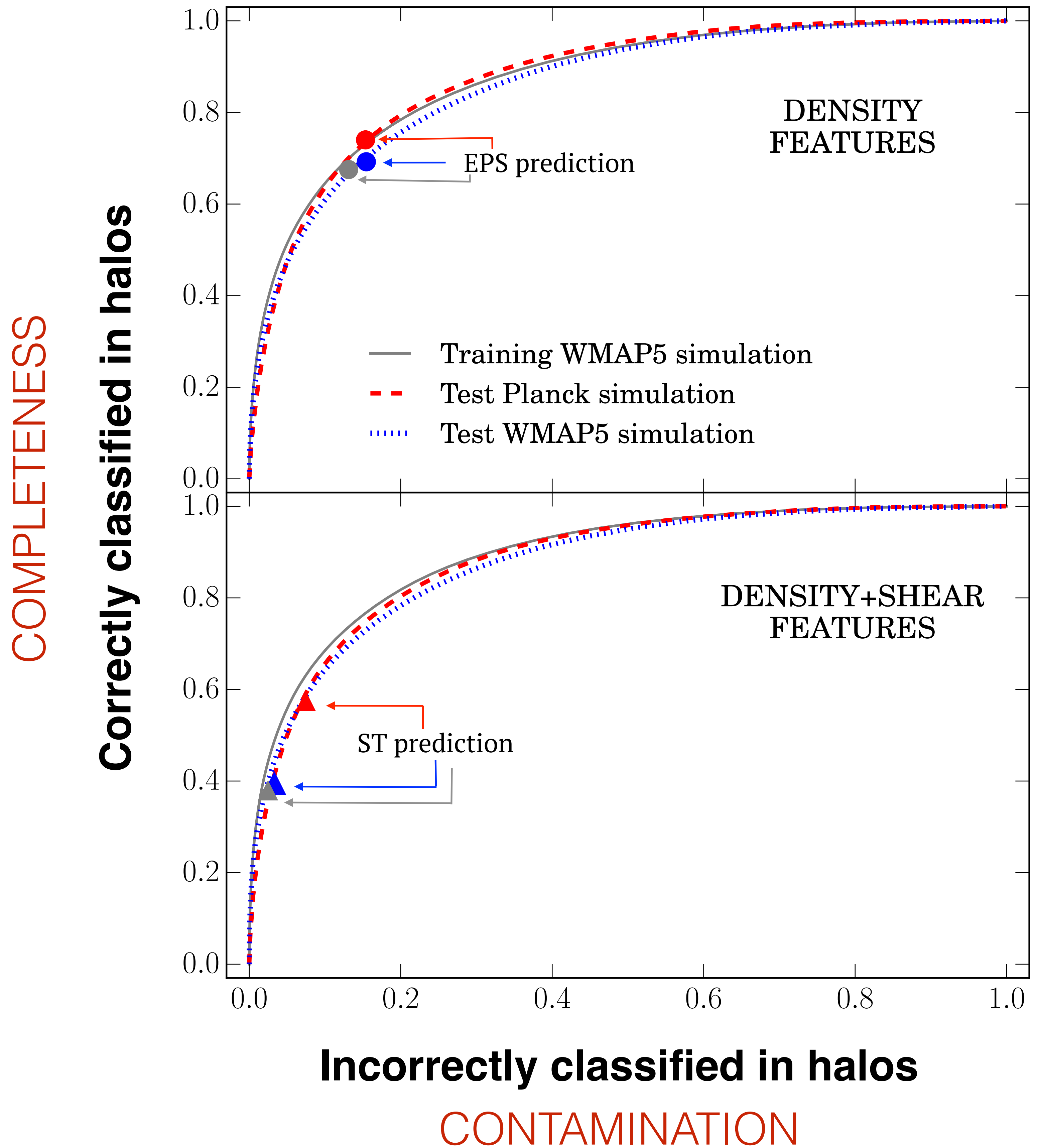
fraction of samples

$$\mathrm{Imp}(X) = \frac{1}{N_T} \sum_{T} \sum_{t \in T : s_t = X} p(t) \Delta i(t)$$

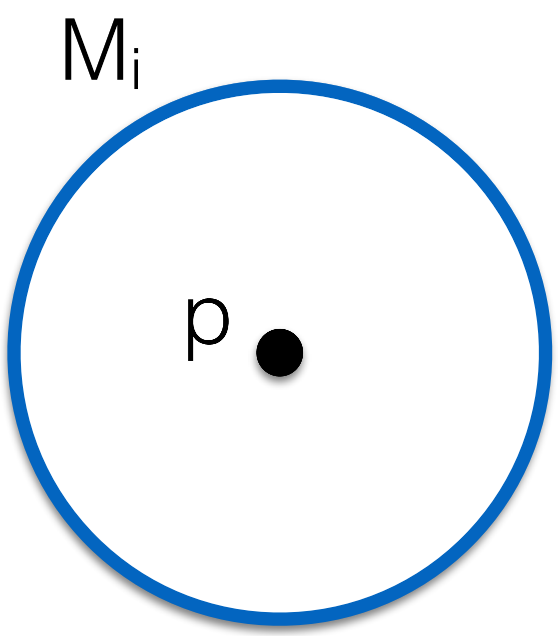
Number of trees

Impurity decrease
(Entropy or Gini impurity)

Test on independent simulations



Density features



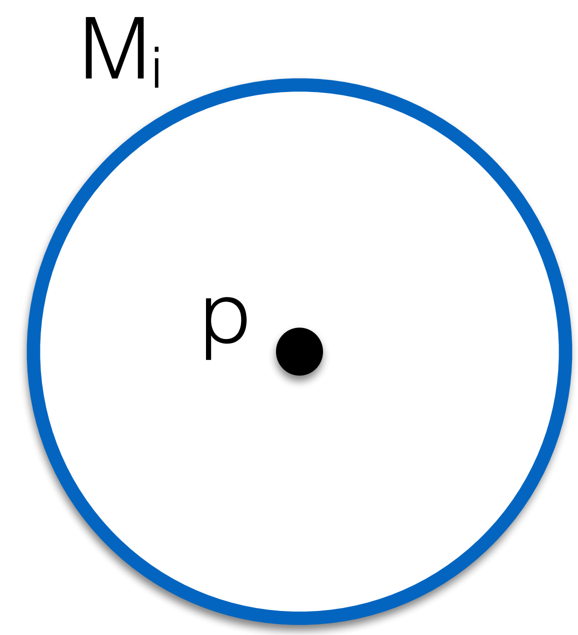
1. Smooth the density field ρ_i with a top-hat window function at mass scale M_i centred on particle p

2. Feature = density contrast,

$$\delta_i = \frac{\rho_i - \bar{\rho}}{\bar{\rho}}$$

Do the same procedure for 50 mass scales

Tidal shear features



1. Smoothed density contrast δ_i at mass scale M_i centred on particle p

2. Solve Poisson's equation $\nabla^2 \Phi_i = \delta_i$

3. The tidal shear tensor

$$T_i^{\alpha\beta} = \frac{\partial^2 \Phi_i}{\partial x^\alpha \partial x^\beta}$$

, with eigenvalues $\lambda_{i,1}$, $\lambda_{i,2}$, $\lambda_{i,3}$

4. Features = two independent linear combinations of the **eigenvalues** (*ellipticity* and *prolateness)*

Tidal shear features

Define $t_{i,j} = \lambda_{i,j} - \delta_i/3$, where λ are the tidal shear eigenvalues.

Two new features per particle at mass scale M_i :

- ***Ellipticity***

$$e_i = 3(t_{i,1} - t_{i,3})$$

- ***Prolateness***

$$p_i = 3(t_{i,1} + t_{i,3})$$

Do the same procedure for 50 mass scales