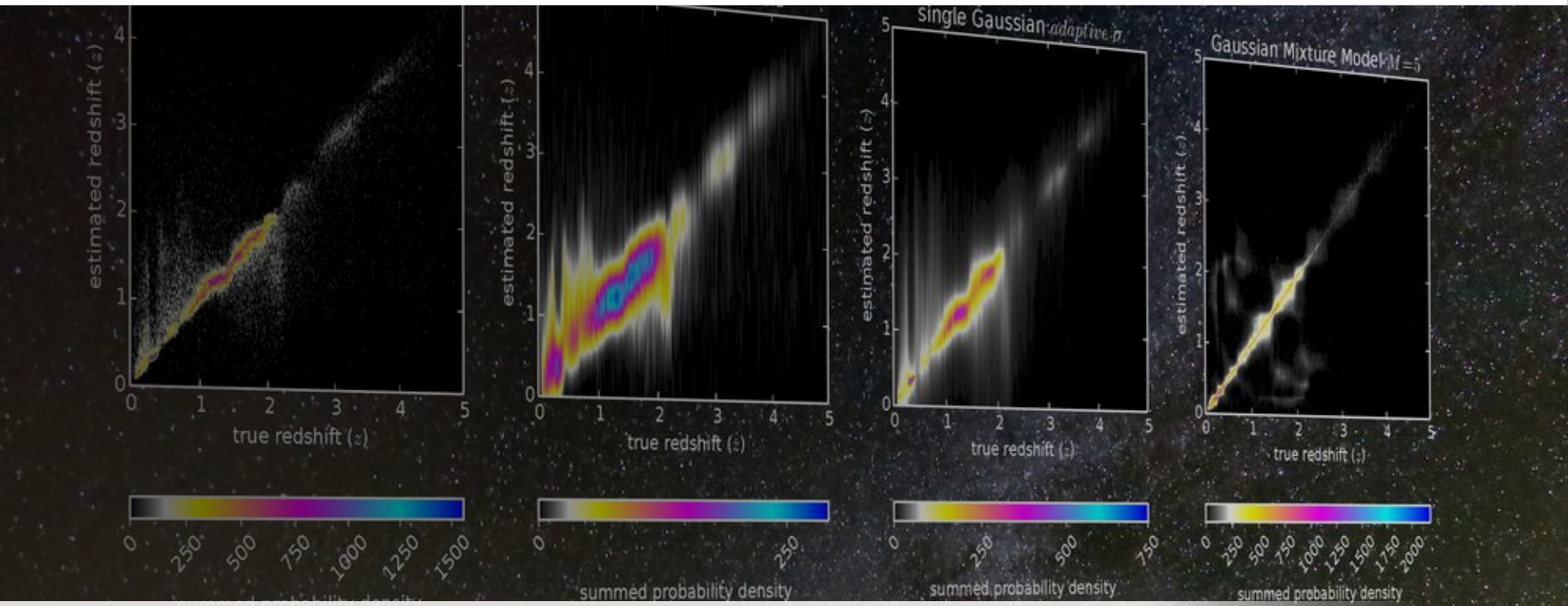
Heidelberg Institute for
Theoretical Studies



Multimodal Redshift Estimation

proper use of probability density functions

Regression Problems in Astronomy

analysis of large catalogs demands efficient solving of regression problems

$$f(\vec{x}) \rightarrow y, \text{where } \vec{x} \in \mathbb{R}^n, y \in \mathbb{R}$$

model

features

estimate

Improvements

new / improved algorithms

feature / metric selection

domain shift

missing / censored features

how about **uncertainties**?

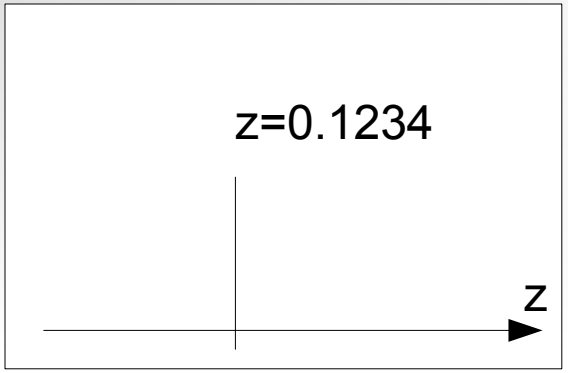
Uncertainties

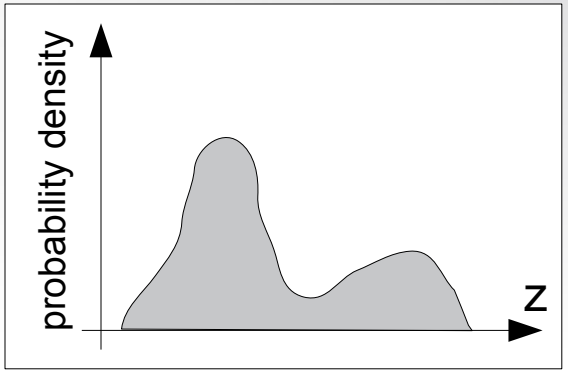
$$f(\vec{x}) \rightarrow y, \text{where } \vec{x} \in \mathbb{R}^n, y \in \mathbb{R}$$

model uncertainty

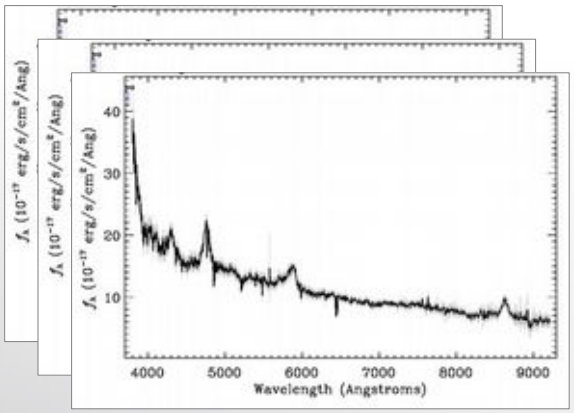
input uncertainty

uncertain estimate



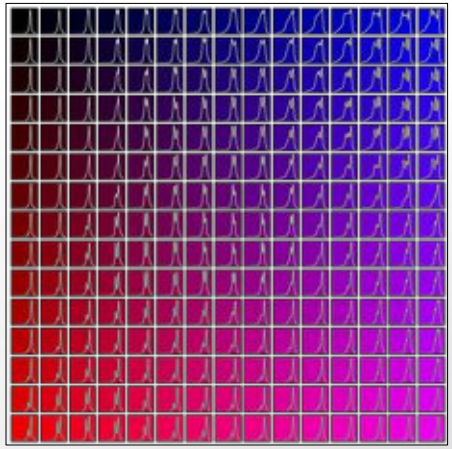


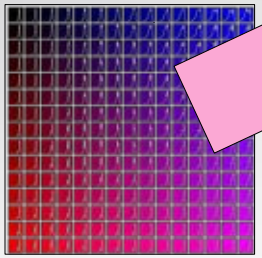
Spectral Model



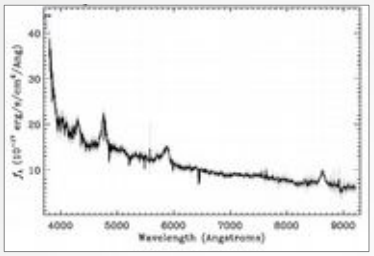
rest frame, z=0

PPCA

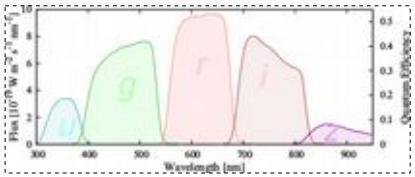




shift by z



+



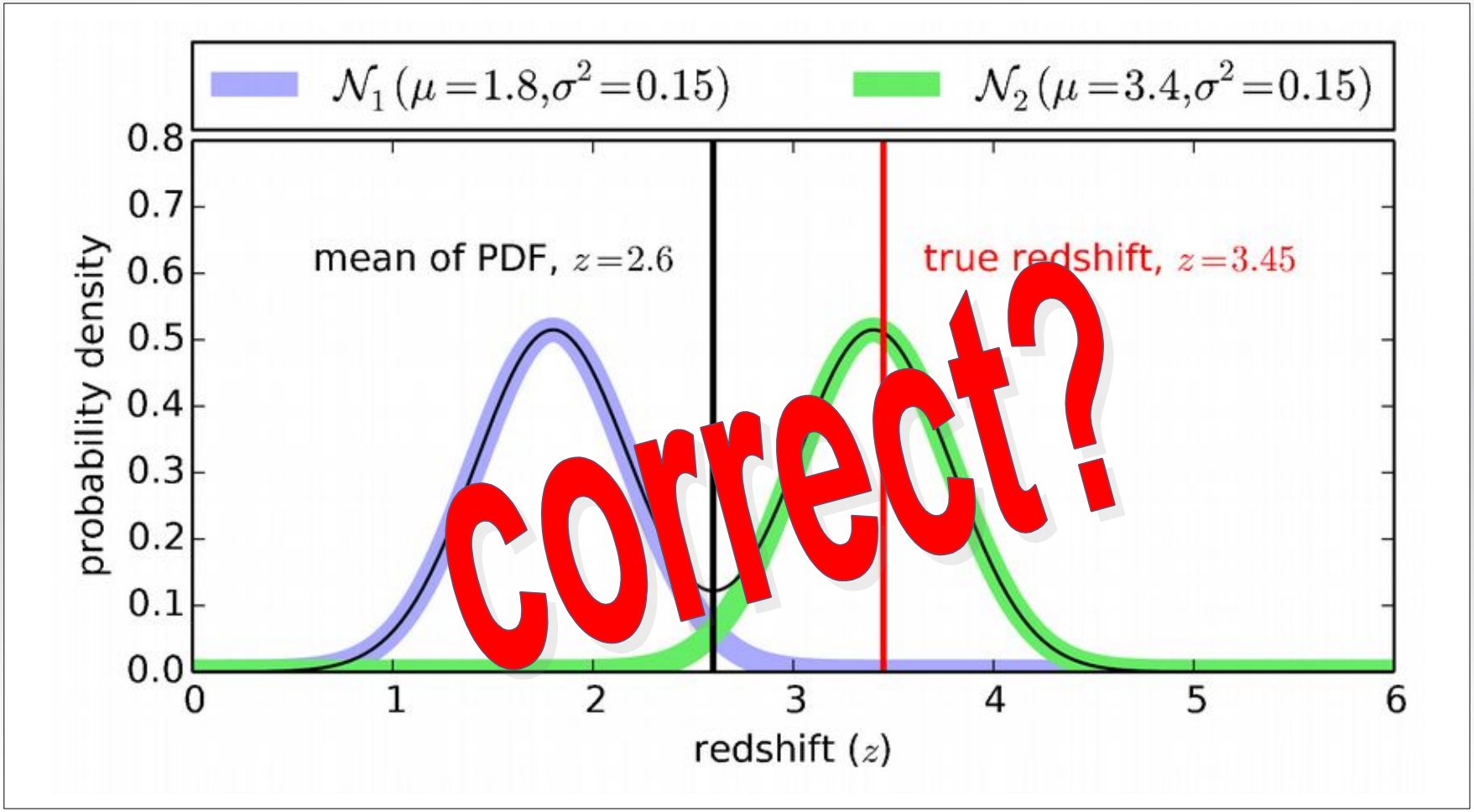
= u,g,r,i,z + noise

compare

observation

Evaluation Tools

simplification / just use the mean



Evaluation Tools

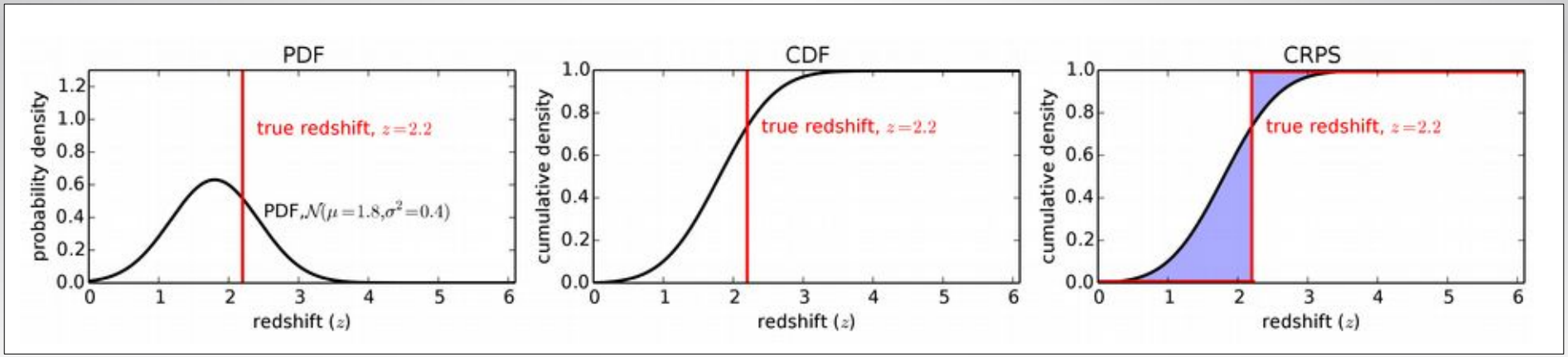
stacking PDFs

Proper Evaluation Tools / CRPS

continuous rank probability score

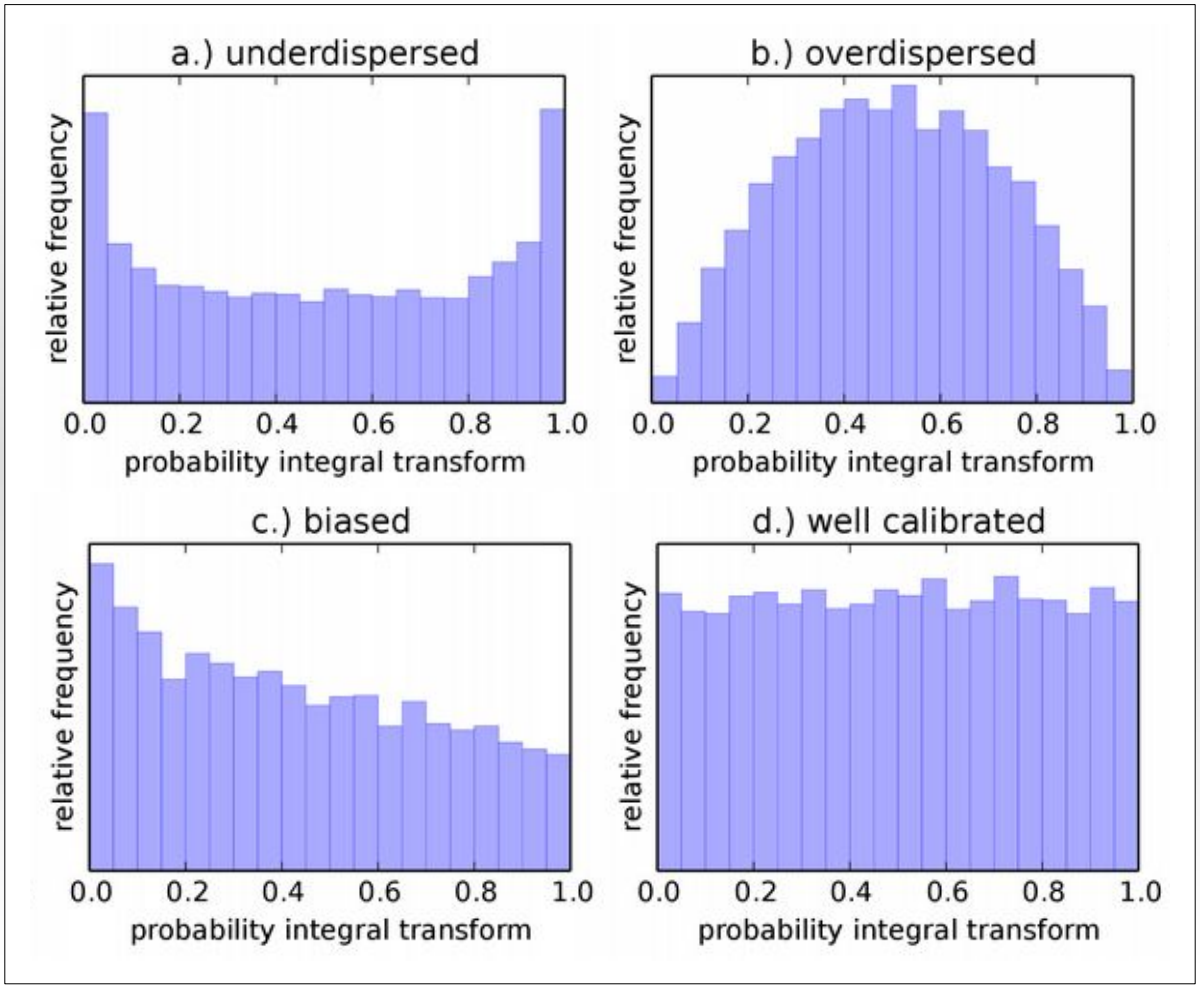
$$CRPS = \frac{1}{N}\sum_{t=1}^{N} crps(CDF_t, z_t),$$

$$\text{with } crps(CDF_t, z_t) = \int_{-\infty}^{+\infty} [CDF_t(z) - CDF_{z_t}(z)]^2 \, dz$$

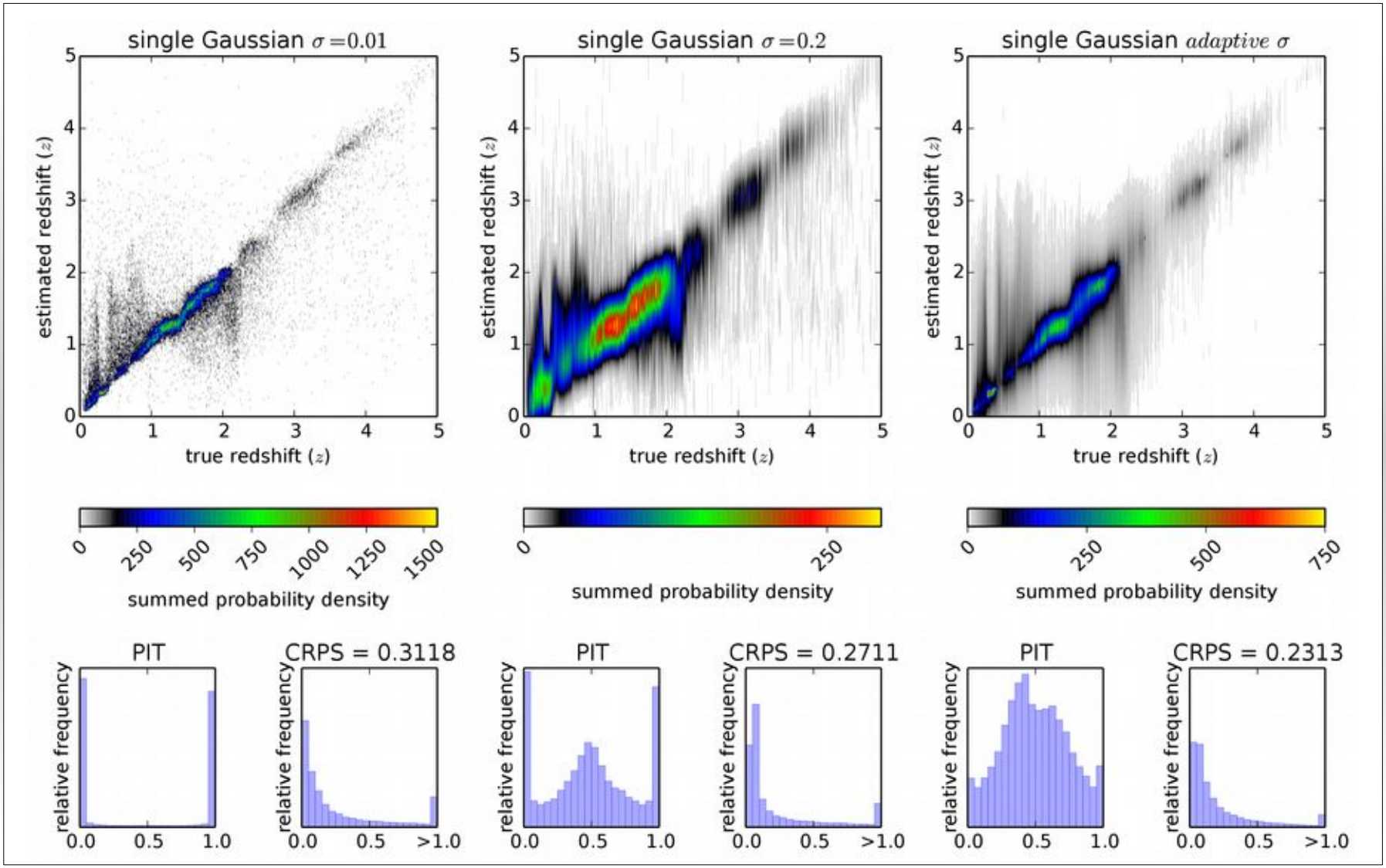


Proper Evaluation Tools / PIT

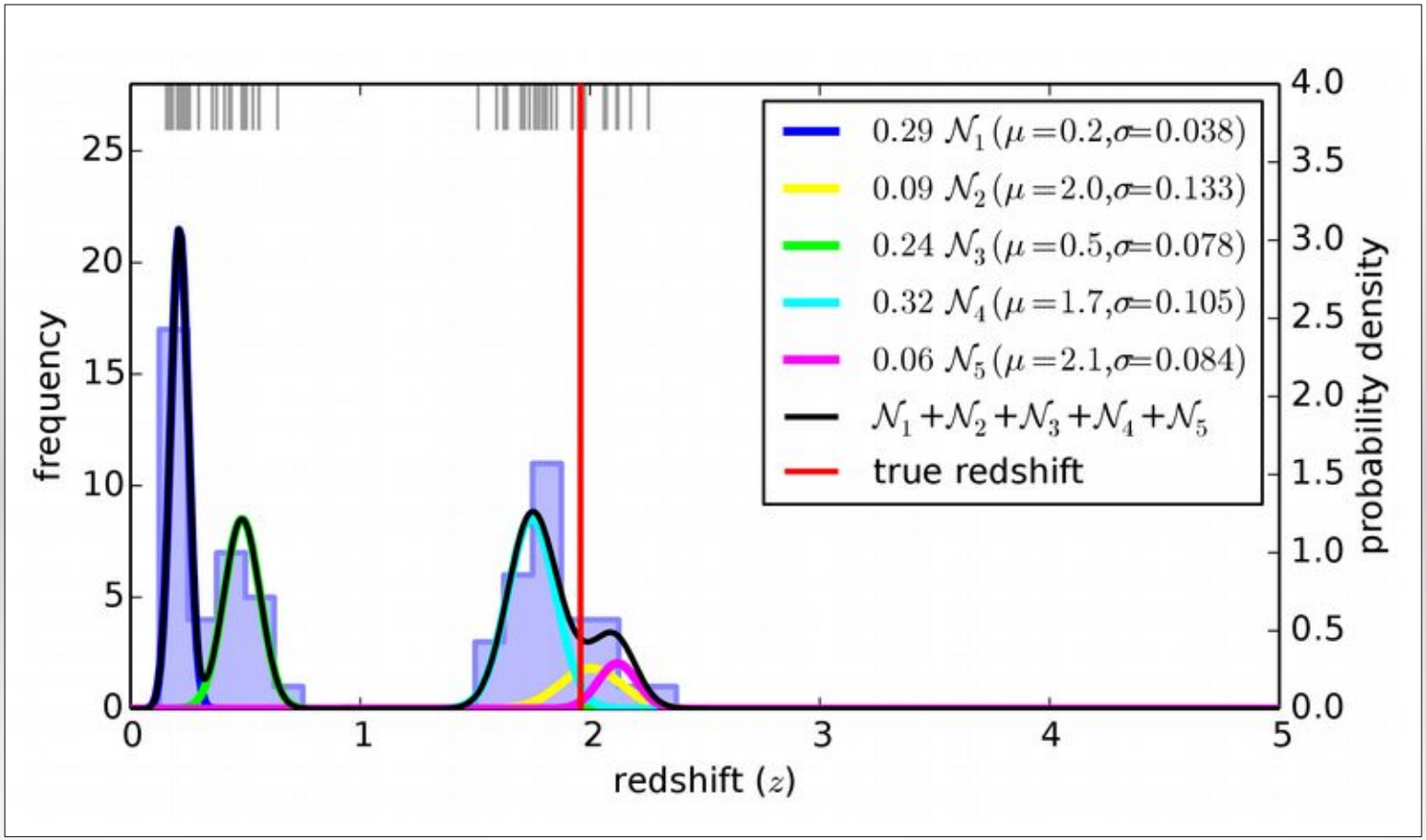
probability integral transform



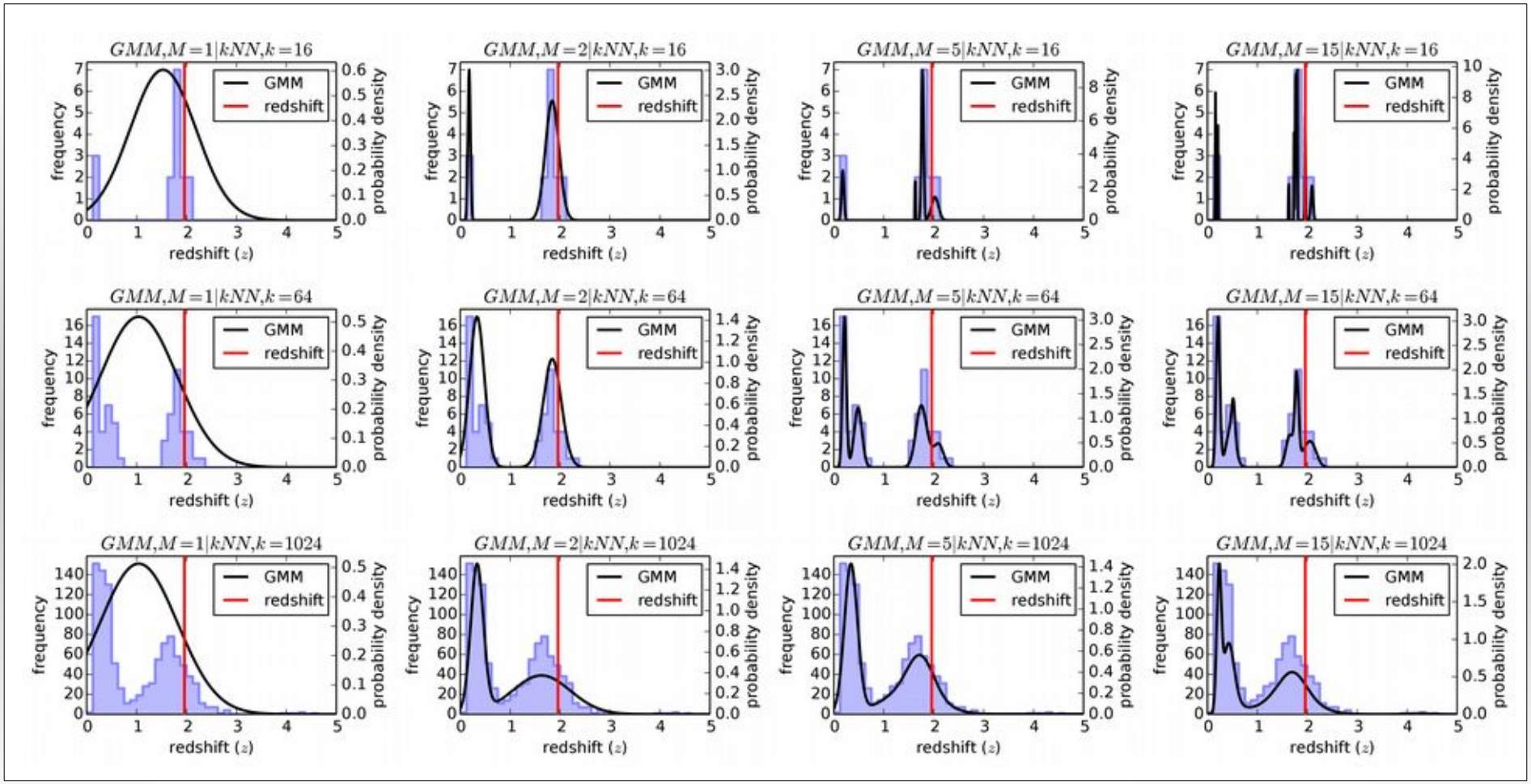
Results


single Gaussian $\sigma = 0.01$
single Gaussian $\sigma = 0.2$
single Gaussian $adaptive\ \sigma$
estimated redshift (z)
true redshift (z)
summed probability density
PIT
CRPS = 0.3118
CRPS = 0.2711
CRPS = 0.2313
relative frequency

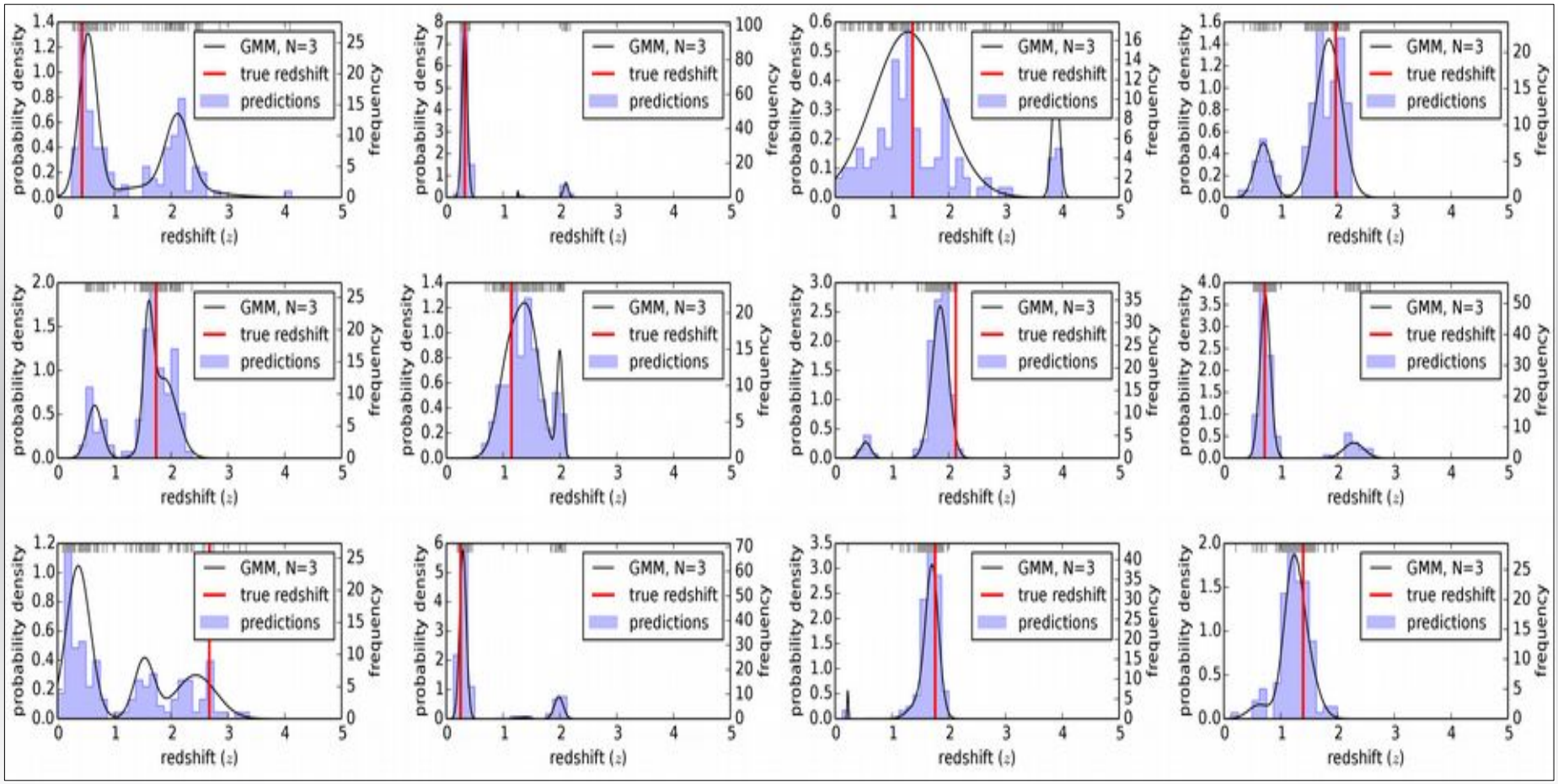
Multi-Modalities



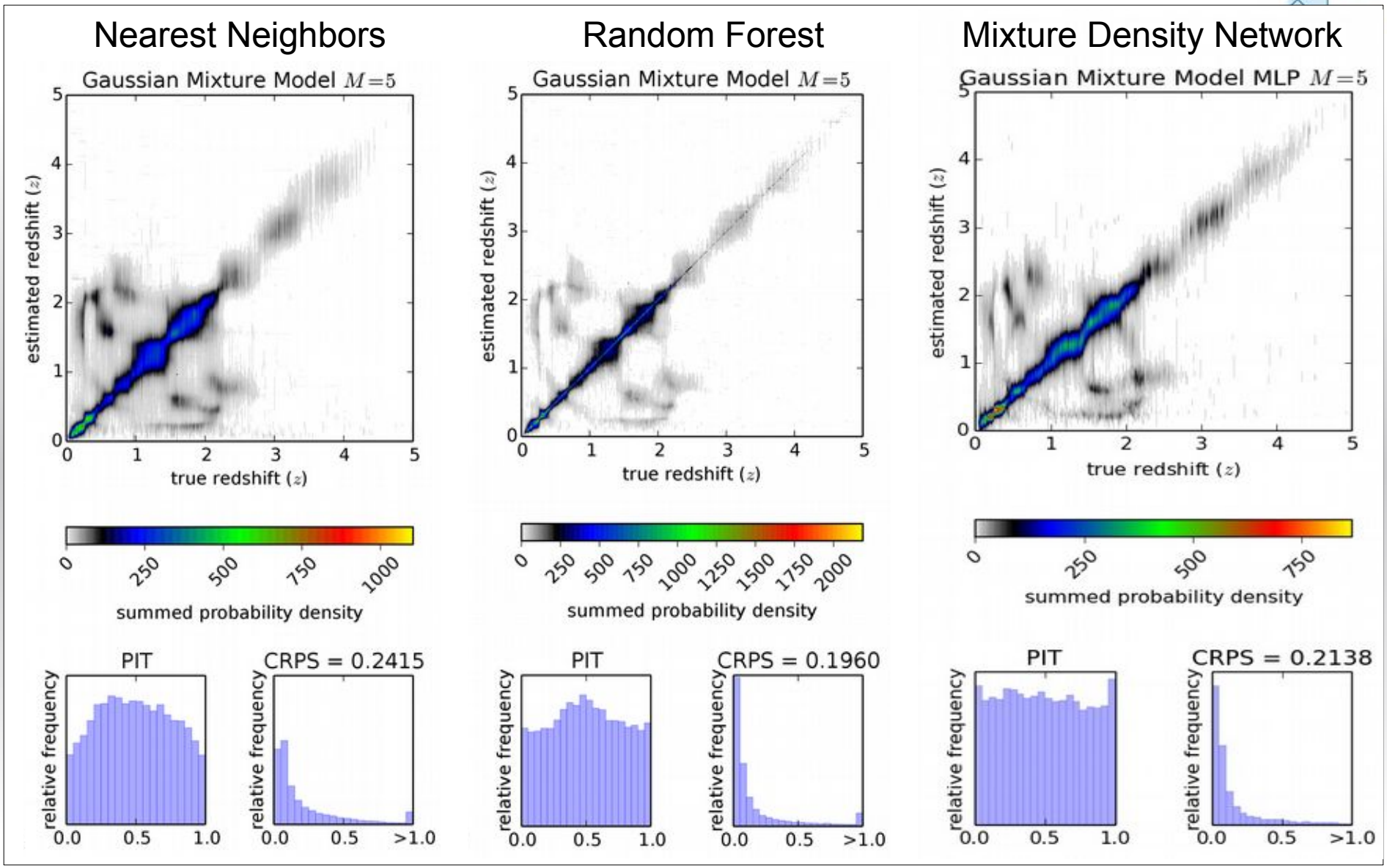
Multi-Modalities



Multi-Modalities

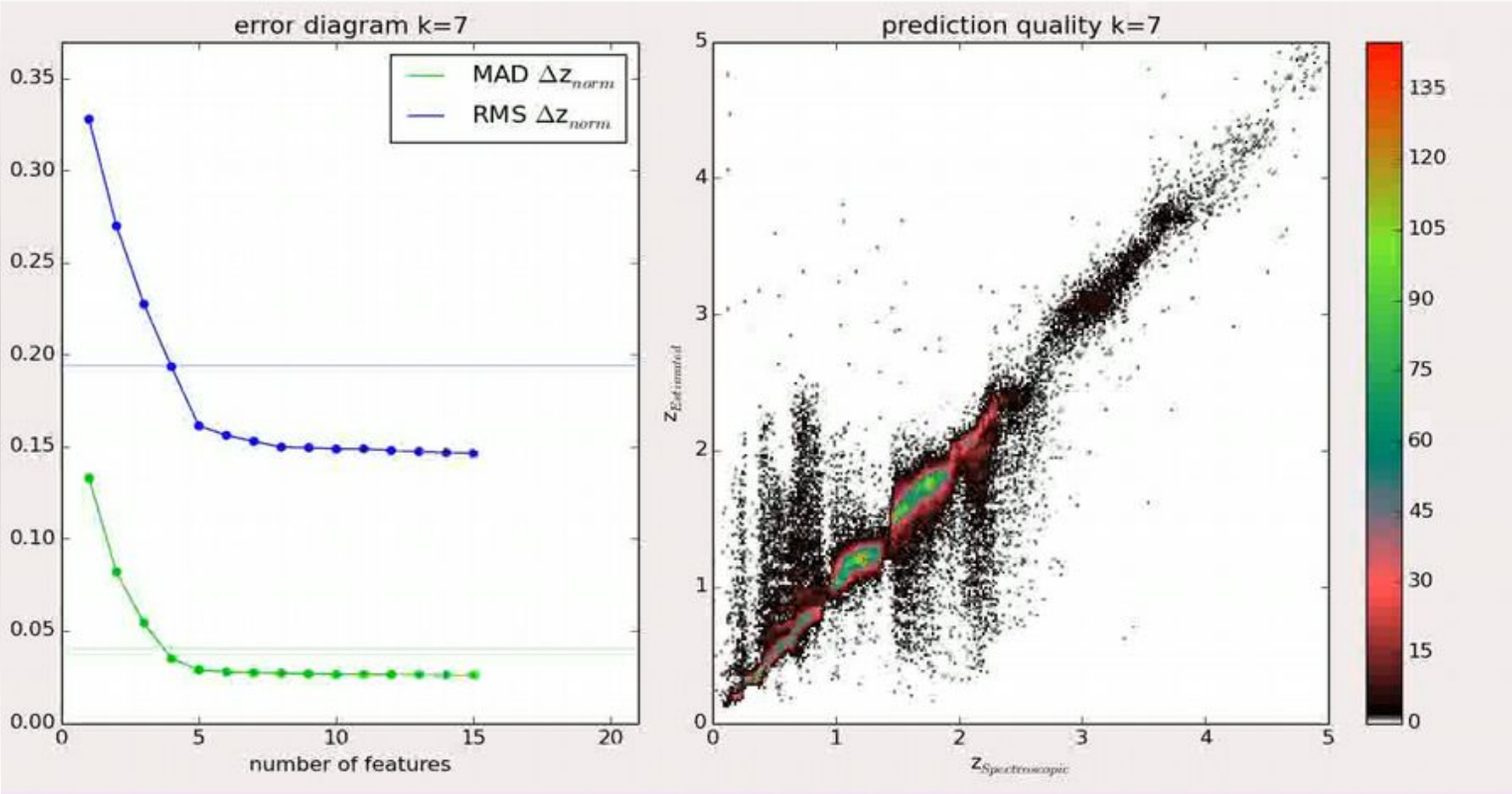


Results


Nearest Neighbors
Random Forest
Mixture Density Network
Gaussian Mixture Model $M=5$
Gaussian Mixture Model $M=5$
Gaussian Mixture Model MLP $M=5$
estimated redshift (z)
true redshift (z)
summed probability density
PIT
CRPS = 0.2415
CRPS = 0.1960
CRPS = 0.2138
relative frequency

Forward Selection

apply greedy forward selection



Forward Selection

resulting features:

$$u_{psf} - g_{petrosian}$$

$$dered(z_{psf}) - dered(i_{petrosian})$$

$$dered(g_{psf}) - dered(r_{model})$$

$$\frac{dered(r_{psf}) - dered(z_{model})}{\sqrt{\sigma^2_{g_{model}} + \sigma^2_{r_{model}}}}$$

$$dered(r_{model}) - dered(i_{model})$$

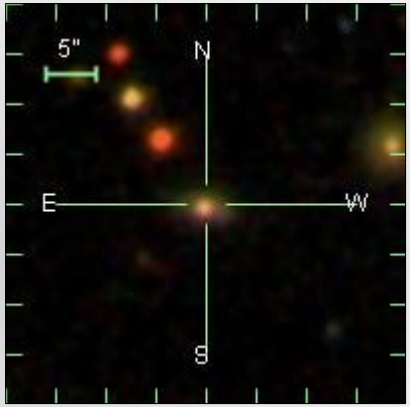
$$i_{psf} - i_{petrosian}$$

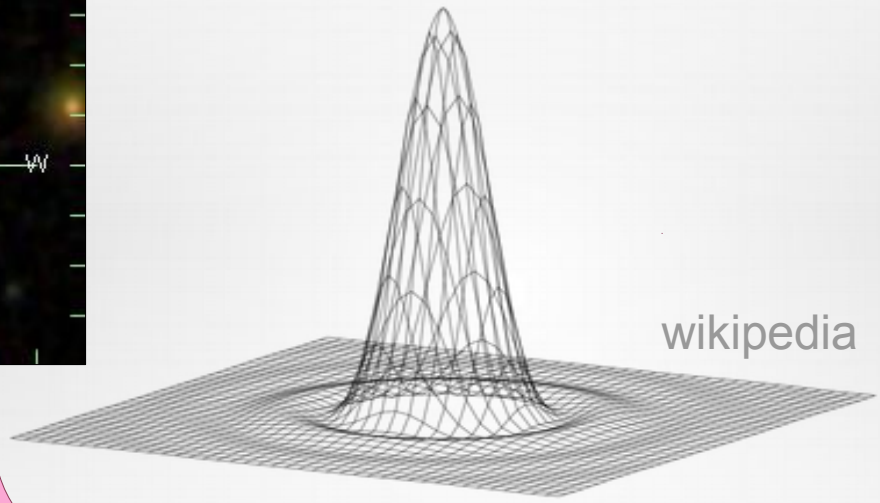
$$dered(z_{psf}) - dered(r_{petrosian})$$

$$\frac{g_{model} - g_{petrosian}}{\sqrt{\sigma^2_{g_{petrosian}} + \sigma^2_{r_{petrosian}}}}$$

Lessons Learned

features stored in catalogs are **not** the best for **machine learning**



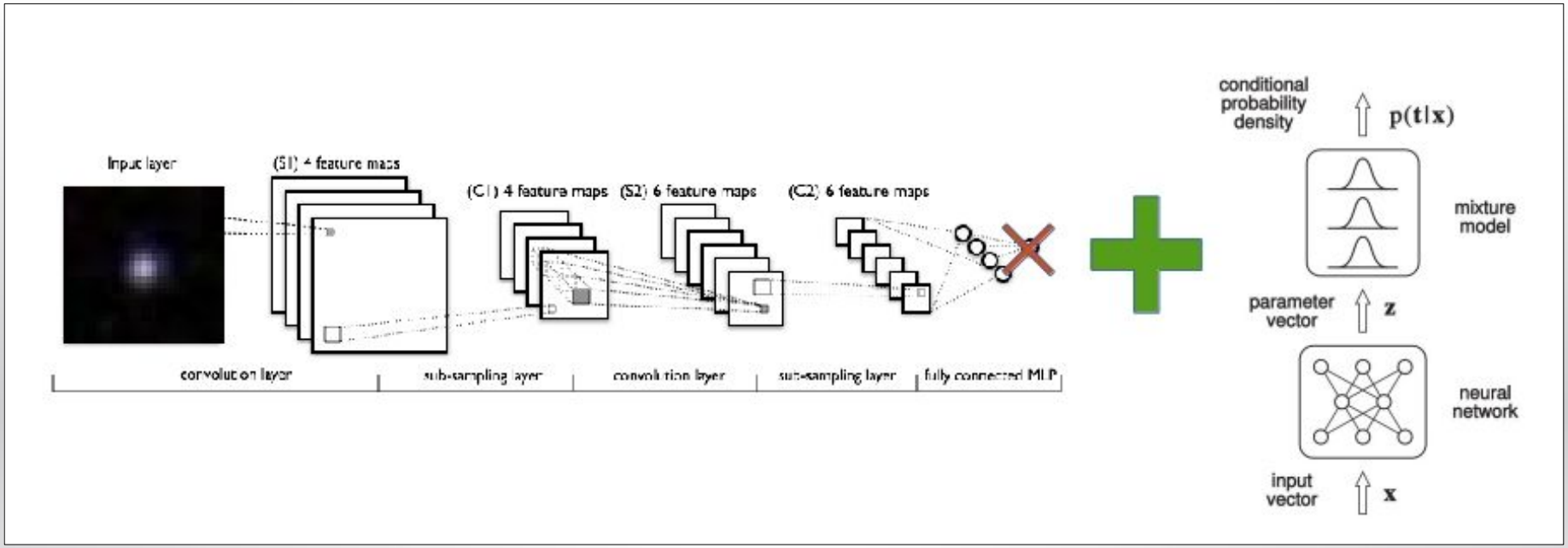


wikipedia

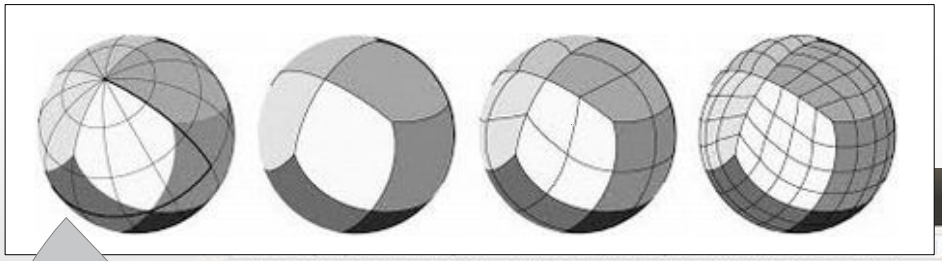
magnitudes

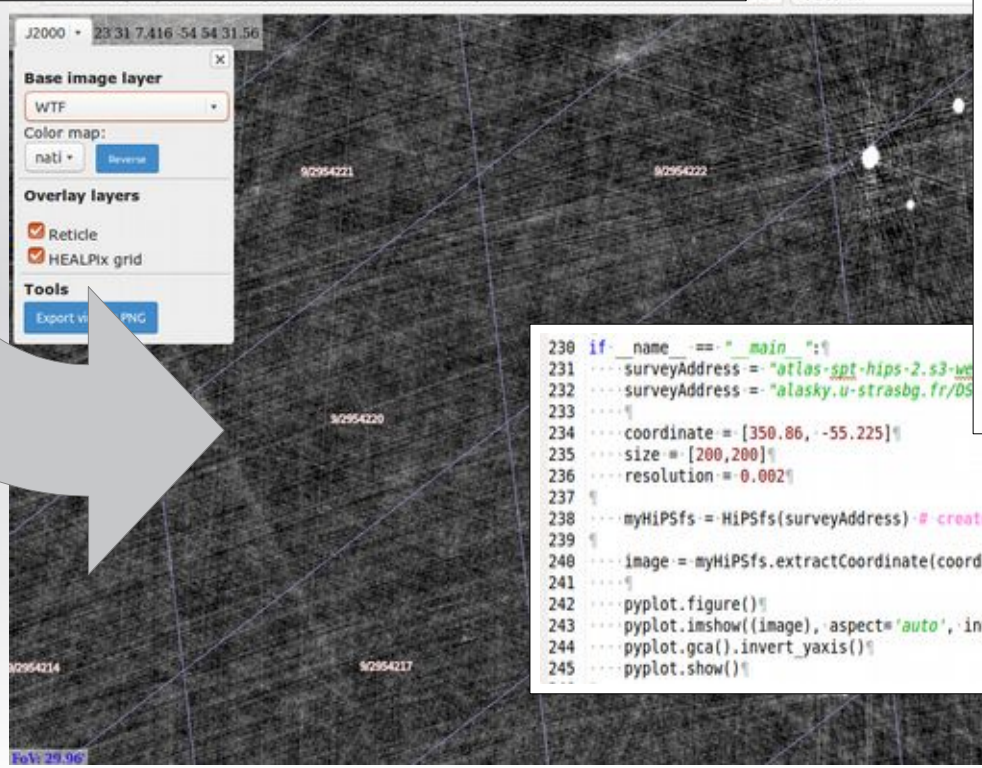
in (u, g, r, i, z)

DCN meet MDN

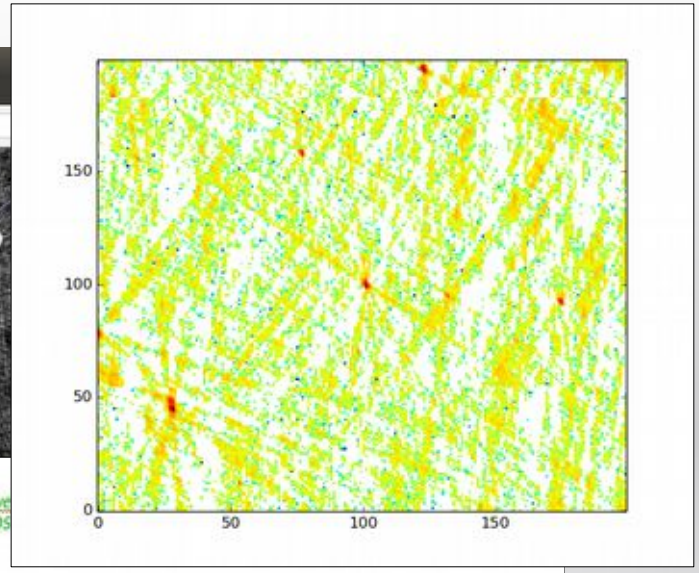

Input layer
(S1) 4 feature maps
(C1) 4 feature maps
(S2) 6 feature maps
(C2) 6 feature maps
convolution layer
sub-sampling layer
convolution layer
sub-sampling layer
fully connected MLP
conditional probability density
$p(\mathbf{t}|\mathbf{x})$
mixture model
parameter vector
$\mathbf{z}$
neural network
input vector
$\mathbf{x}$

Healpix / HiPS / IVOA

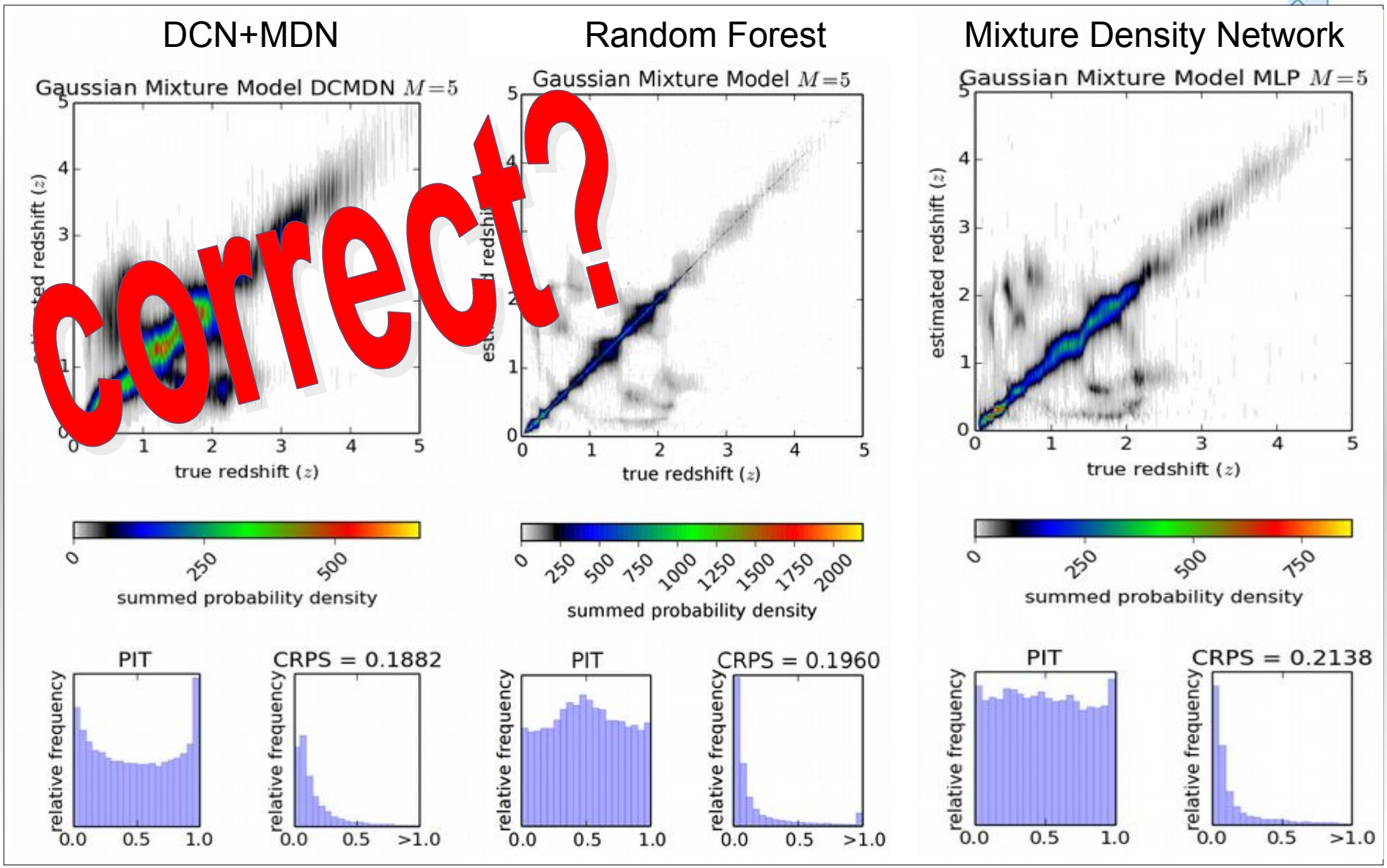





```
if __name__ == "__main__":
    surveyAddress = "atlas-spt-hips-2.s3-we
    surveyAddress = "alasky.u-strasbg.fr/DS

    coordinate = [350.86, -55.225]
    size = [200,200]
    resolution = 0.002

    myHiPSfs = HiPSfs(surveyAddress) # create access

    image = myHiPSfs.extractCoordinate(coordinate, size, resolution, nested=True) # extract data array

    pyplot.figure()
    pyplot.imshow((image), aspect='auto', interpolation="nearest")
    pyplot.gca().invert_yaxis()
    pyplot.show()
```



Results


DCN+MDN
Random Forest
Mixture Density Network
Gaussian Mixture Model DCMDN $M=5$
Gaussian Mixture Model $M=5$
Gaussian Mixture Model MLP $M=5$
estimated redshift (z)
true redshift (z)
summed probability density
PIT
CRPS = 0.1882
CRPS = 0.1960
CRPS = 0.2138
relative frequency
correct?

Conclusion

uncertainty is very **important** to further **improve** the science quality

we must start to use **proper** tools to **evaluate** the performance