Using Cosmic Voids to Illuminate Galaxy Properties

Christina Kreisch
Department of Astrophysical Sciences

Abstract

- Aim: predict median halo mass from only void properties
- There is a complex, non-linear relationship between void properties and the median halo mass
- Random Forest regression outperforms all linear models, with a mean squared error $\mathcal{O}(10^{-6})$
- Relationship between the median halo mass and void features is more complex than 2nd order polynomial or power law

Motivation & Previous Work

- Directly measuring galaxy features can be difficult
  - Seek alternative method
- As large underdone regions that house few halos, voids provide a different environment for halos
  - Impact galaxy evolution
- Void properties are sensitive to their tracers (Kreisch et al. 2018, Pollina et al. 2016)
  - Should be relationship between halo properties and void properties
- Active work on relationship, such as halo properties as a function of distance to center of void (Habouzit et al. in prep.)

Data & Methods

- Data from IllustrisTNG simulation (Springel et al. 2018)
  - Redshift z=0
  - Boxlength 300 Mpc
  - N-body simulation + hydrodynamics
  - Includes only halo properties
- Obtain void catalog by running VIDE (Sutter et al. 2015) on halo catalog
  - Finds voids from the halo distribution
  - Outputs void features: radius, ellipticity, etc.
- Regression with:
  - Linear Regression
  - Bayesian Ridge Regression, features selected from Elastic Net
  - Random Forest Regression
- Standardize all parameters before regression

Probing Halo Properties

PCA 0: EVR = 0.2940

	Component	Weight
7	GroupWindMass	0.361959
6	GroupStarMetallicity	0.342809
5	GroupSFR	0.334031
15	hasWind	0.317440
1	GroupStellarMass	0.299451
14	hasSF	0.299415
0	GroupBHMass	0.293638
4	GroupMass	0.286832
13	hasBH	0.285080
2	GroupBHMdot	0.231045
12	GroupStellarMassFraction	0.227405
3	GroupGasMetallicity	0.070208
11	Vpec	-0.003480
9	VY	-0.000197
8	VX	0.000161
10	VZ	0.000130

PCA 1: EVR = 0.1701

	Component	Weight
4	GroupMass	0.448101
0	GroupBHMass	0.444417
1	GroupStellarMass	0.439940
14	hasSF	-0.320800
13	hasBH	-0.314541
15	hasWind	-0.285350
6	GroupStarMetallicity	-0.258725
12	GroupStellarMassFraction	-0.200031
2	GroupBHMdot	0.076224
5	GroupSFR	0.071476
3	GroupGasMetallicity	-0.069487
7	GroupWindMass	-0.046647
11	Vpec	0.001916
10	VZ	-0.000248
8	VX	-0.000026
9	VY	0.000021

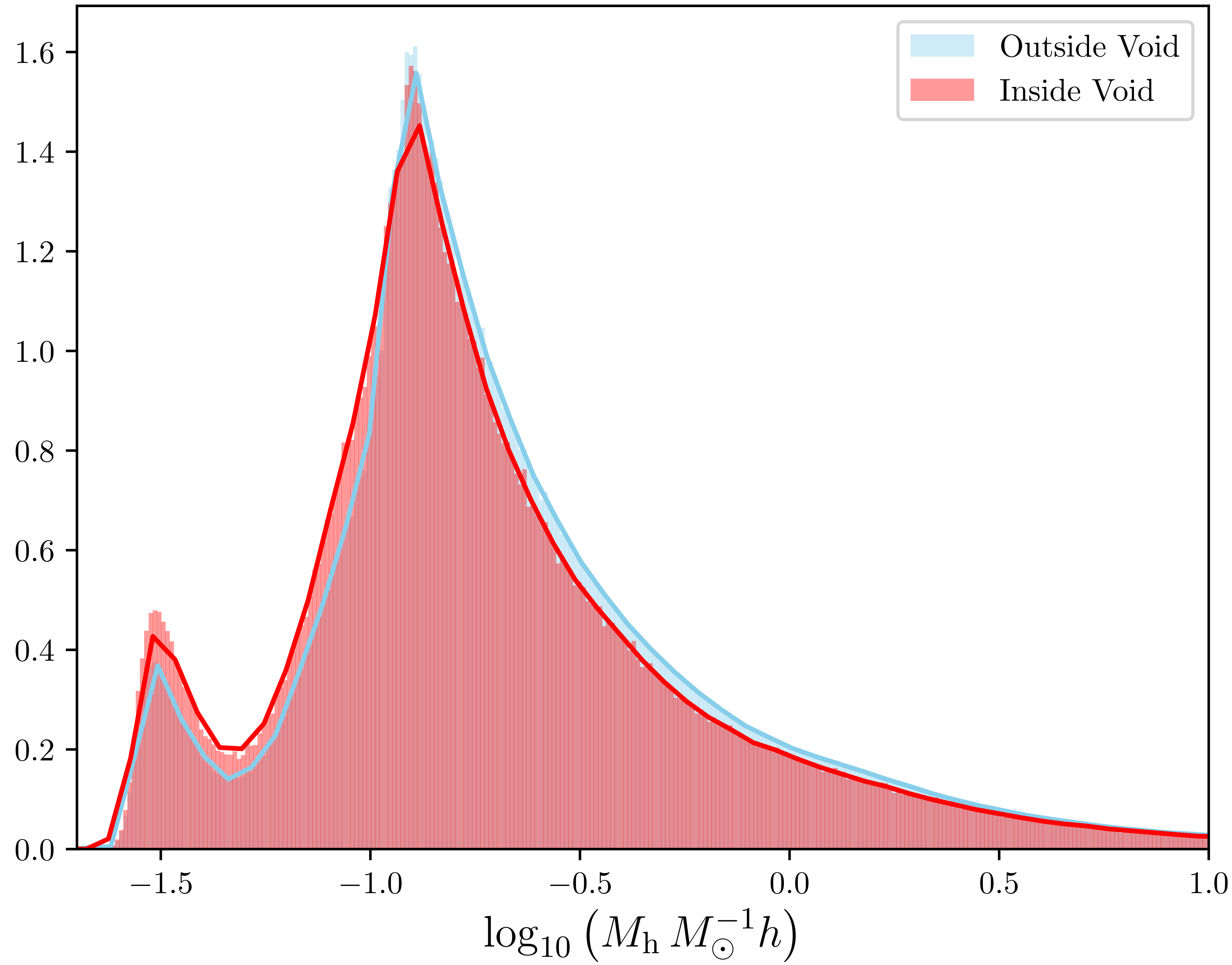
PCA 2: EVR = 0.0702

	Component	Weight
2	GroupBHMdot	-0.570992
5	GroupSFR	-0.458801
13	hasBH	0.320360
7	GroupWindMass	-0.308943
14	hasSF	0.284231
4	GroupMass	0.225279
0	GroupBHMass	0.210482
1	GroupStellarMass	0.198500
11	Vpec	-0.145379
3	GroupGasMetallicity	-0.133183
15	hasWind	0.071248
10	VZ	0.053215
6	GroupStarMetallicity	0.042378
12	GroupStellarMassFraction	0.028162
8	VX	-0.027290
9	VY	-0.005933

PCA 3: EVR = 0.0684

	Component	Weight
11	Vpec	0.701499
3	GroupGasMetallicity	0.544937
10	VZ	-0.340744
8	VX	0.145931
2	GroupBHMdot	-0.131348
5	GroupSFR	-0.115315
7	GroupWindMass	-0.113043
12	GroupStellarMassFraction	0.109012
9	VY	0.074093
0	GroupBHMass	0.053906
4	GroupMass	0.053527
1	GroupStellarMass	0.051693
6	GroupStarMetallicity	0.051074
15	hasWind	-0.040855
14	hasSF	-0.010210
13	hasBH	0.003740

- 4 components produced PCs with the most physical sense
- 1st PC: stars and star formation activity in galaxy
- 2nd PC: mass of galaxies (frequently considered the most important galaxy feature)
- 3rd PC: black hole activity
- 4th PC: galaxy motion + metallicity
- Striking: 1st + 2nd PCs agree with Connolly et al. 1995 that used galaxy spectra
- Start with the most fundamental galaxy feature: galaxy (halo) mass



- Galaxies inside voids tend to be less massive than galaxies outside voids → seems promising void features have impact on mass
- Many galaxies within a single void → aim to predict population parameters
- Goal: predict median galaxy mass of galaxy population within void

Predicting Median Halo Mass

—Linear Void Features—

Feature	LR	BR + EN	RF
voidDensityContrast	-0.288656	—	0.402250
voidEllipticity	-0.167343	-0.109672 ± 0.001221	0.342885
voidRadius	-0.005088	0.027225 ± 0.001610	0.244195
voidCentralDen	-0.002654	-0.016113 ± 0.001333	0.007460
voidNumChildren	0.044095	0.010645 ± 0.001566	0.003210

Table 3: Feature weights for linear features

- Void Features:
  - Density Contrast: contrast between inner and outer densities
  - Ellipticity
  - Radius: average size of void
  - Number of Children: number of sub voids
  - Central Density: density within 1/4 radius of center
- See Table 1: Cannot predict individual mass as well
  - Makes sense: large scatter in halo mass for halo pop. in void
  - Justifies prediction of median halo mass
- See Table 1: Random Forest is astonishingly accurate

Why is the Random Forest so Accurate?

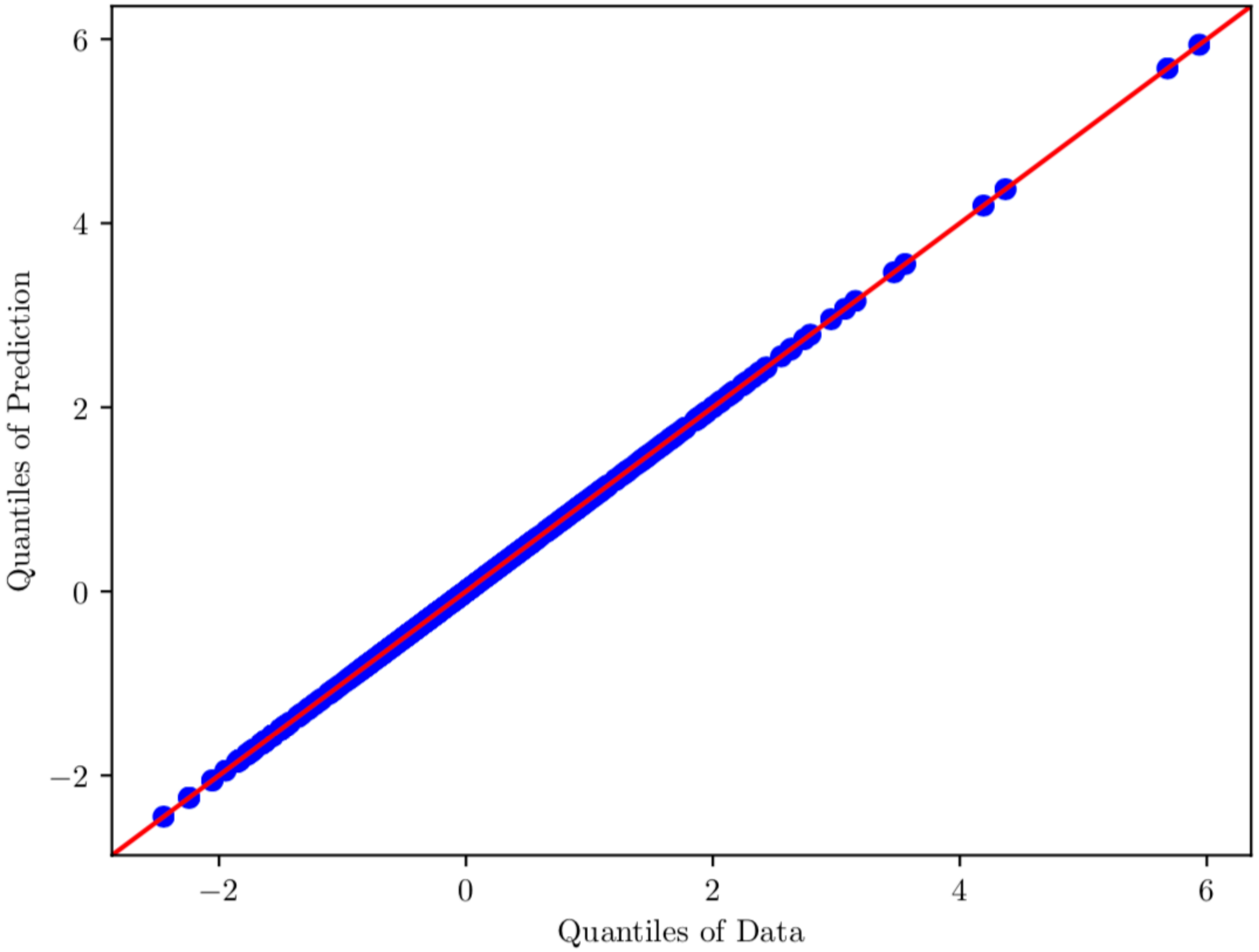


Figure 10: RF linear

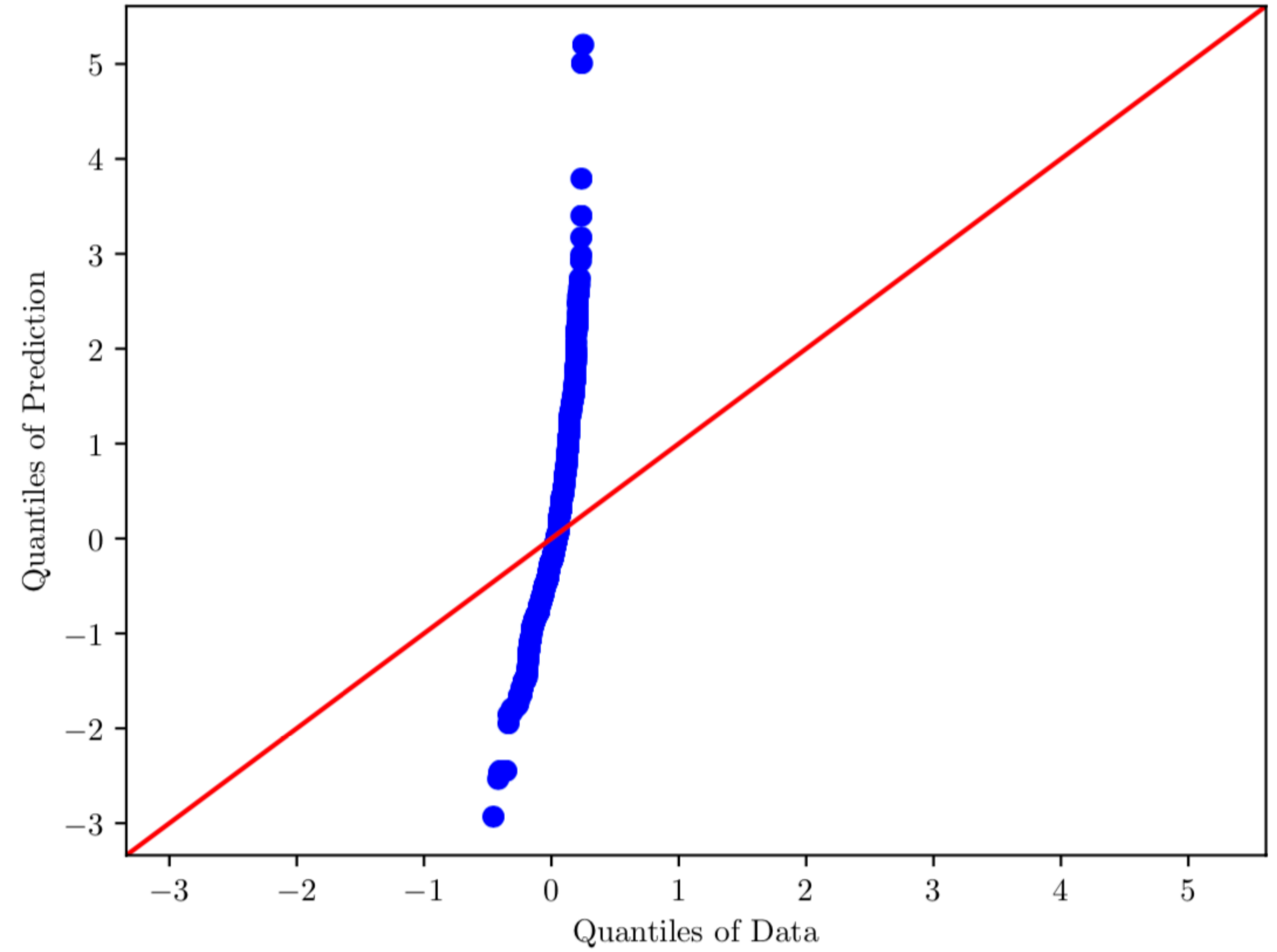
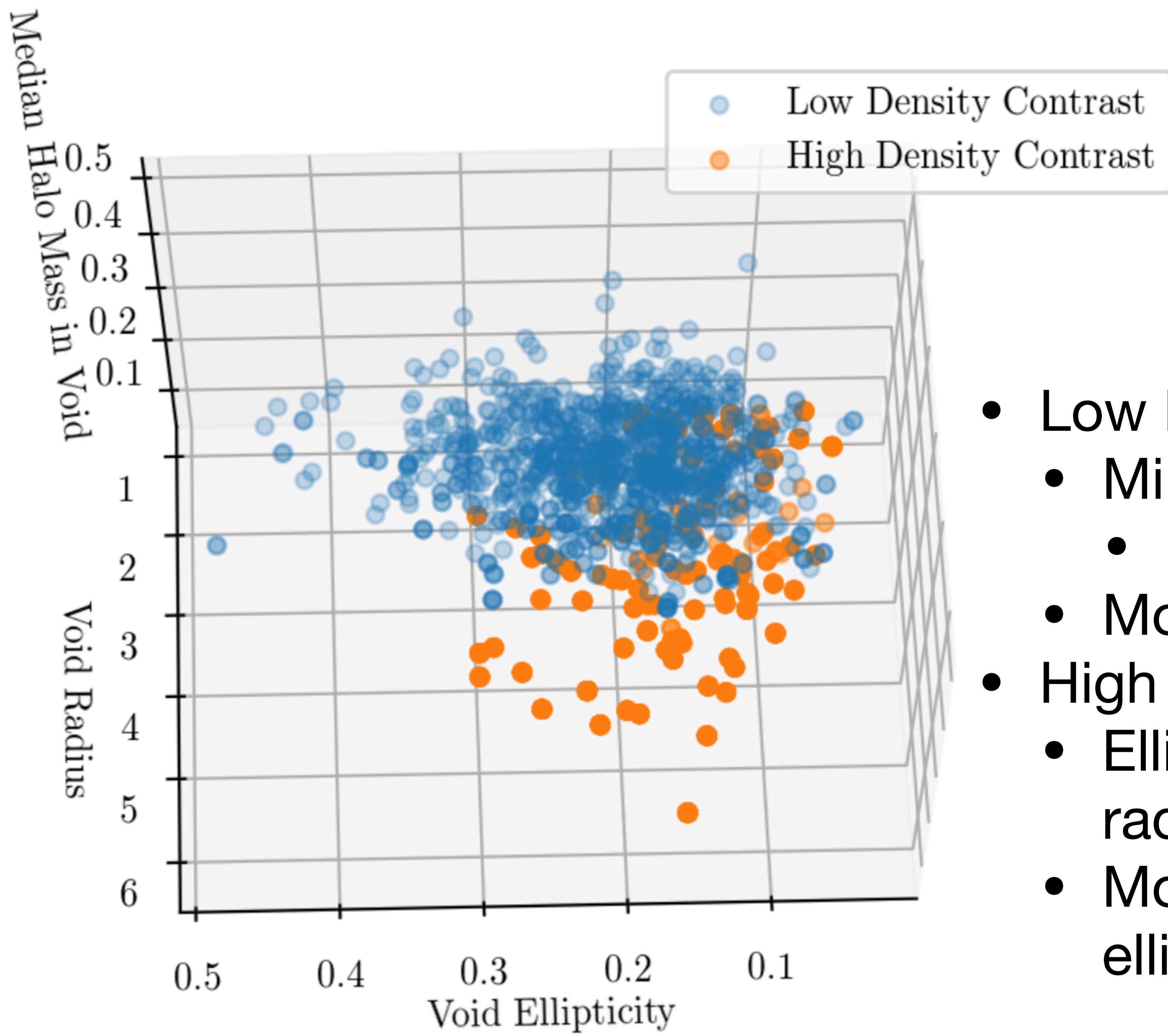


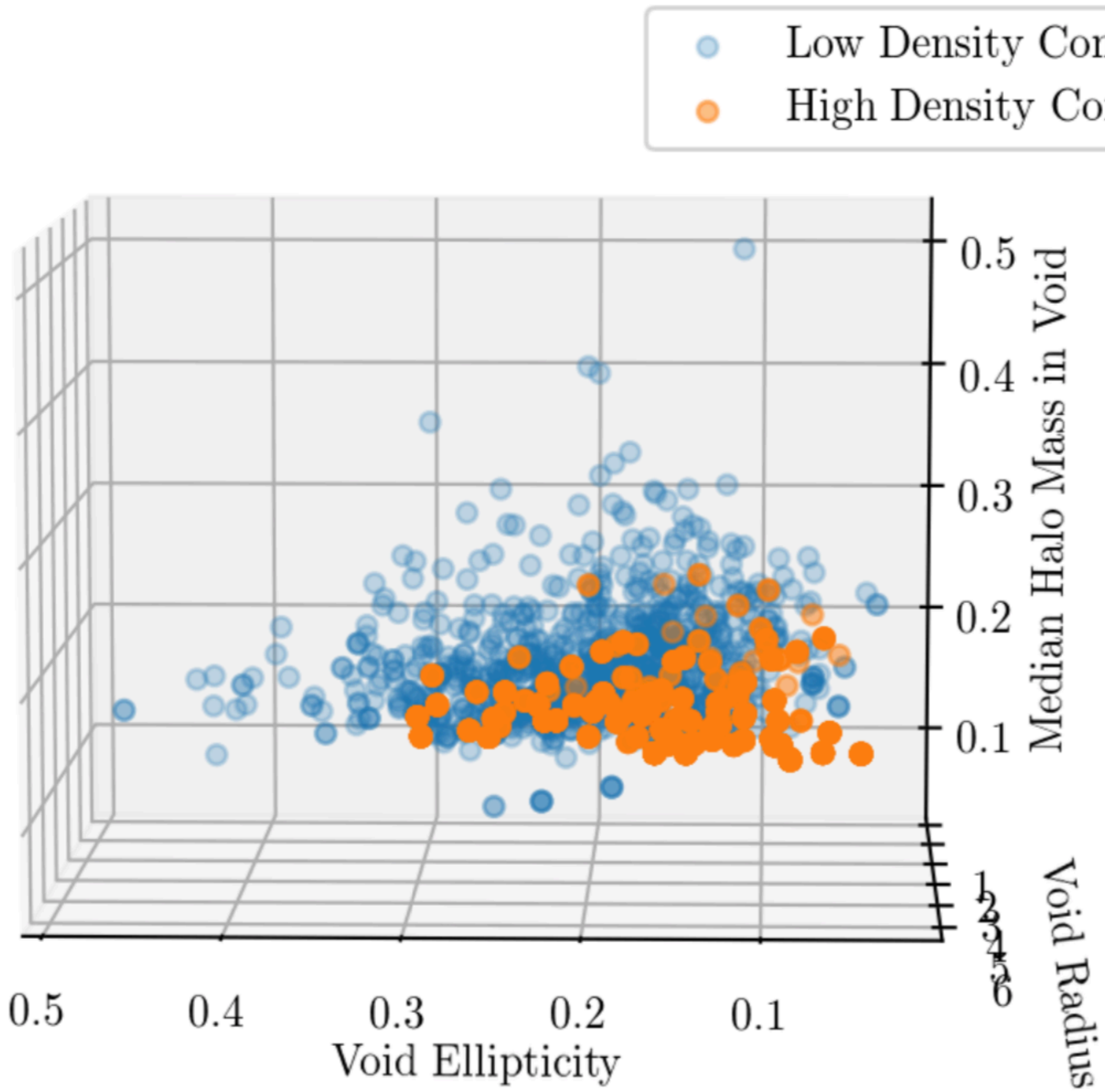
Figure 11: BR + EN linear qq

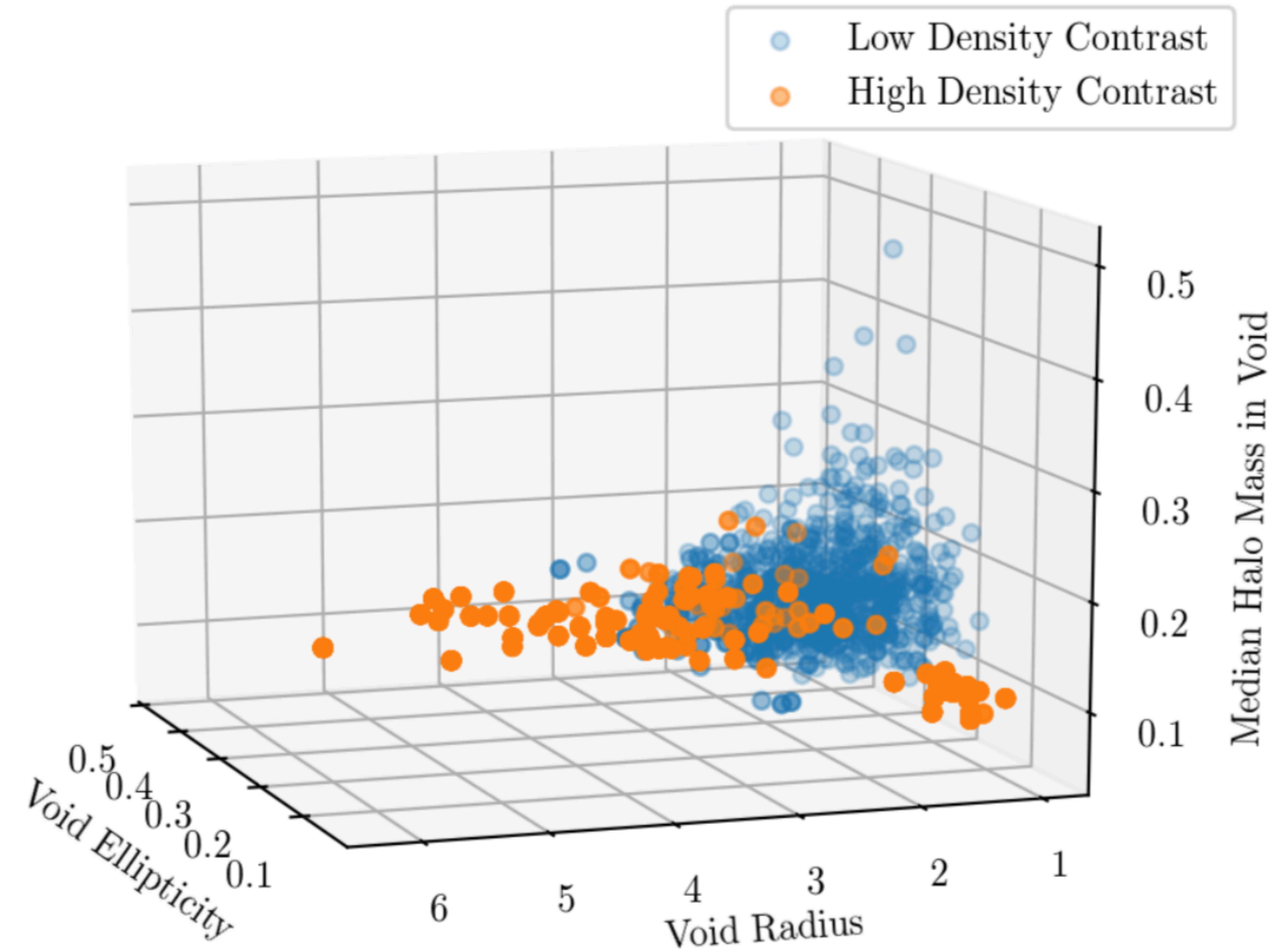
- Random Forest predicted mass distribution almost perfect agreement with true distribution
- Bayesian Regression has heavier tails than true distribution

Mass Trends with Void Features



- Low Density Contrast:
  - Minimal spread in radius
    - Most small
  - Most spread in ellipticity
- High Density Contrast:
  - Ellipticity increases with radius
  - More scatter for high ellipticity and radius





- High Density Contrast:
  - Complex relationship
  - Small + round: low mass
  - Radius + ellipticity ↑: mass increases then levels off

- Low Density Contrast:
  - Higher mass than for High Density Contrast
  - Mass decreases and has less scatter as ellipticity increases
  - Rounder: larger mass

Linear Regression: fits low density contrast population (Table 2)— why radius not top feature

Predicting Median Halo Mass

—2nd Order Polynomial Features—

- Non-linearities are clearly present in data (see visualizations)
- 2nd order: start simple, preference to lower order in astrophysics
- MSE improves for all models (see Table 1)
    - Confirmation of non-linearities in data
- Random Forest still most accurate
- Ellipticity x Radius term most important for Random Forest
    - Speaks to relationship in high density contrast voids
- Top features for Bayesian Regression only include ellipticity, radius, and density contrast (density contrast not present before)
    - Convergence of important features
- Takeaways:

1. Data are strongly non-linear and require strongly non-linear models
2. Most important features for predicting mass are indeed perhaps those selected by the Random Forest

Feature	LR	BR + EN	RF
voidEllipticity voidRadius	-0.116405	**-0.235247 $\pm$ 0.004730**	**0.215216**
voidDensityContrast	**-1.317766**	**-0.875281 $\pm$ 0.005303**	**0.138871**
voidEllipticity voidDensityContrast	0.121850	**0.606563 $\pm$ 0.005240**	**0.133671**
voidDensityContrast^2	**0.637442**	—	**0.121085**
voidRadius voidDensityContrast	**0.350251**	—	**0.111713**
voidEllipticity^2	-0.014487	—	0.079694
voidEllipticity	-0.162293	-0.170988 $\pm$ 0.003660	0.078789
voidRadius	-0.131148	**0.332322 $\pm$ 0.008559**	0.058743
voidRadius^2	0.109956	**-0.259320 $\pm$ 0.009312**	0.050404
voidDensityContrast voidCentralDen	0.260904	—	0.003452
voidRadius voidCentralDen	**-0.692097**	—	0.002401
voidEllipticity voidNumChildren	0.104119	—	0.002399
voidCentralDen^2	-0.111892		0.001275
voidRadius voidNumChildren	-0.048491	0.224913 $\pm$ 0.017066	0.000968
voidDensityContrast voidNumChildren	-0.157952	—	0.000530
voidEllipticity voidCentralDen	0.140832	—	0.000401
voidNumChildren	0.110550	0.010304 $\pm$ 0.009558	0.000216
voidNumChildren^2	-0.012982	-0.143589 $\pm$ 0.007340	0.000130
voidCentralDen	**0.385748**	—	0.000032
voidNumChildren voidCentralDen	0.046160	—	0.000009

Table 4: Feature weights for 2nd order polynomial.

Predicting Median Halo Mass

—Power Law Features—

- Classic functional form in astrophysics
- Use only ellipticity, radius, density contrast (convergence)
- Regression problem now:

$$\tilde{M}_{\rm h} = \delta^{\alpha} e^{\beta} r_{\rm eff}^{\gamma}$$

$\tilde{M}_{\rm h}$ is the median halo mass,
δ is the void density contrast,
e is the void ellipticity,
$r_{\rm eff}$ is the void radius.

which can be rewritten as

$$\log_{10}\tilde{M}_{\rm h} = \alpha\log_{10}\delta + \beta\log_{10}e + \gamma\log_{10}r_{\rm eff}$$

Feature	LR	BR	RF
voidDensityContrast, α	-0.430270	-0.430262 $\pm$ 0.001195	0.430152
voidEllipticity, β	0.207581	-0.177645 $\pm$ 0.001175	0.331069
voidRadius, γ	-0.177650	0.207577 $\pm$ 0.001129	0.238779

Table 5: Feature weights for power law fit

- MSE improves for both linear models
    - Data extremely non-linear
    - Perhaps combination of power laws and polynomials is best
- MSE degrades for Random Forest
    - Limited to only 3 features- lack of freedom to fit non-linearities

- Takeaways:

1. The Random Forest is our only model capable of capturing the complex non-linearities in the data
2. Relationship between mass and void features is far more complicated than simple power law

What do Voids Add?

- Run same analysis but include halo population features
  - E.g.: median, standard deviation, skew, and kurtosis of star formation rate, black hole accretion, gas metallicity, etc.
- MSE for linear models is better with halo features (Table 1)
- MSE for Random Forest degrades with halo features

- Takeaways:

1. Mass scales more linearly with halo features
2. Halo features unnecessary for non-linear models

Metric: Mean Squared Error

Features	LR	EN	BR + EN	RF
Linear	0.9072	1.02758	0.98629	5.316×10^{-6}
Linear, individual mass	0.9999	1.0	—	0.9996
Linear halo + void features	0.3076	0.9802	0.4727	2.084×10^{-5}
Deg 2 Poly	0.8446	1.01779	0.88572	4.456×10^{-6}
Power Law	0.82076	1.63769	0.82076 (no EN)	5.066×10^{-6}

Table 1: MSE

Features	Low Density Contrast	High Density Contrast
Linear	0.0062	1.1518
Deg 2 Poly	0.93384	0.84726
Power Law	0.95695	0.82075

Table 2: LR MSE split on density contrast

Discussion & Conclusions

- Random Forest provides the best predictive power among our models
- There is a complex non-linear relationship between the mass and void properties
- Using polynomial and power law features improves the predictive power of linear models
- Halo features are unnecessary if using non-linear models