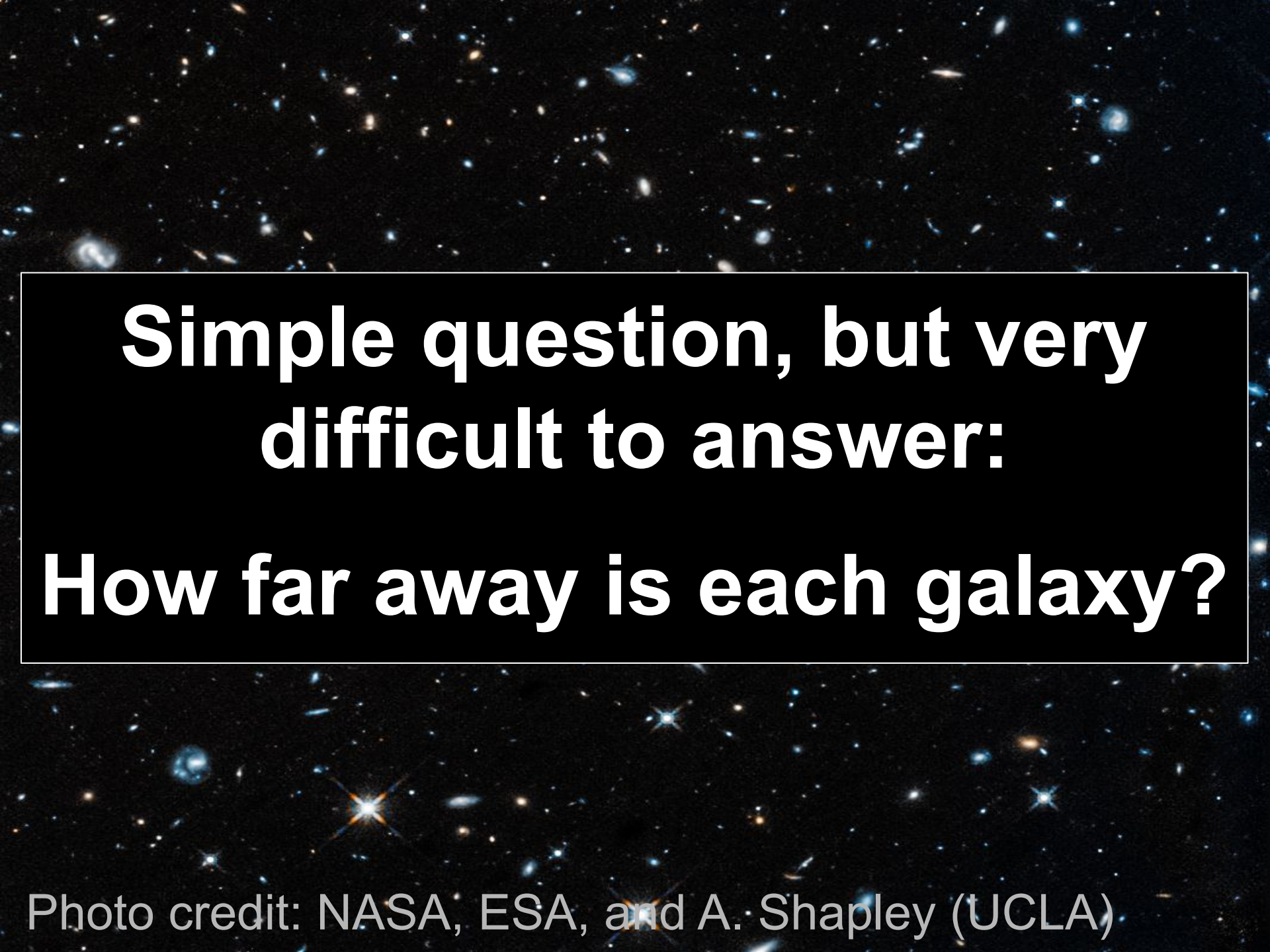
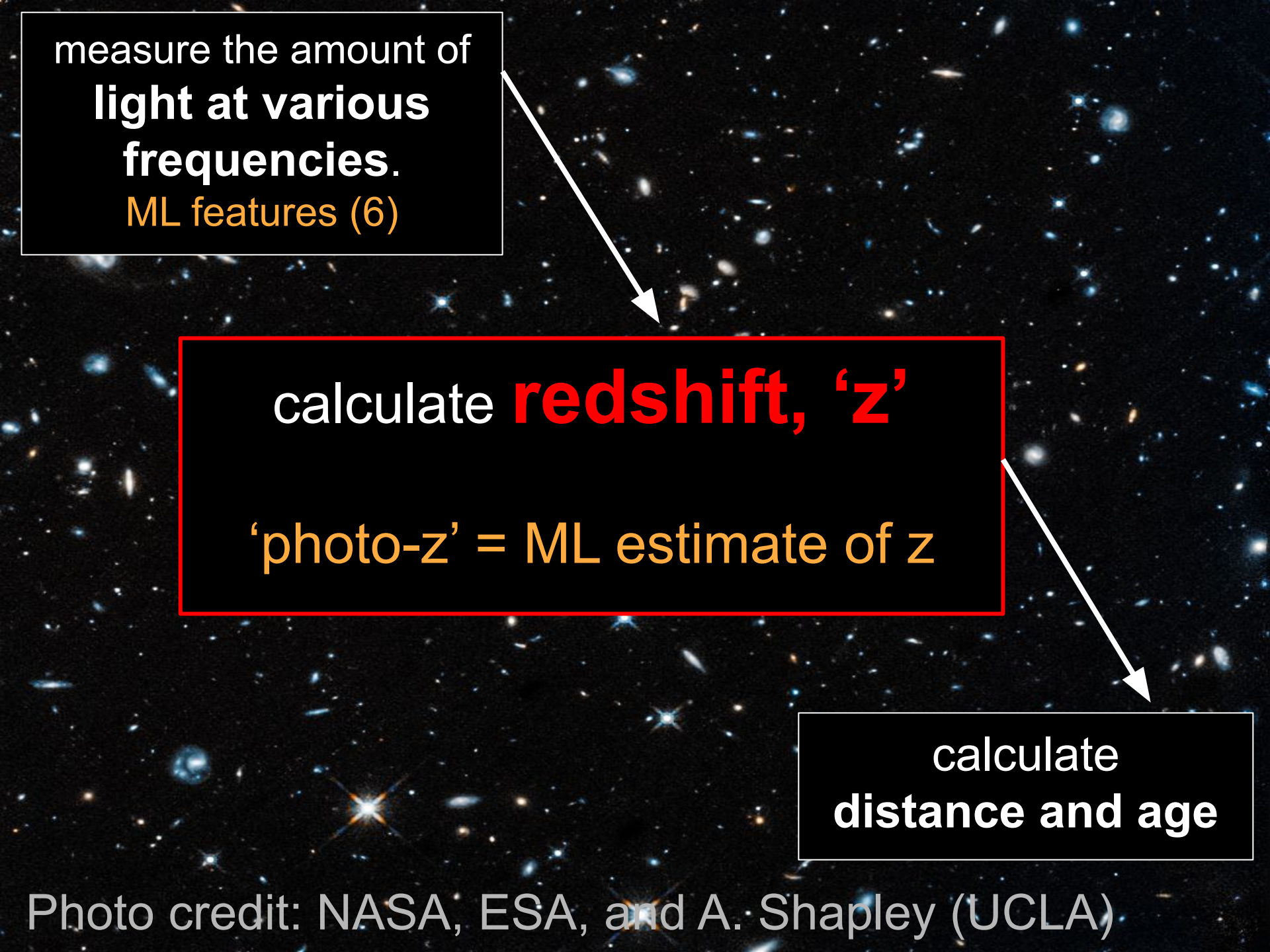
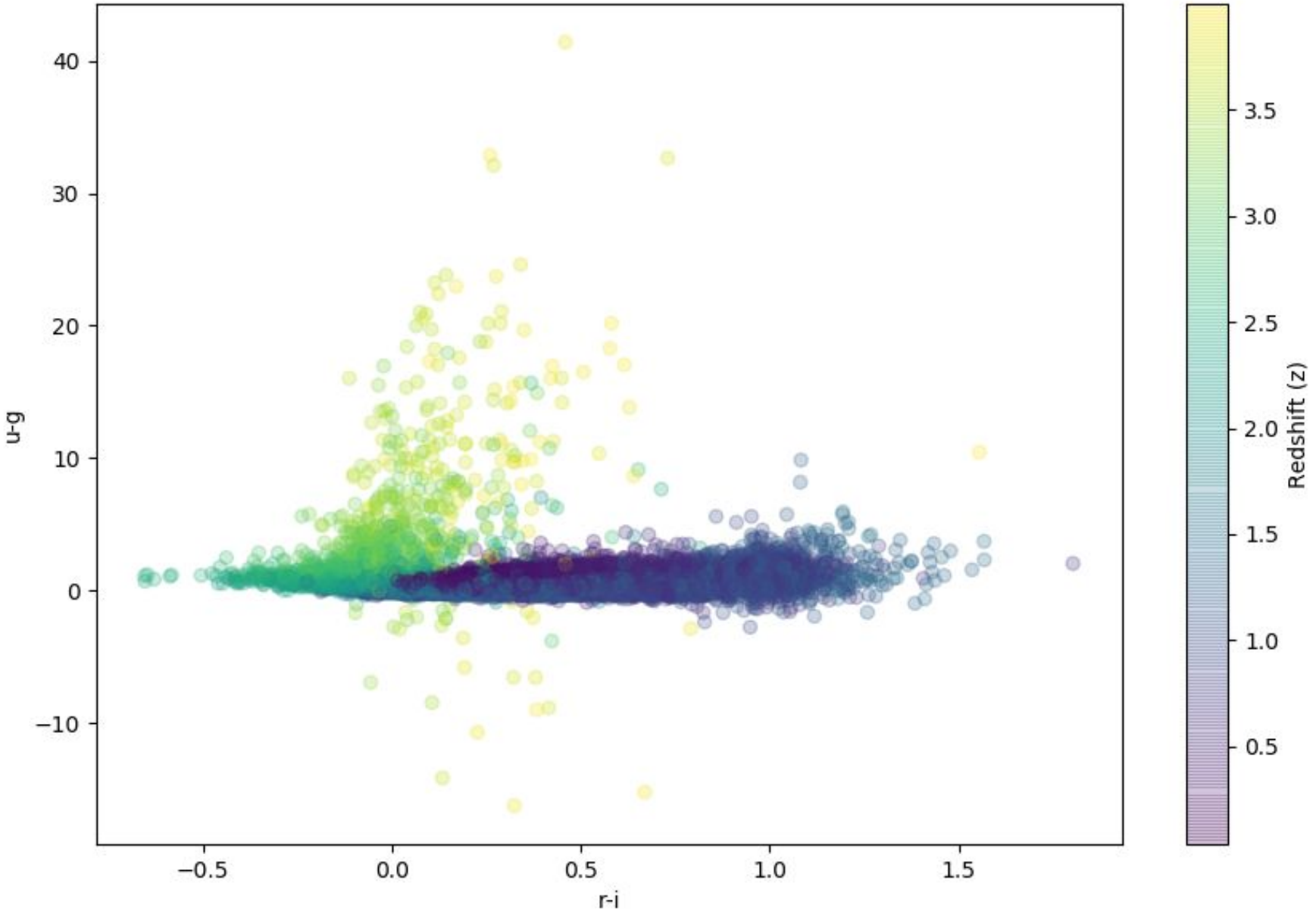
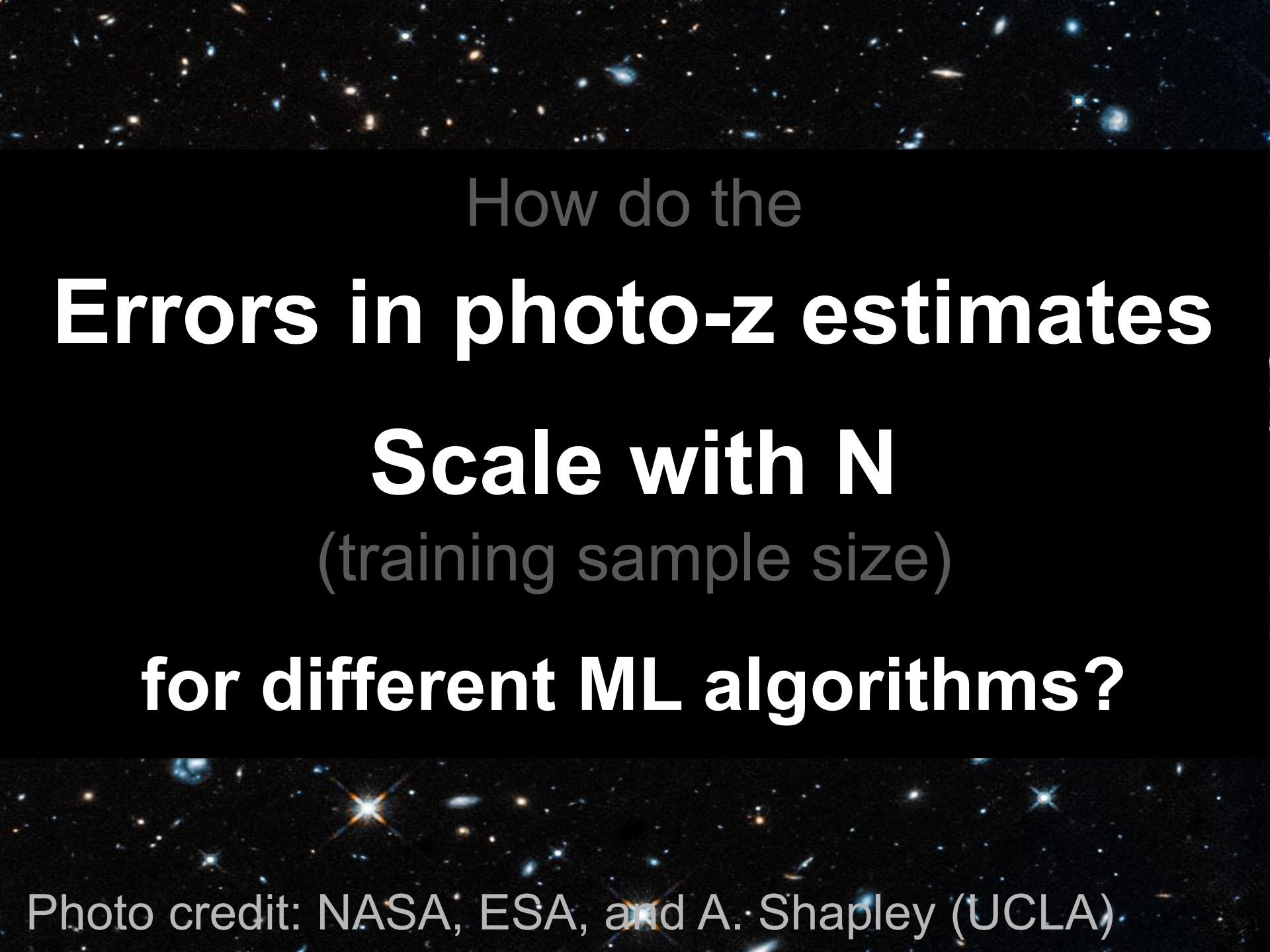
Photo credit: NASA, ESA, and A. Shapley (UCLA)


Simple question, but very difficult to answer:
How far away is each galaxy?
Photo credit: NASA, ESA, and A. Shapley (UCLA)


measure the amount of
light at various frequencies.
ML features (6)
calculate **redshift, 'z'**
'photo-z' = ML estimate of z
calculate
distance and age
Photo credit: NASA, ESA, and A. Shapley (UCLA)

2 features with strongest correlations with Redshift




How do the
Errors in photo-z estimates
Scale with N
(training sample size)
for different ML algorithms?
Photo credit: NASA, ESA, and A. Shapley (UCLA)

Error Metric:

$$\Delta z \equiv \frac{photo_z - true_z}{1 + true_z}$$

2 Statistics of the metric:

$$\mathrm{NMAD} = 1.48 \times \mathrm{median}(|\Delta z|)$$

$$\mathrm{OUT10} = \frac{1}{N} \sum_{n=1}^{N} \left[|\Delta z_n| > 0.1 \right]$$

Tested Algorithms

2 **Neural Net** architectures:

- 2 hidden layers, 10 units each
- 3 hidden layers, 15 units each

Random Forest Regressor

GPz

GPz

Assume: $y = f(\mathbf{x}) + \epsilon$ (epsilon = Gaussian noise)

Kernel density estimators

(radial basis functions)

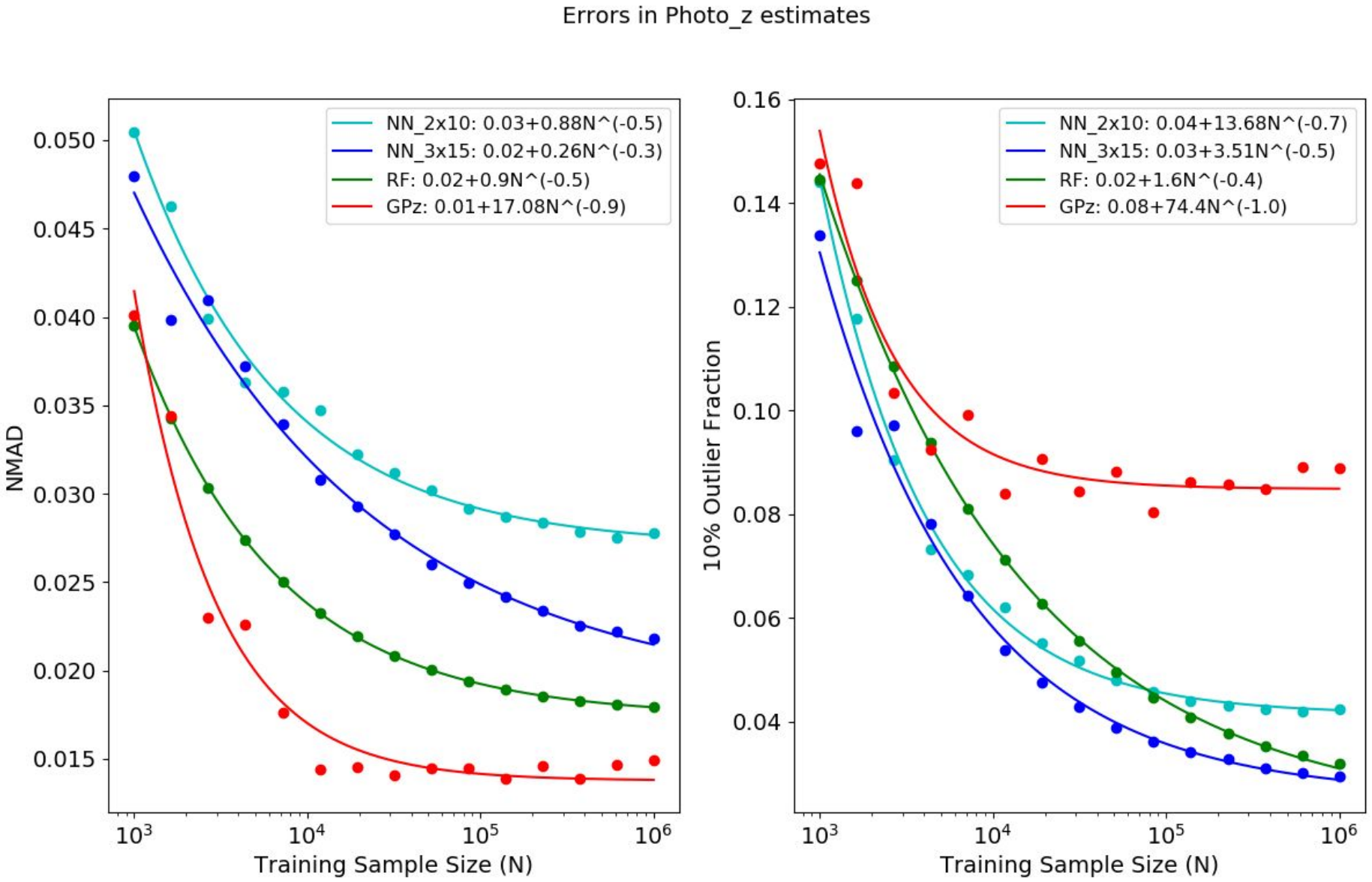
Calculate:

$$p(f|y,\mathbf{X}) = \frac{p(y|f)p(f|\mathbf{X})}{p(y|\mathbf{X})}$$

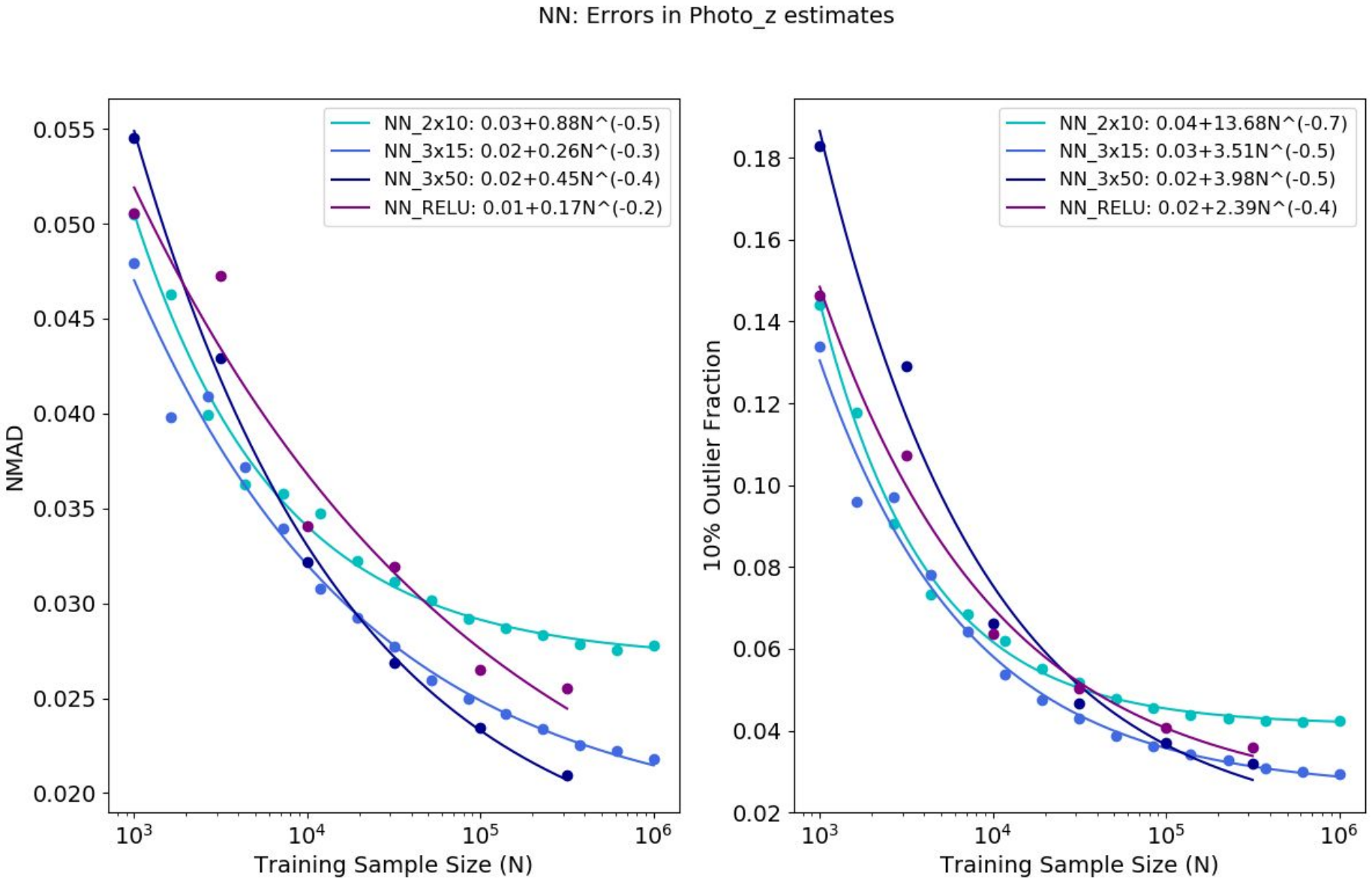
$p(y|f) \sim N(f, \sigma^2)$

normalization constant

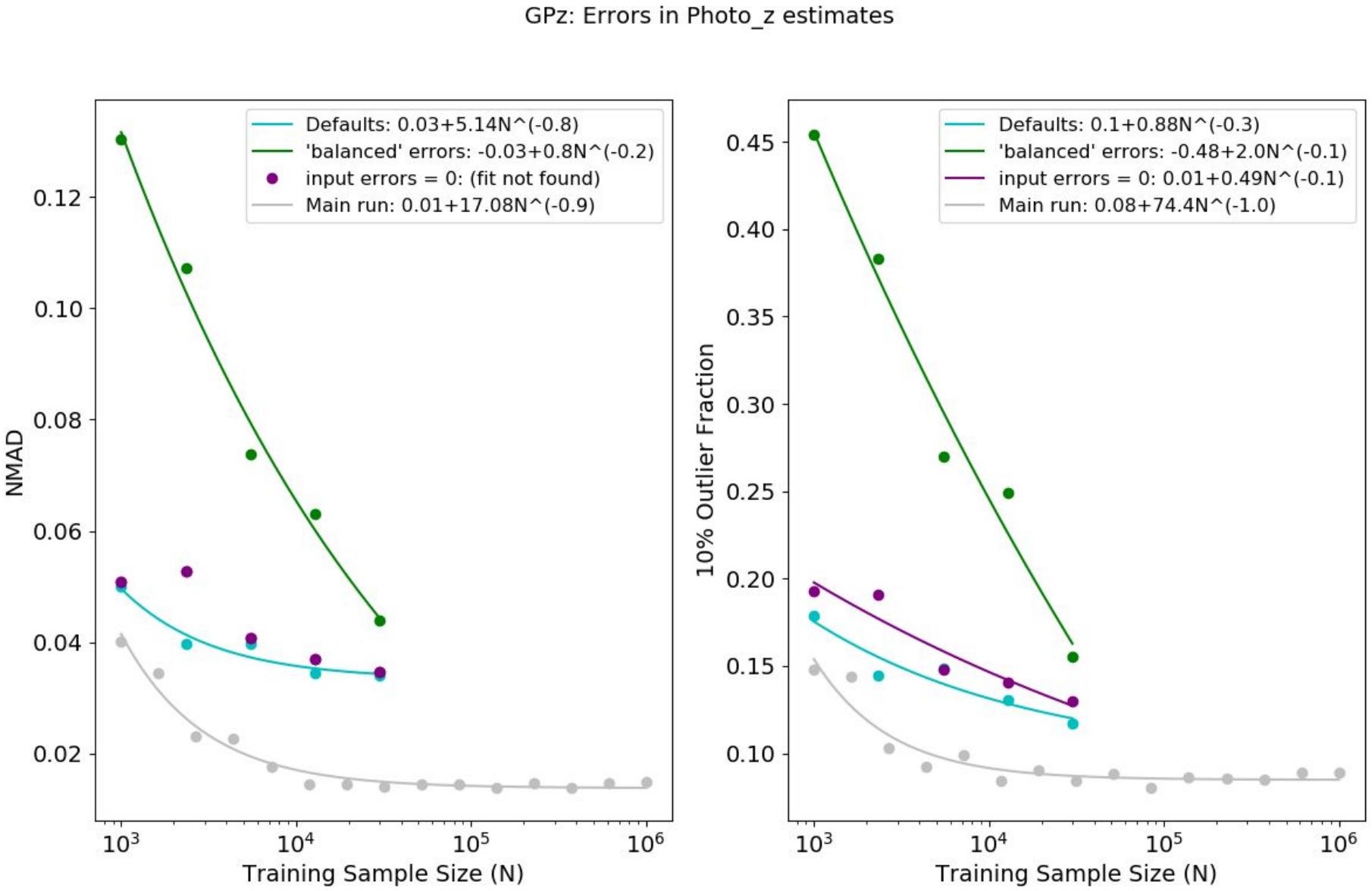
Results


Errors in Photo_z estimates
NN_2x10: 0.03+0.88N^(-0.5)
NN_3x15: 0.02+0.26N^(-0.3)
RF: 0.02+0.9N^(-0.5)
GPz: 0.01+17.08N^(-0.9)
NMAD
Training Sample Size (N)
NN_2x10: 0.04+13.68N^(-0.7)
NN_3x15: 0.03+3.51N^(-0.5)
RF: 0.02+1.6N^(-0.4)
GPz: 0.08+74.4N^(-1.0)
10% Outlier Fraction
Training Sample Size (N)

Results


NN: Errors in Photo_z estimates
NN_2x10: 0.03+0.88N^(-0.5)
NN_3x15: 0.02+0.26N^(-0.3)
NN_3x50: 0.02+0.45N^(-0.4)
NN_RELU: 0.01+0.17N^(-0.2)
NMAD
Training Sample Size (N)
NN_2x10: 0.04+13.68N^(-0.7)
NN_3x15: 0.03+3.51N^(-0.5)
NN_3x50: 0.02+3.98N^(-0.5)
NN_RELU: 0.02+2.39N^(-0.4)
10% Outlier Fraction
Training Sample Size (N)

Results


GPz: Errors in Photo_z estimates
Defaults: 0.03+5.14N^(-0.8)
'balanced' errors: -0.03+0.8N^(-0.2)
input errors = 0: (fit not found)
Main run: 0.01+17.08N^(-0.9)
NMAD
Training Sample Size (N)
Defaults: 0.1+0.88N^(-0.3)
'balanced' errors: -0.48+2.0N^(-0.1)
input errors = 0: 0.01+0.49N^(-0.1)
Main run: 0.08+74.4N^(-1.0)
10% Outlier Fraction
Training Sample Size (N)

Future Work

Tune these algorithms (particularly GPz)

Test other algorithms (e.g. ANNz, TPz)

Explore GPz kernel functions and what they can tell us about galaxy types and distributions.