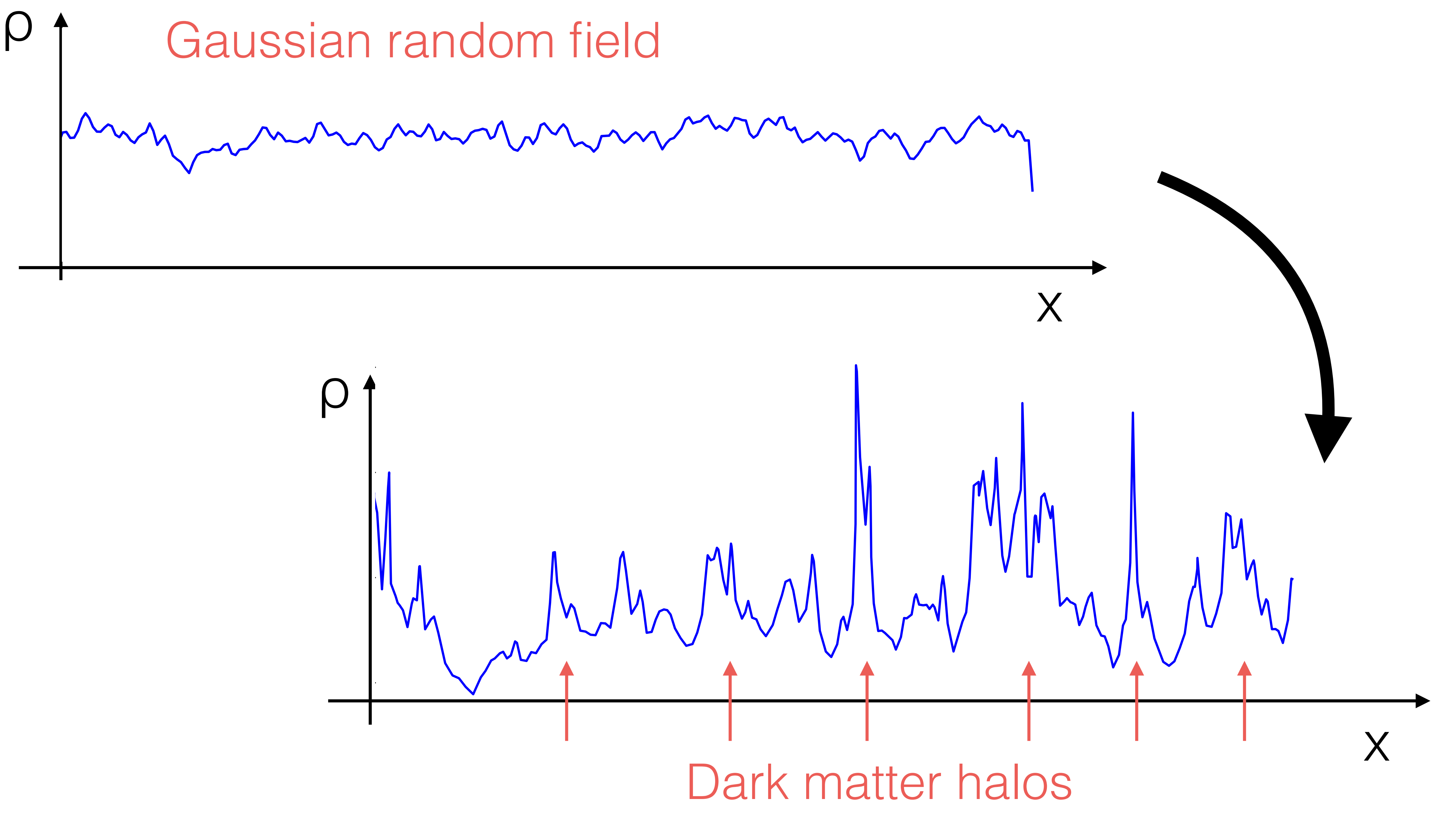
# Machine Learning Cosmological Structure Formation

**Luisa Lucie-Smith**
*University College London*

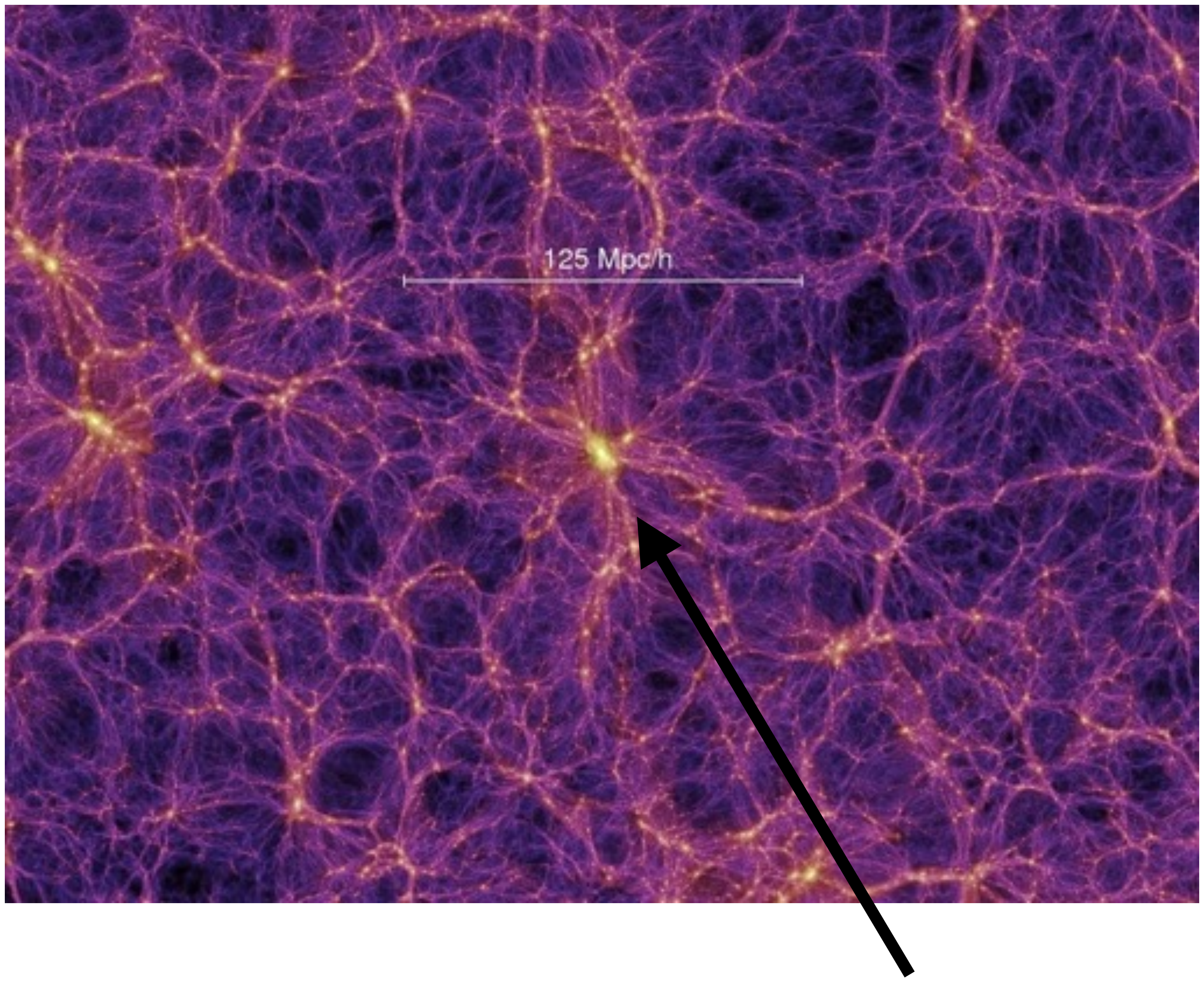
with H.V. Peiris & A. Pontzen

UCL

# The Physics



# N-body simulation




Evolve dark matter (DM) through cosmic time

***slow*** and ***expensive***
+
difficult ***physical*** interpretation

Dark matter halo

*Figure: Springel et al. (2005)*

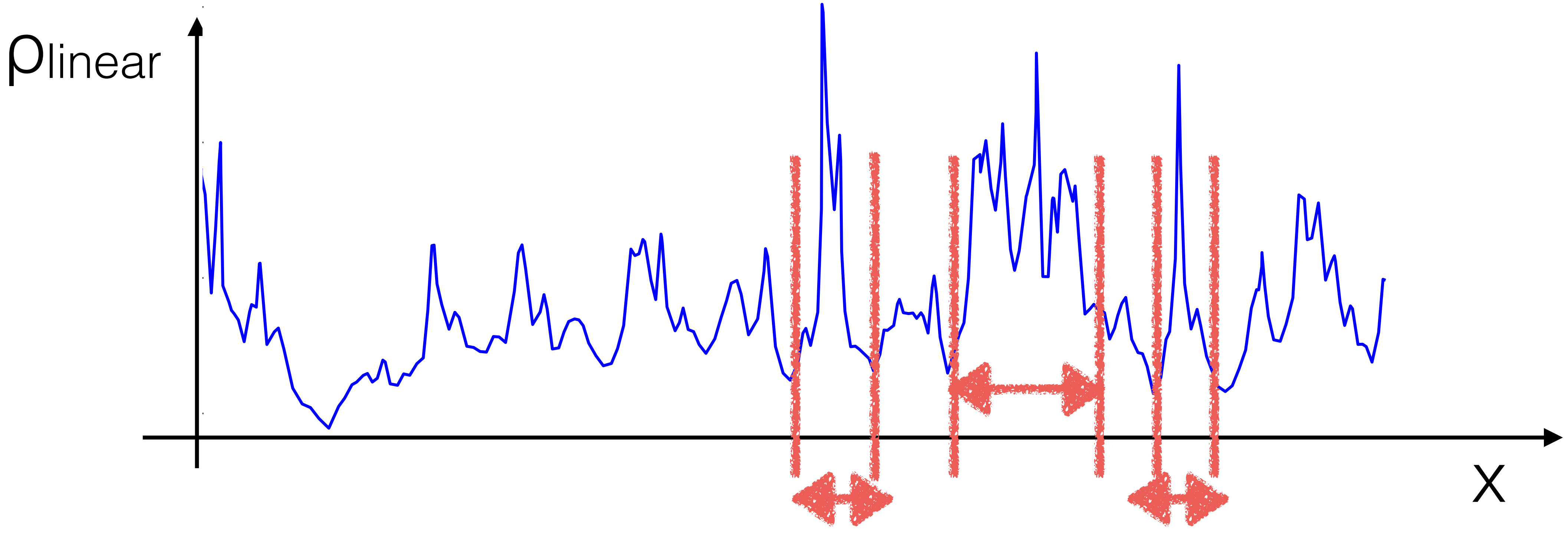
# Outline

1. Build a fast mapping between early Universe and final halo distribution

2. Investigate what physics of early Universe contains relevant information on dark matter halo formation

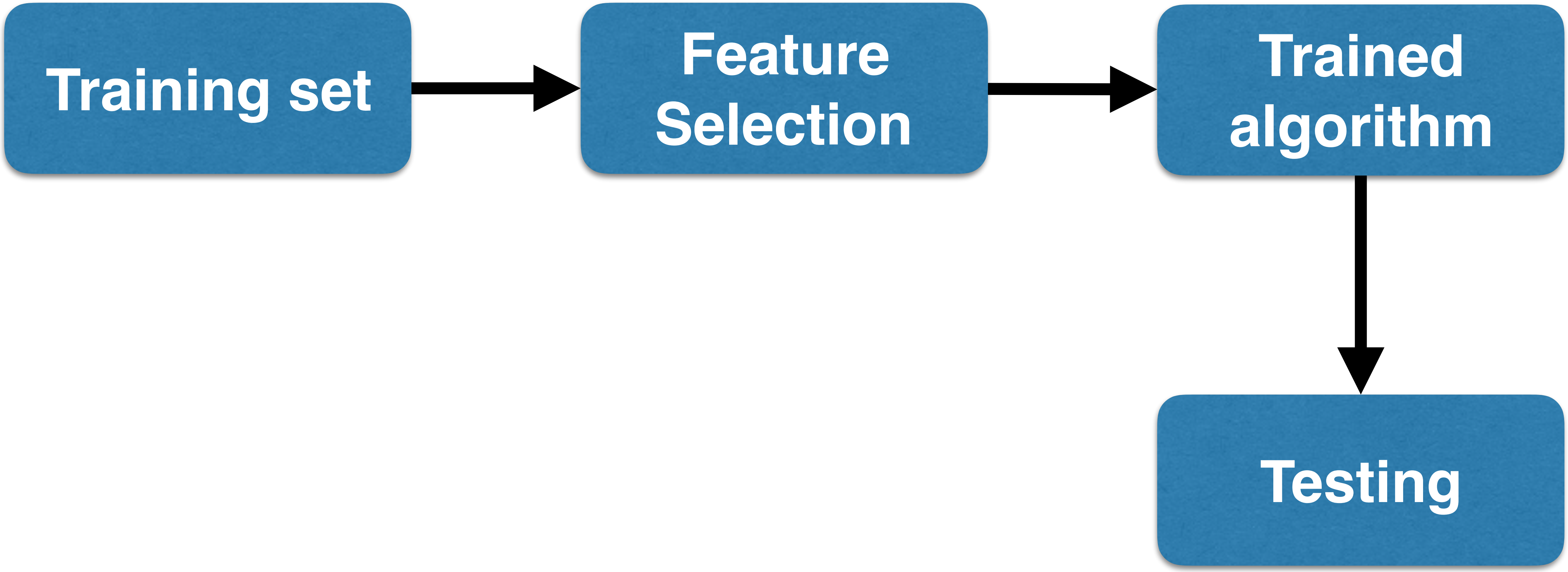
3. How we can go beyond standard spherical collapse

# A machine learning approach




*Can a machine learning algorithm classify whether DM particles in the initial conditions will end up **IN** or **OUT** of halos of mass in range ~$10^{12}\,M_\odot$ - $10^{14}\,M_\odot$ at the end of a simulation?*

# Supervised classification



# Training set: N-body simulation

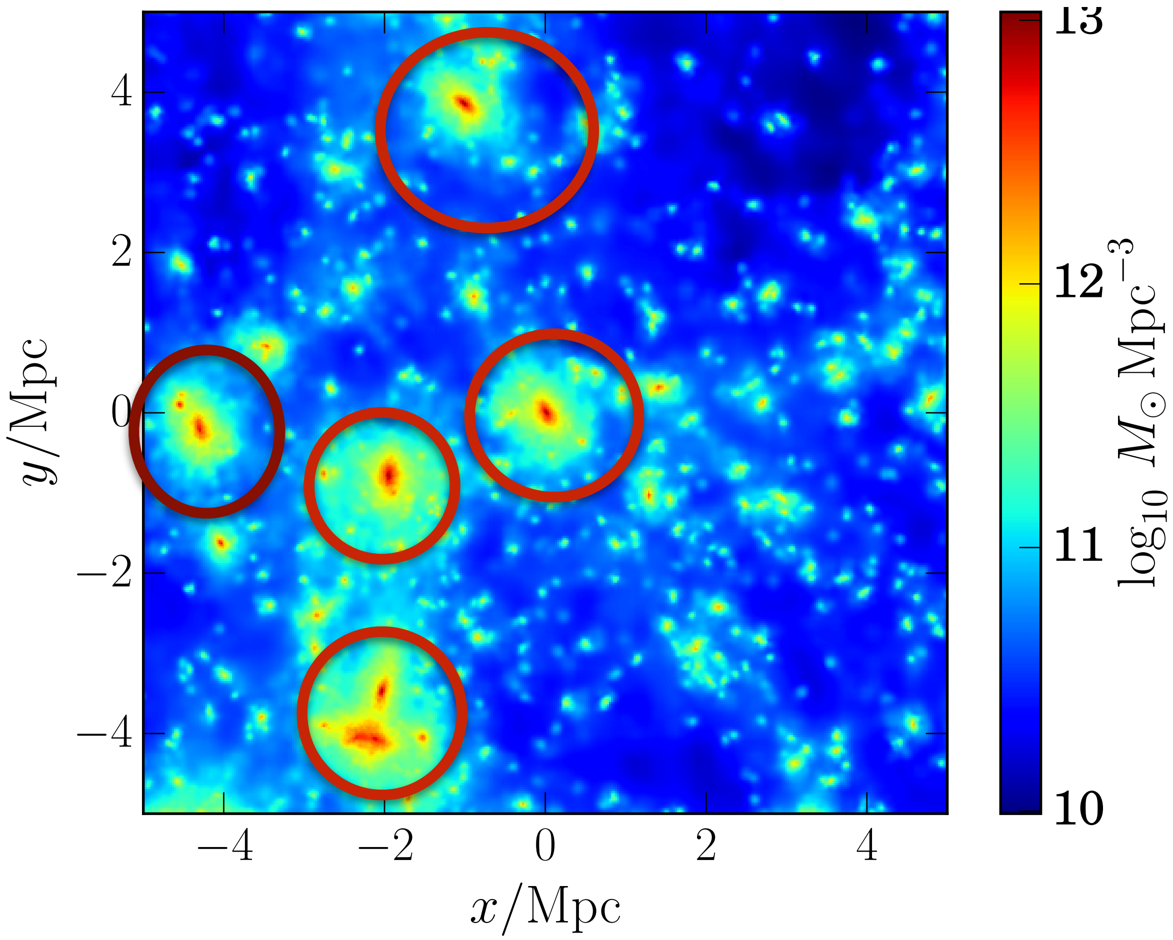
- *Samples*

  A subsample of the simulation's DM particles

- *Class Labels*

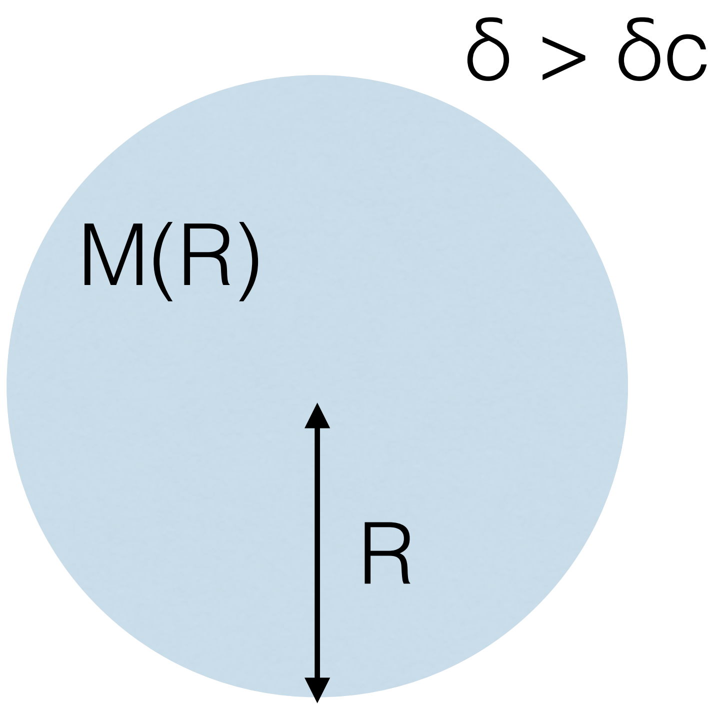
  1. **IN** halos of mass M, s.t. $10^{12}\,M_\odot < M < 10^{14}\,M_\odot$

  2. **OUT**, otherwise.

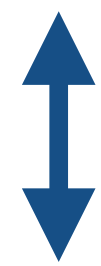


# The density field

***Spherical collapse:***



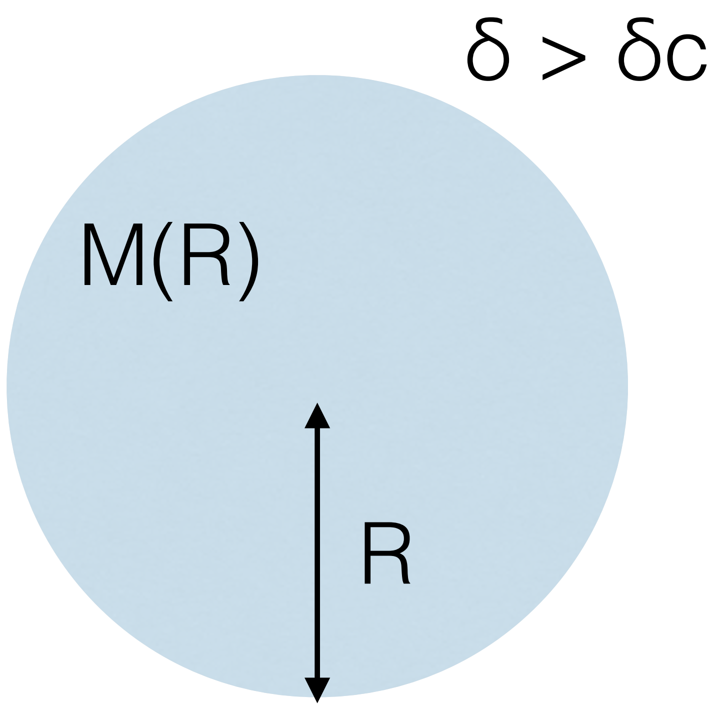

Regions where
density contrast is above some
threshold, $\delta_c$



Dark matter halo of mass M(R)

# The density field

***Spherical collapse:***

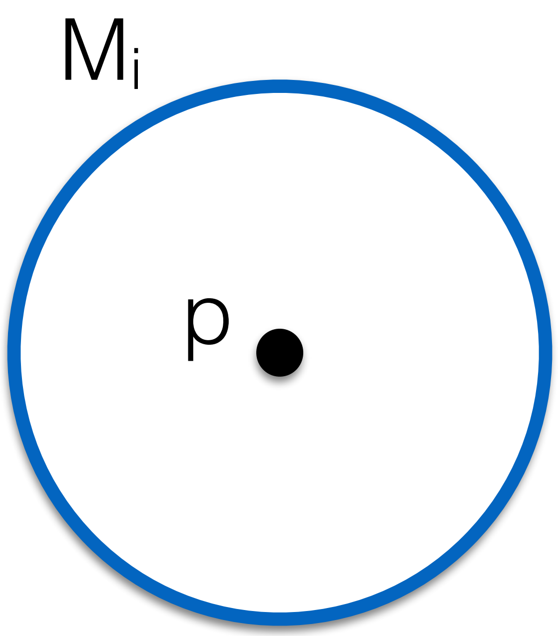



Regions where
density contrast is above some
threshold, $\delta_c$

Dark matter halo of mass M(R)

***Extended Press-Schechter theory:*** *analytic* solution tested against simulations

# Density features



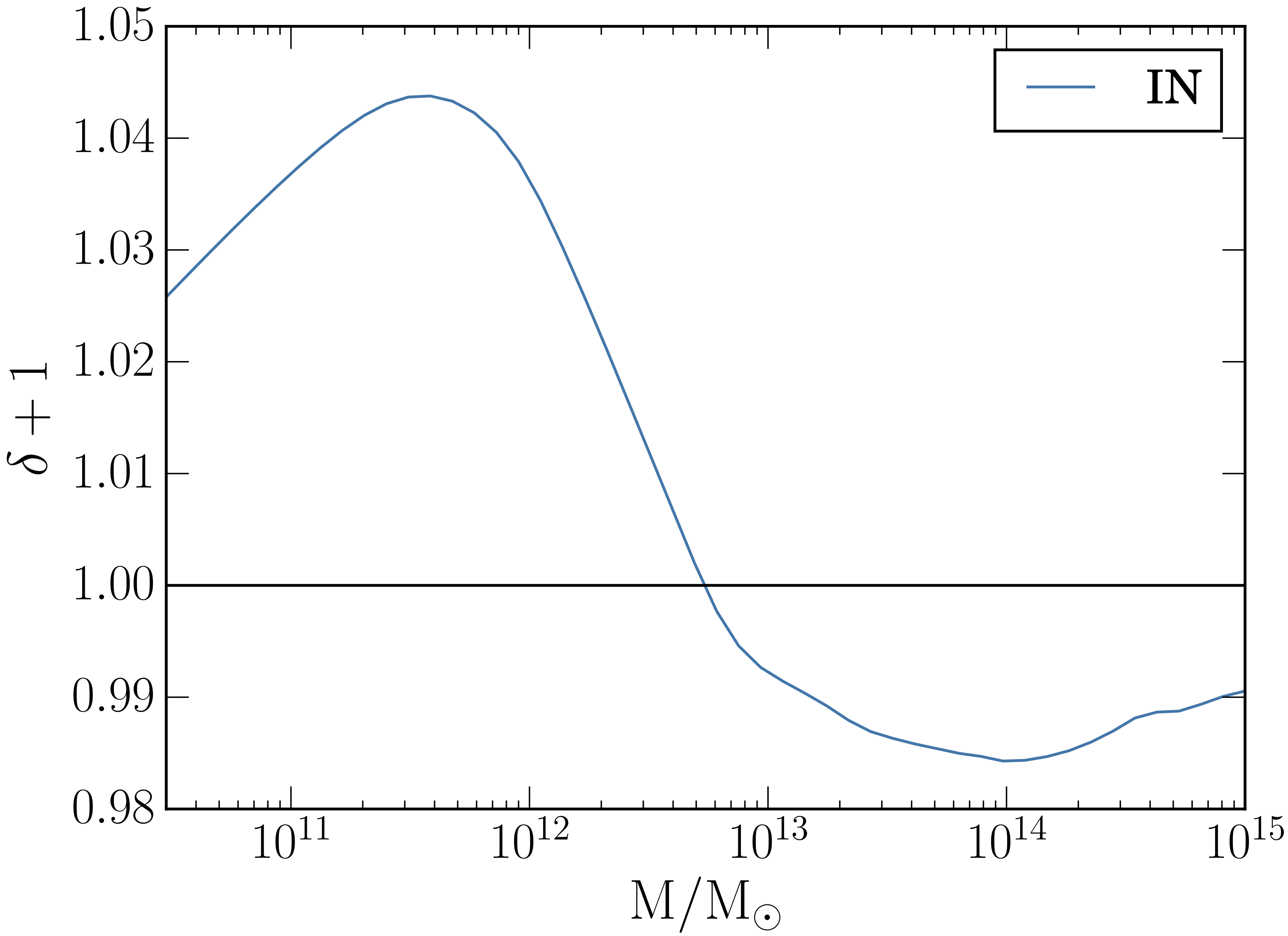

1. Smooth the density field $\rho_i$ with a top-hat window function at mass scale $M_i$ centred on particle p

2. Feature = density contrast,

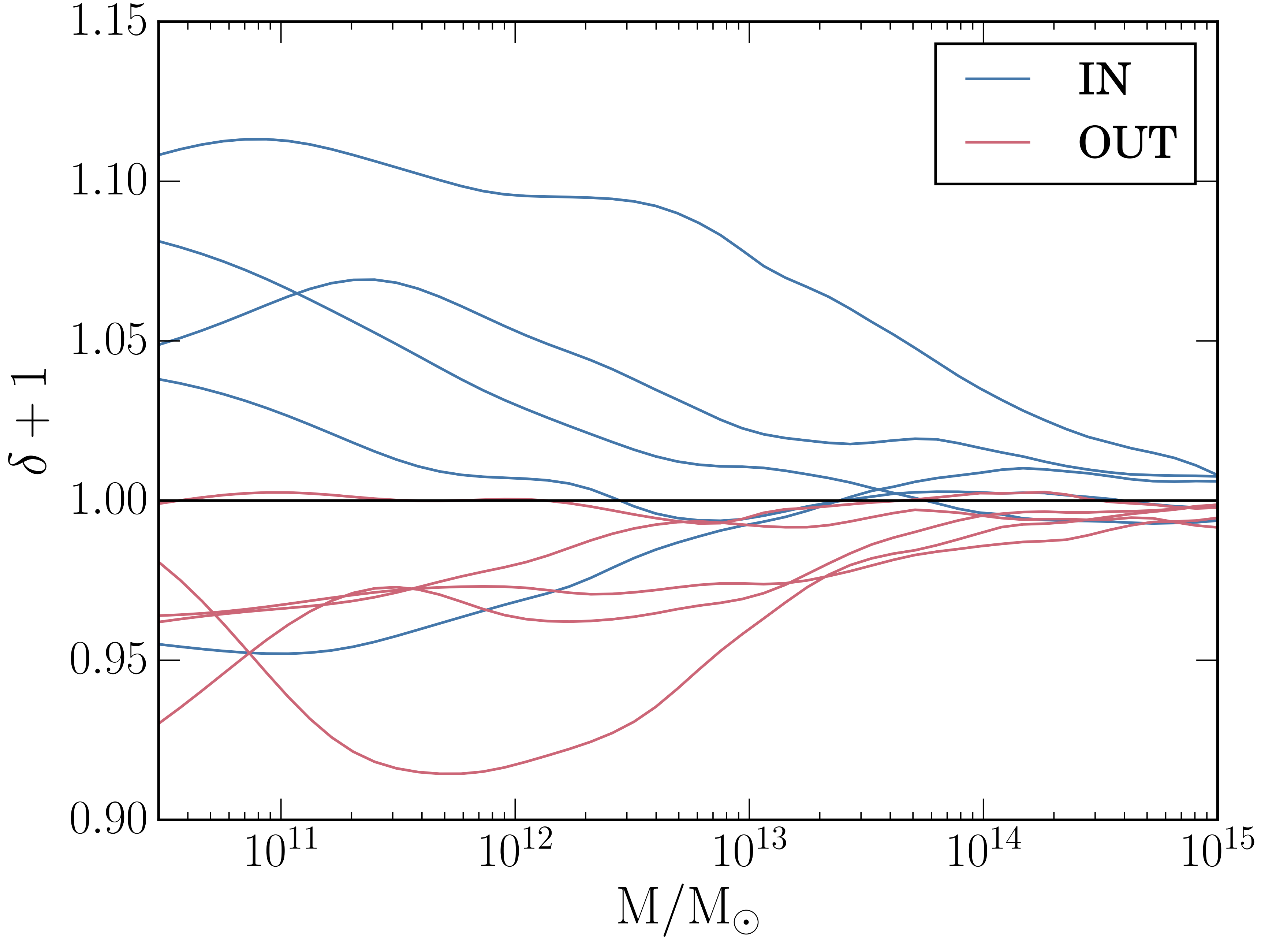
$$\delta_i = \frac{\rho_i - \bar{\rho}}{\bar{\rho}}$$

Do the same procedure for 50 mass scales

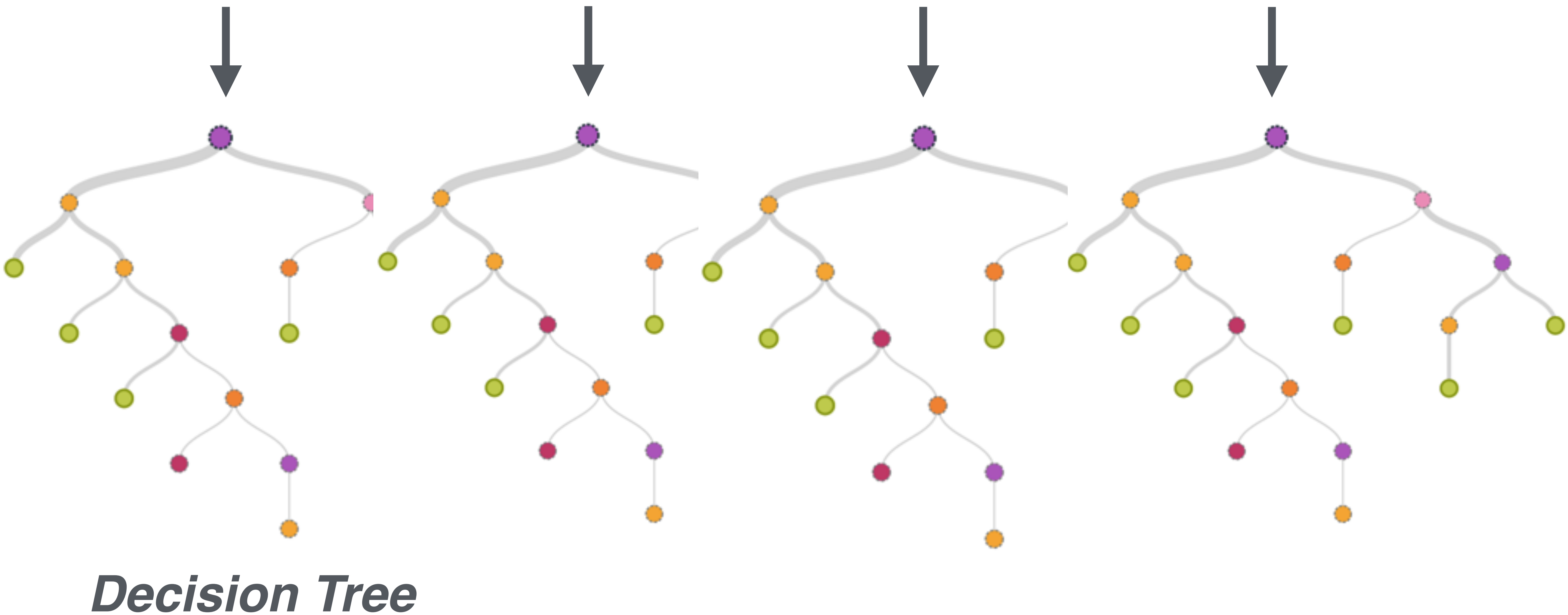
# Density features



# Density features

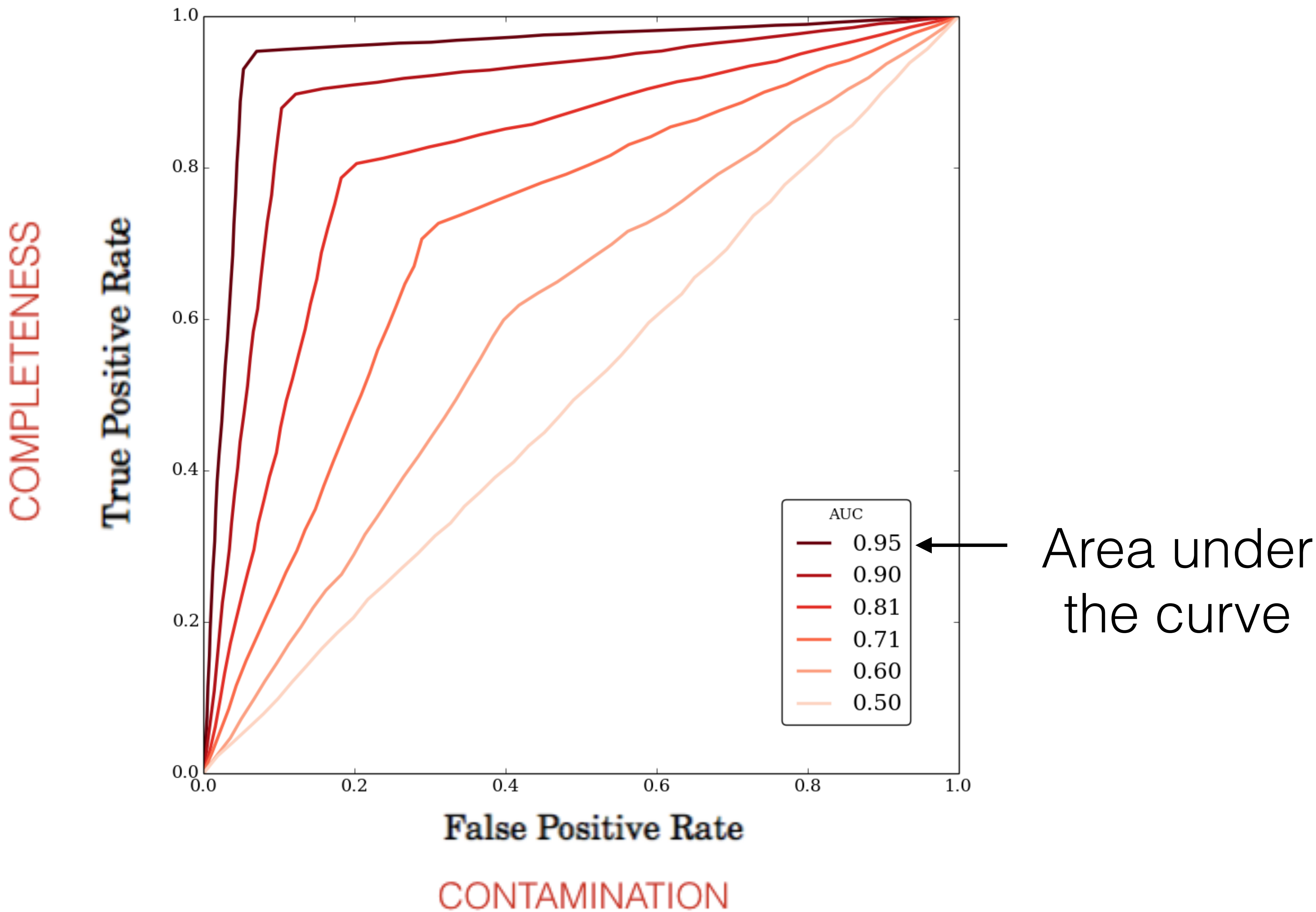


# Random Forests



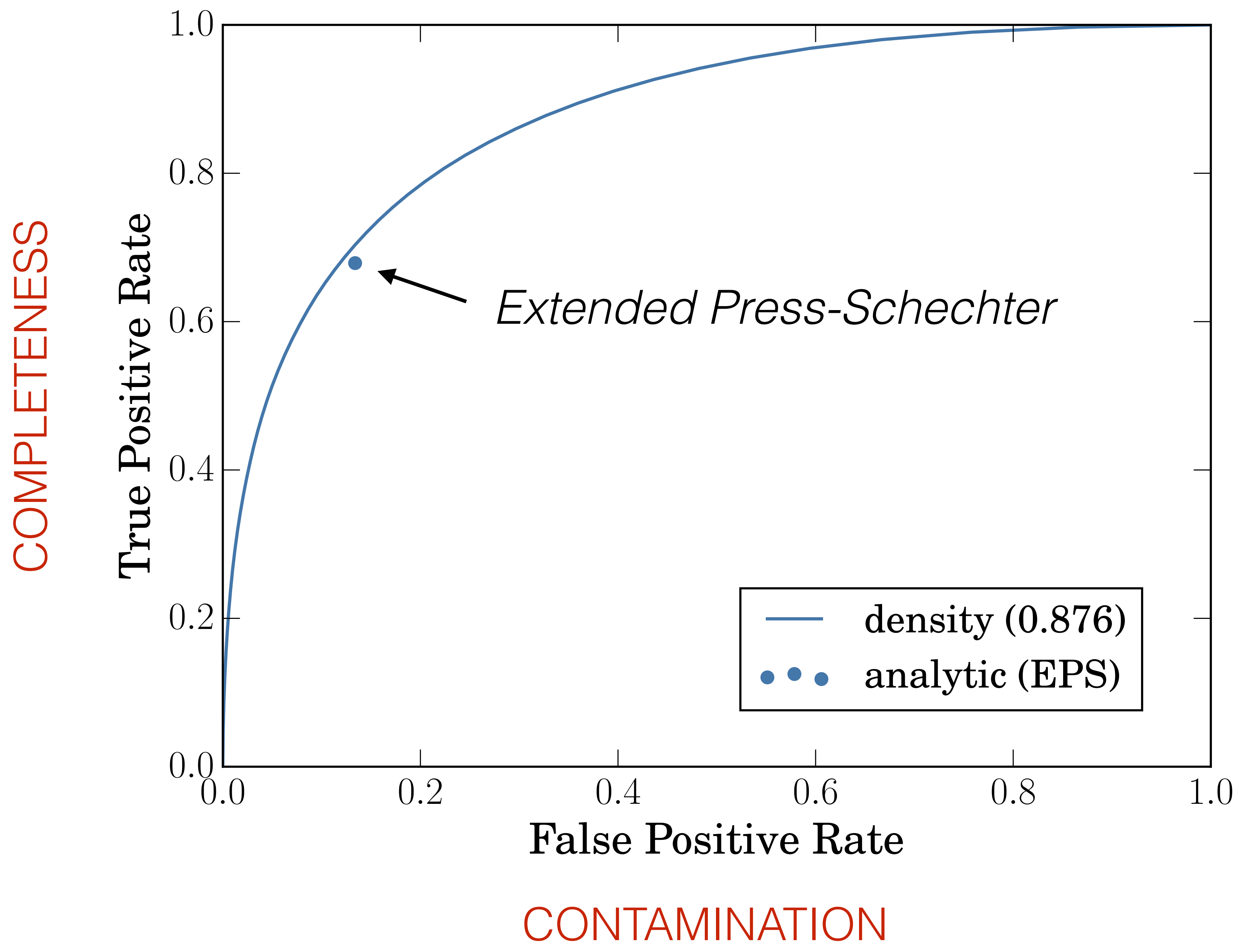

Final prediction =
average probabilistic predictions

# Receiver Operating Characteristic (ROC) curves

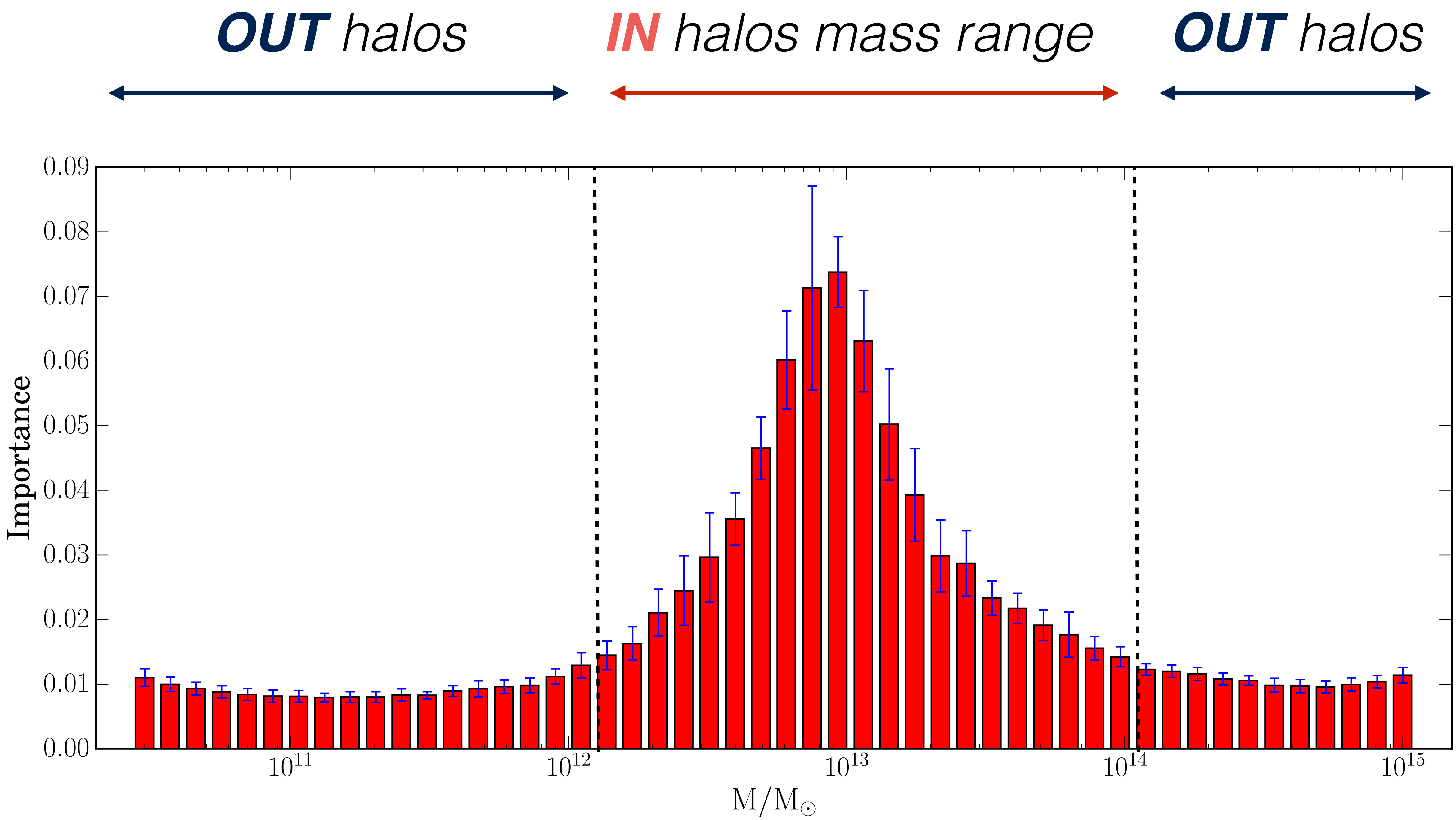



*Credit: Michelle Lochner*

# Machine learning vs extended Press-Schechter



# Density Importances



# Additional physics

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

Difficult analytically ✗

Straightforward with machine learning ✔

↓

Translate the shear field into new features!

# The tidal shear

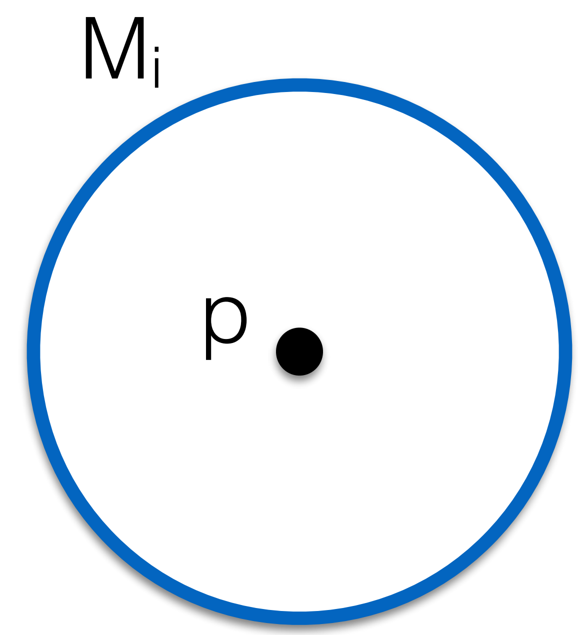
1. Smoothed density contrast $\delta_i$ at mass scale $M_i$ centred on particle p

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

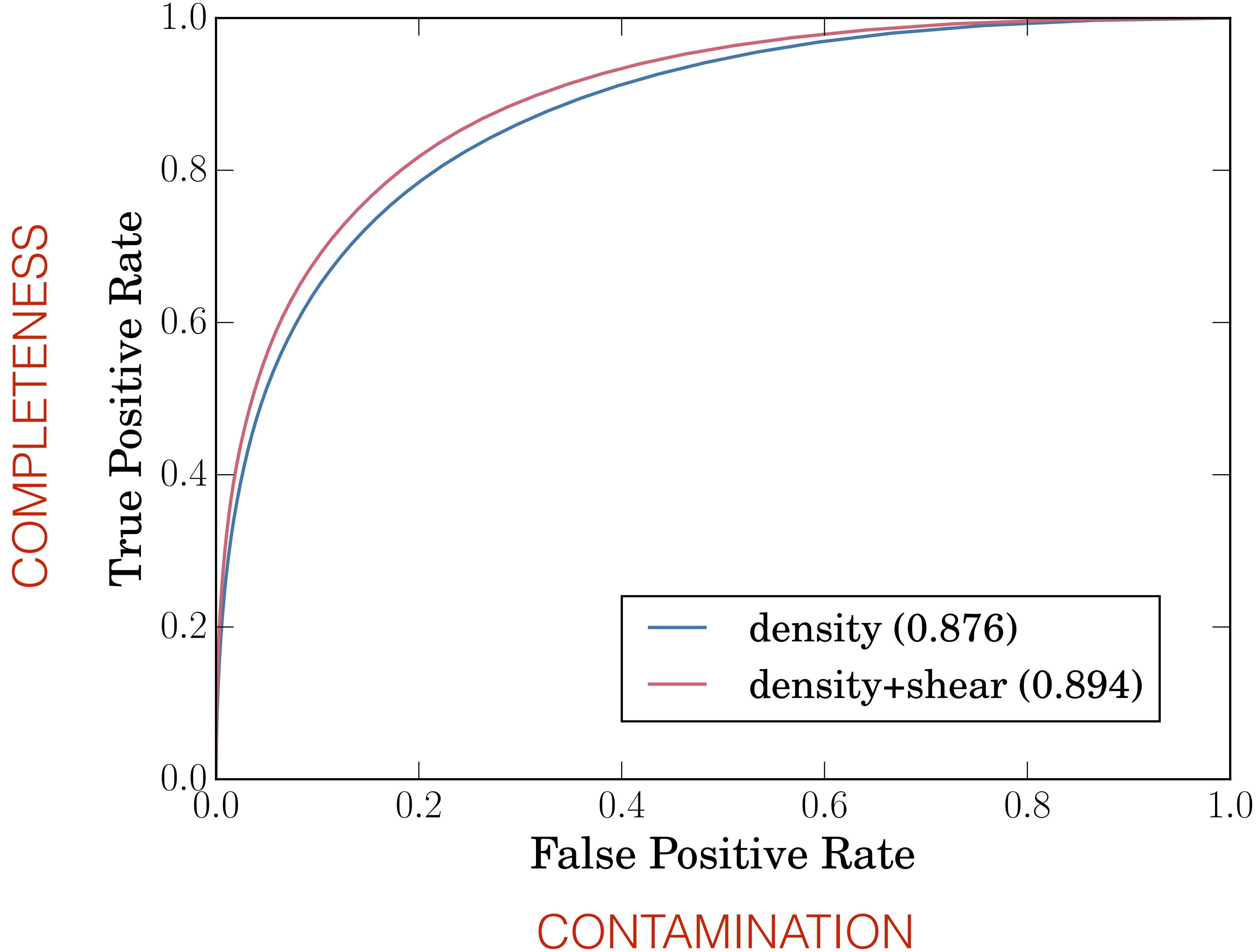
3. The tidal shear tensor

$$T_i^{\alpha\beta} = \frac{\partial^2 \Phi_i}{\partial x^\alpha \partial x^\beta}$$ , with eigenvalues $\lambda_{i,1}$ , $\lambda_{i,2}$ , $\lambda_{i,3}$

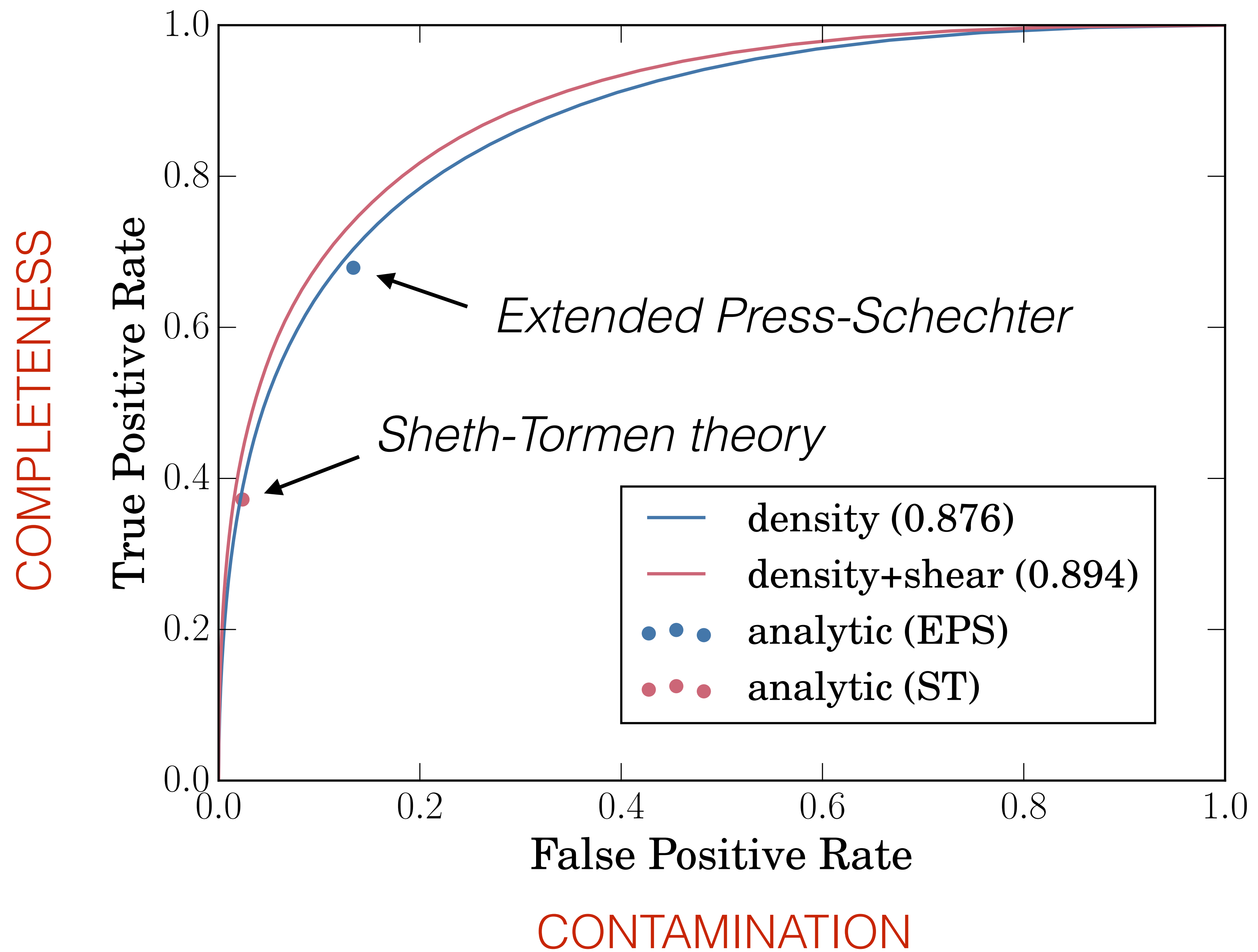
4. Features = two independent linear combinations of the **eigenvalues** (*ellipticity* and *prolateness)*



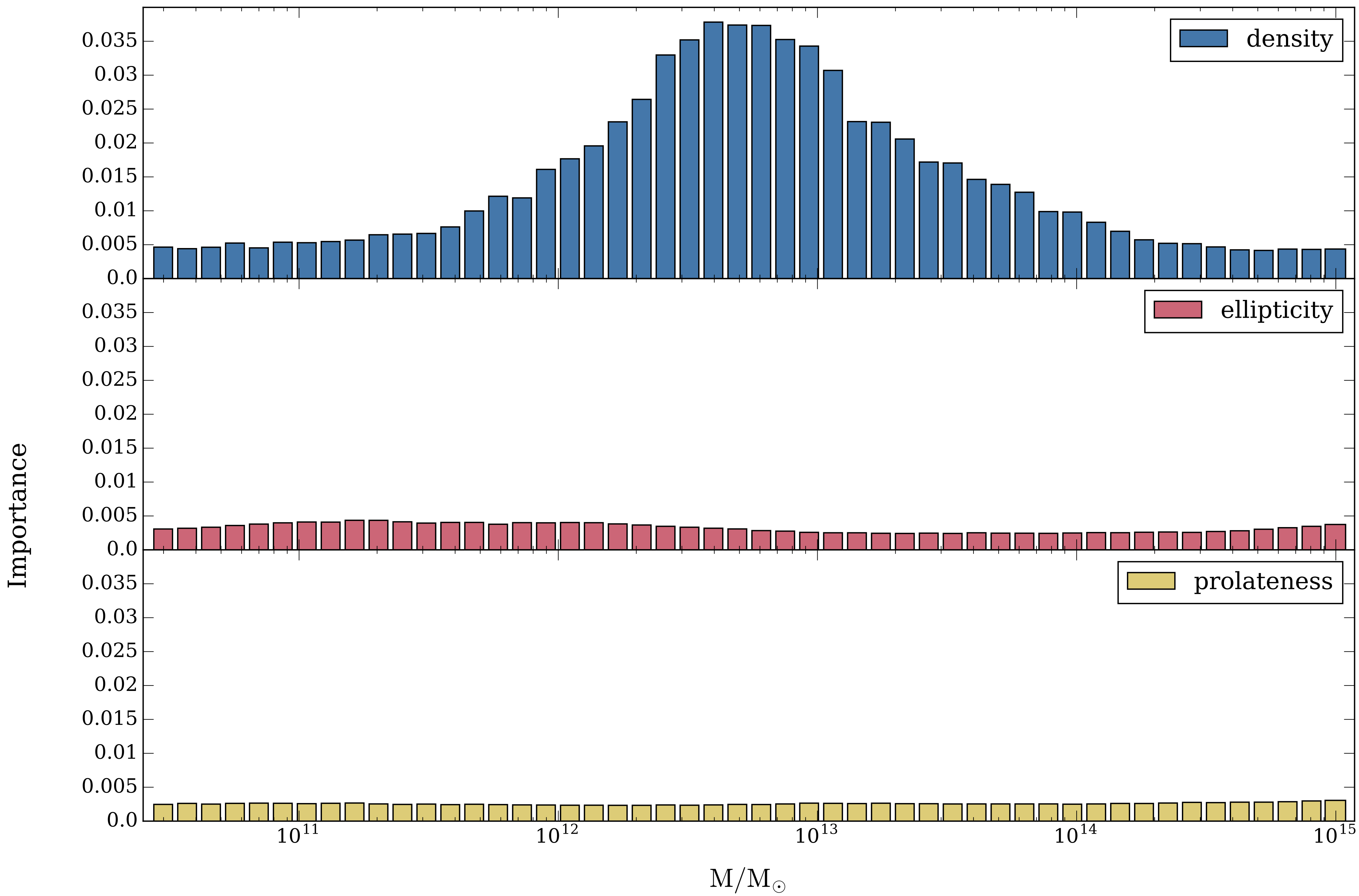
# Adding the shear shows little improvement



# What is the difference between ST and EPS?



# Density + Shear importances



# Conclusions

- Achieve comparable predictions to extended Press-Schechter and Sheth-Tormen theory
- Importance ranking shows which information improves predictions or not
- Incorporating extra physical information beyond spherical collapse should allow better understanding of link between linear and non-linear universe