

Uncertain Photometric Redshifts

deep learning meets probability density functions

Regression Problems in Astronomy



analysis of **large catalogs** demands efficient **solving**

regression of **problems**

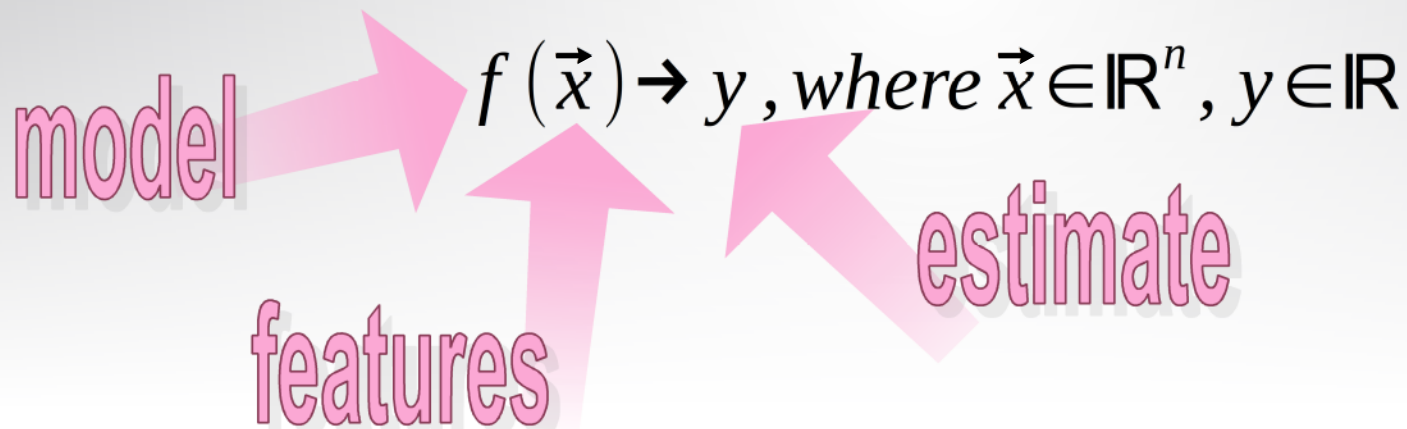
$$f(\vec{x}) \rightarrow y, \text{ where } \vec{x} \in \mathbb{R}^n, y \in \mathbb{R}$$

Regression Problems in Astronomy



analysis of **large catalogs** demands efficient **solving**

regression of **problems**

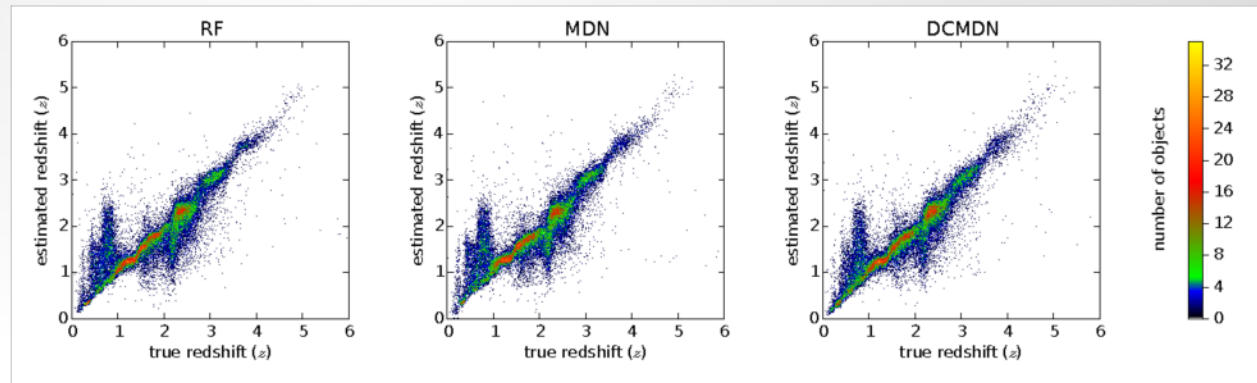
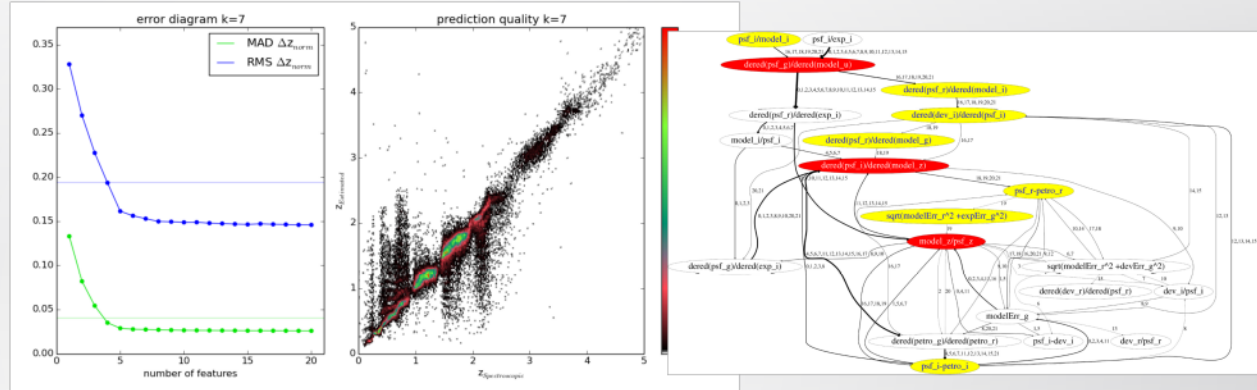
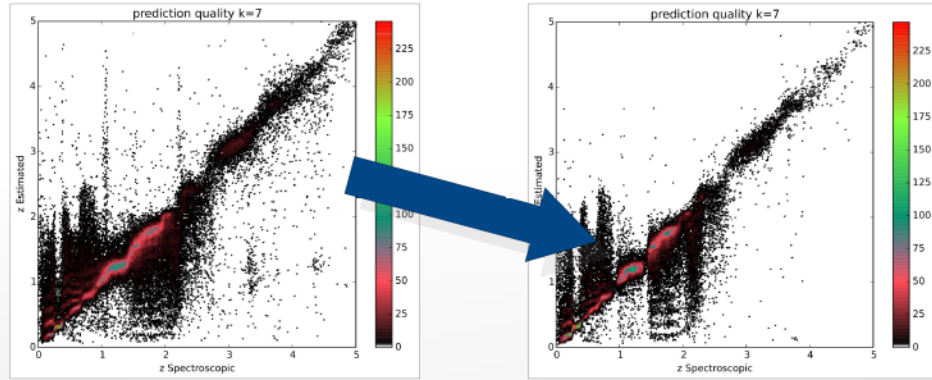


Improvements

data-sets /
missing data /
quality flags

feature selection /
metric selection

new / improved
algorithms



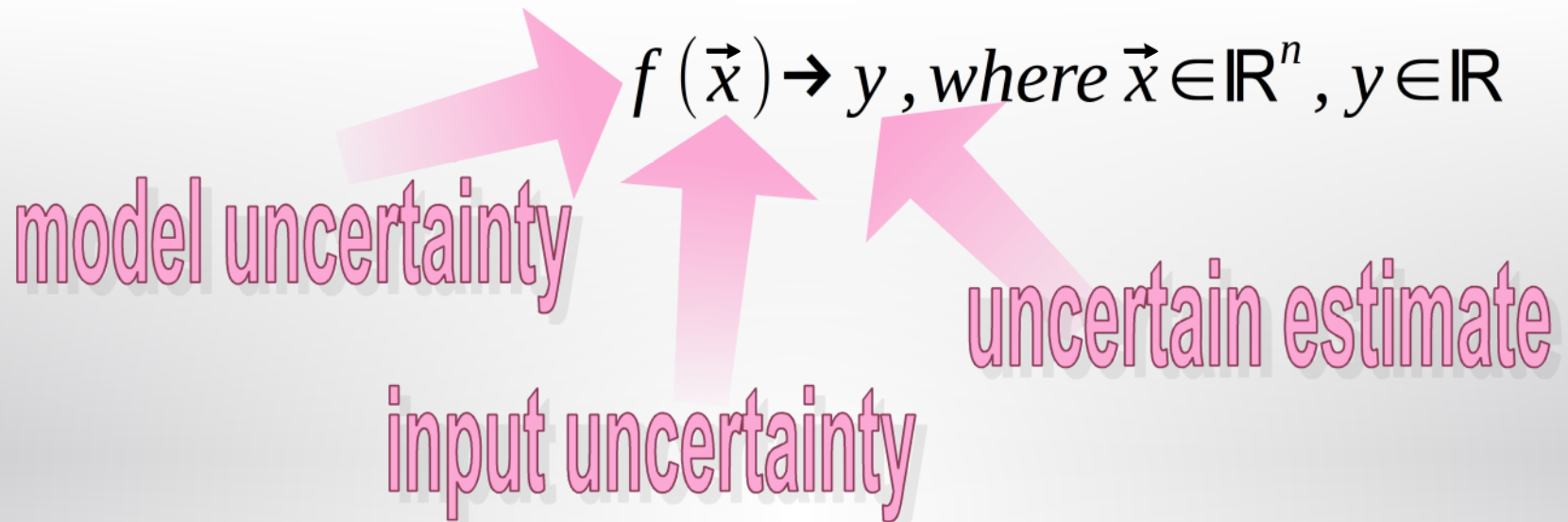
how about **uncertainties**?

Uncertainties

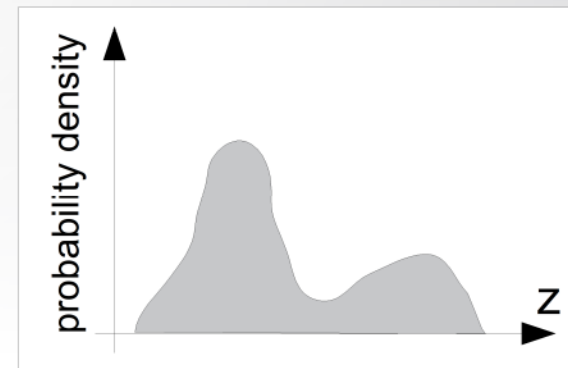
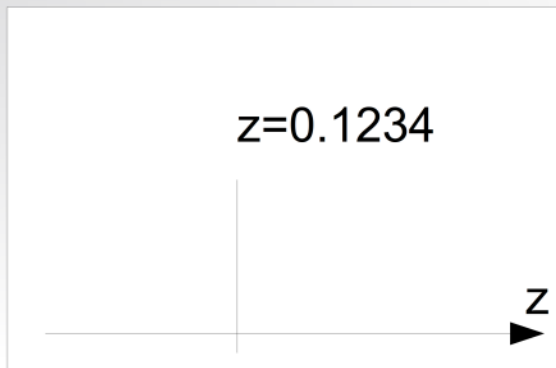
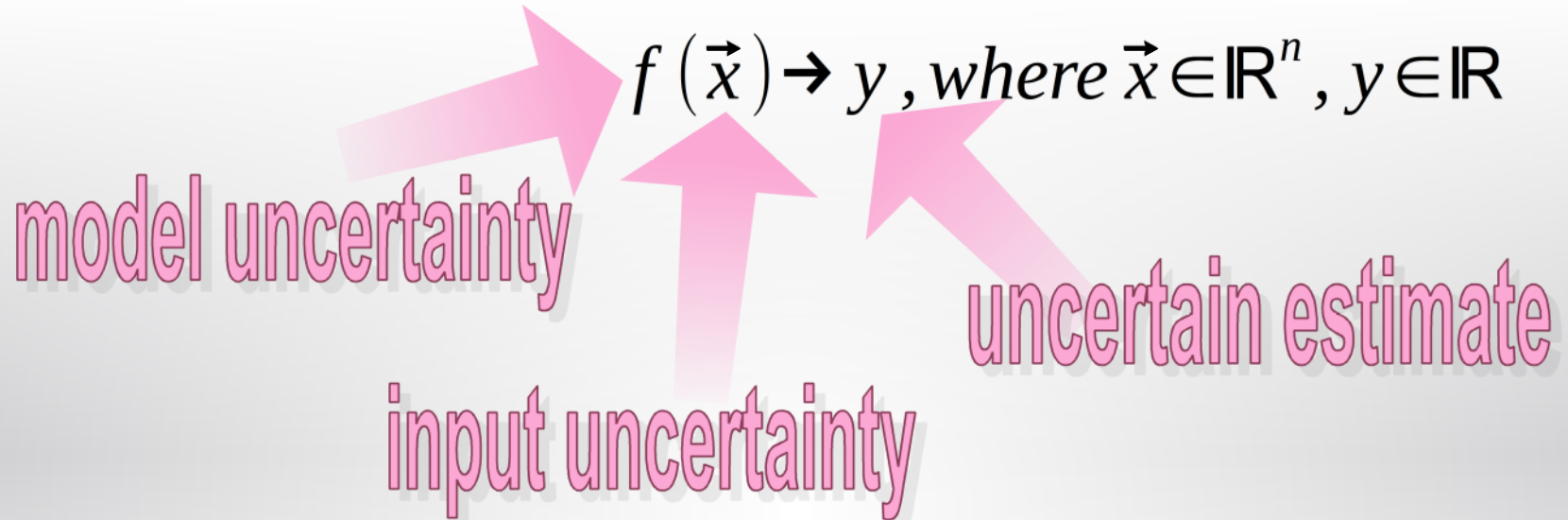


$$f(\vec{x}) \rightarrow y, \text{ where } \vec{x} \in \mathbb{R}^n, y \in \mathbb{R}$$

Uncertainties

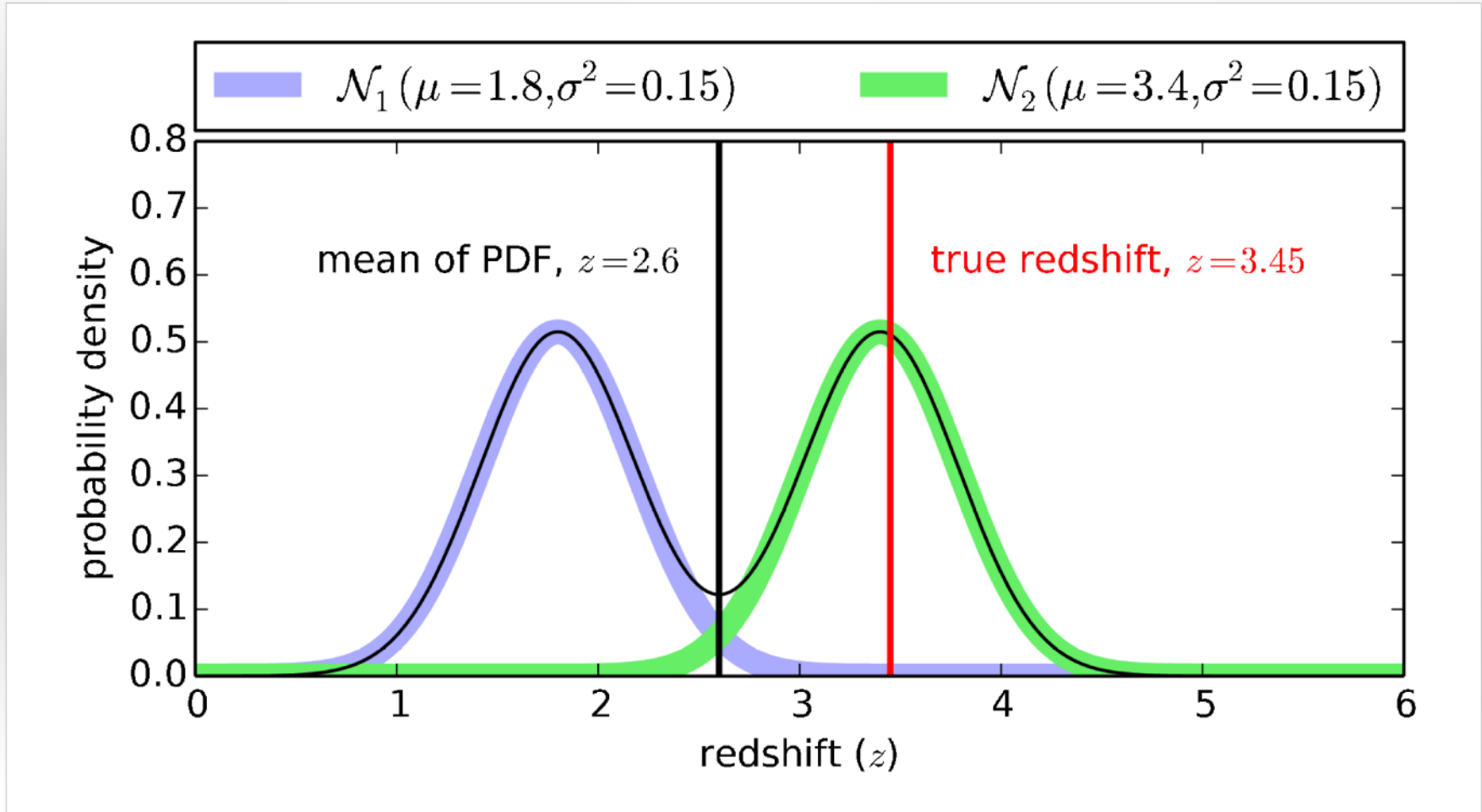


Uncertainties



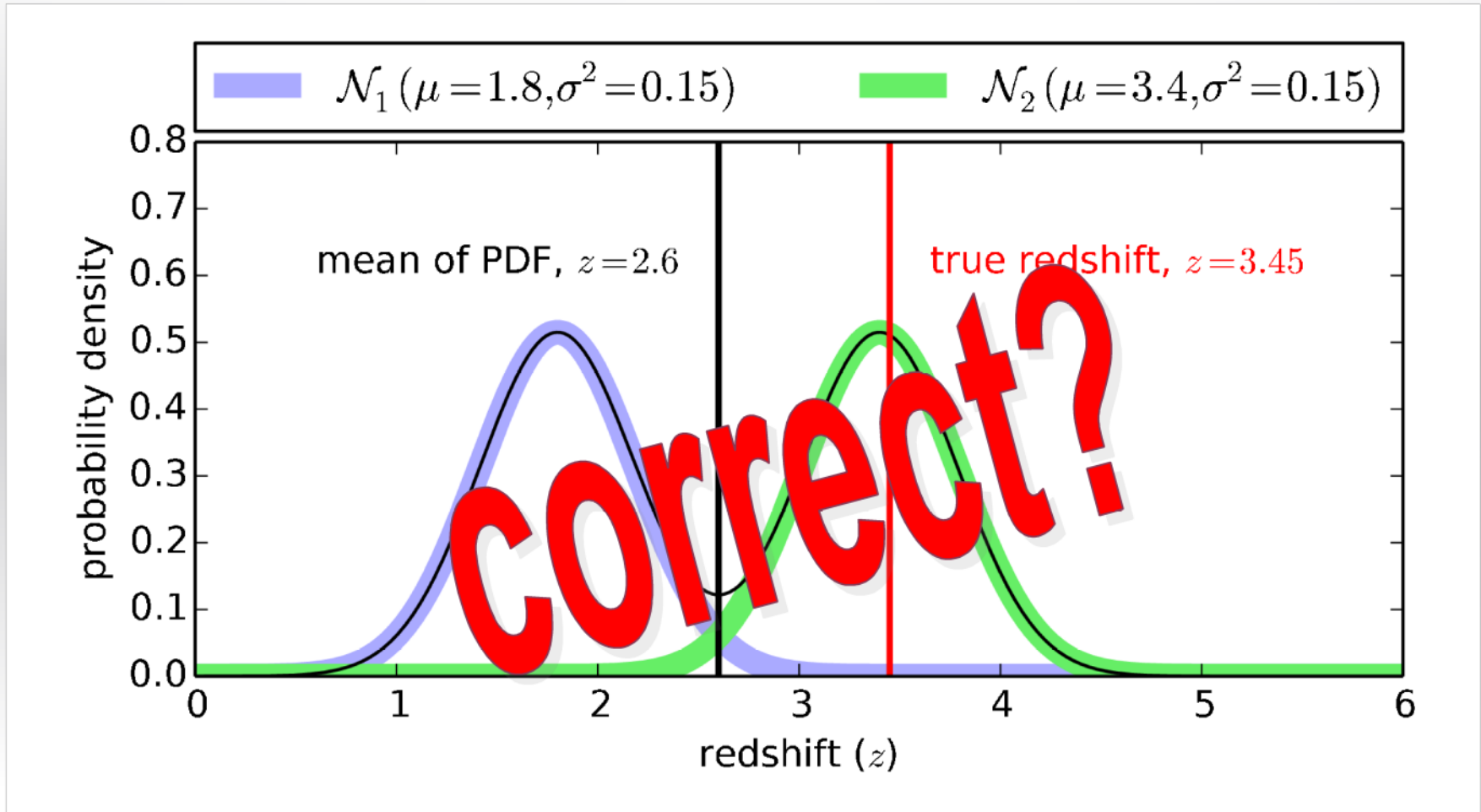
Evaluation Tools

simplification / just use the mean



Evaluation Tools

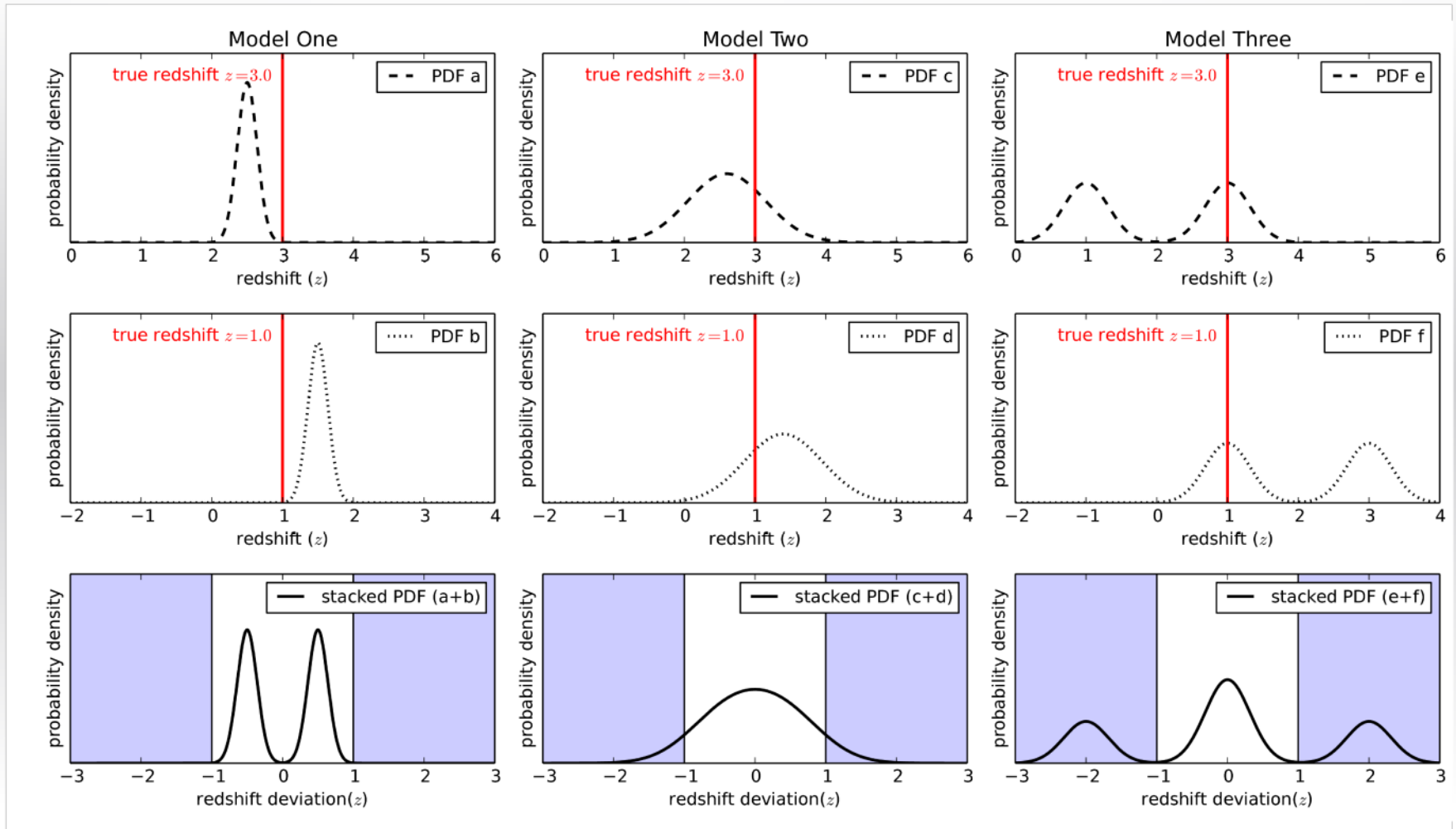
simplification / just use the mean



Evaluation Tools



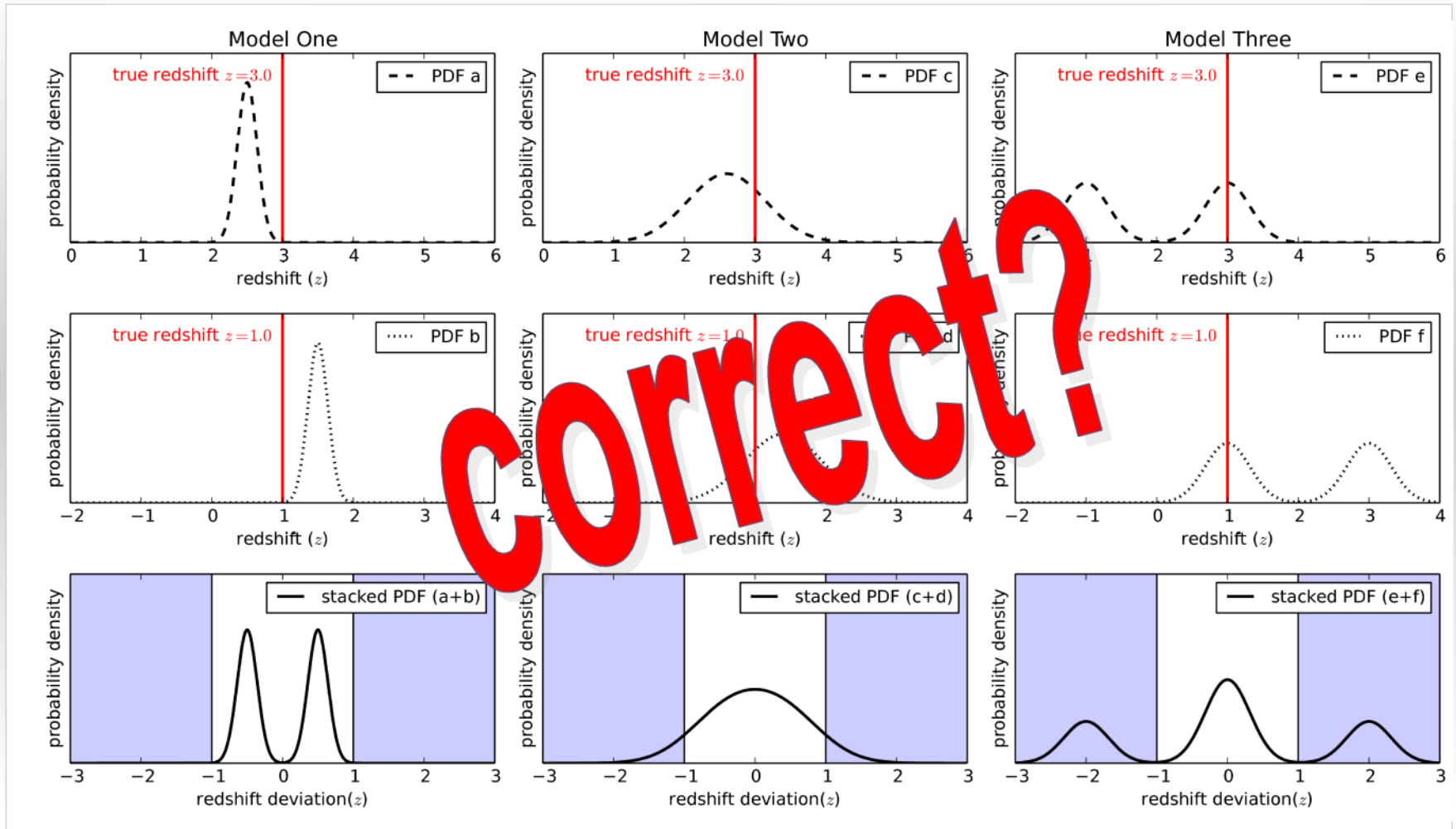
stacking PDFs



Evaluation Tools



stacking PDFs



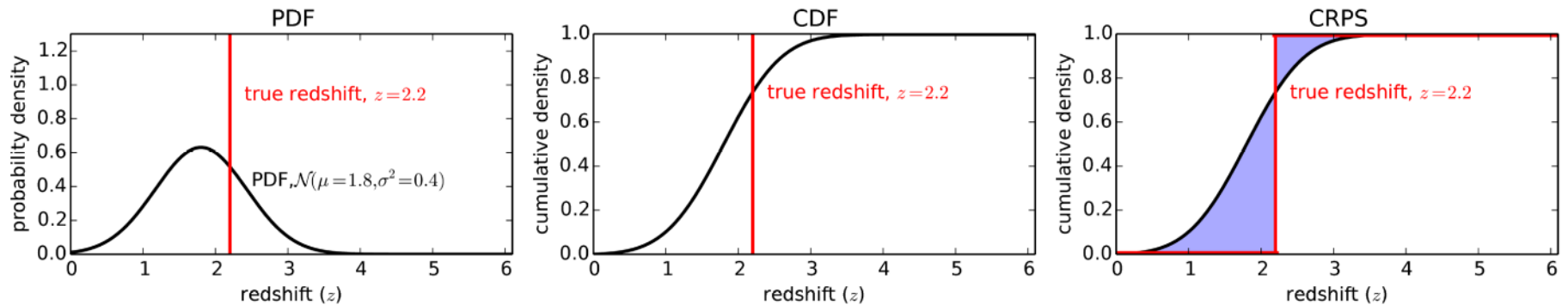
Proper Evaluation Tools / CRPS



continuous ranked probability score

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

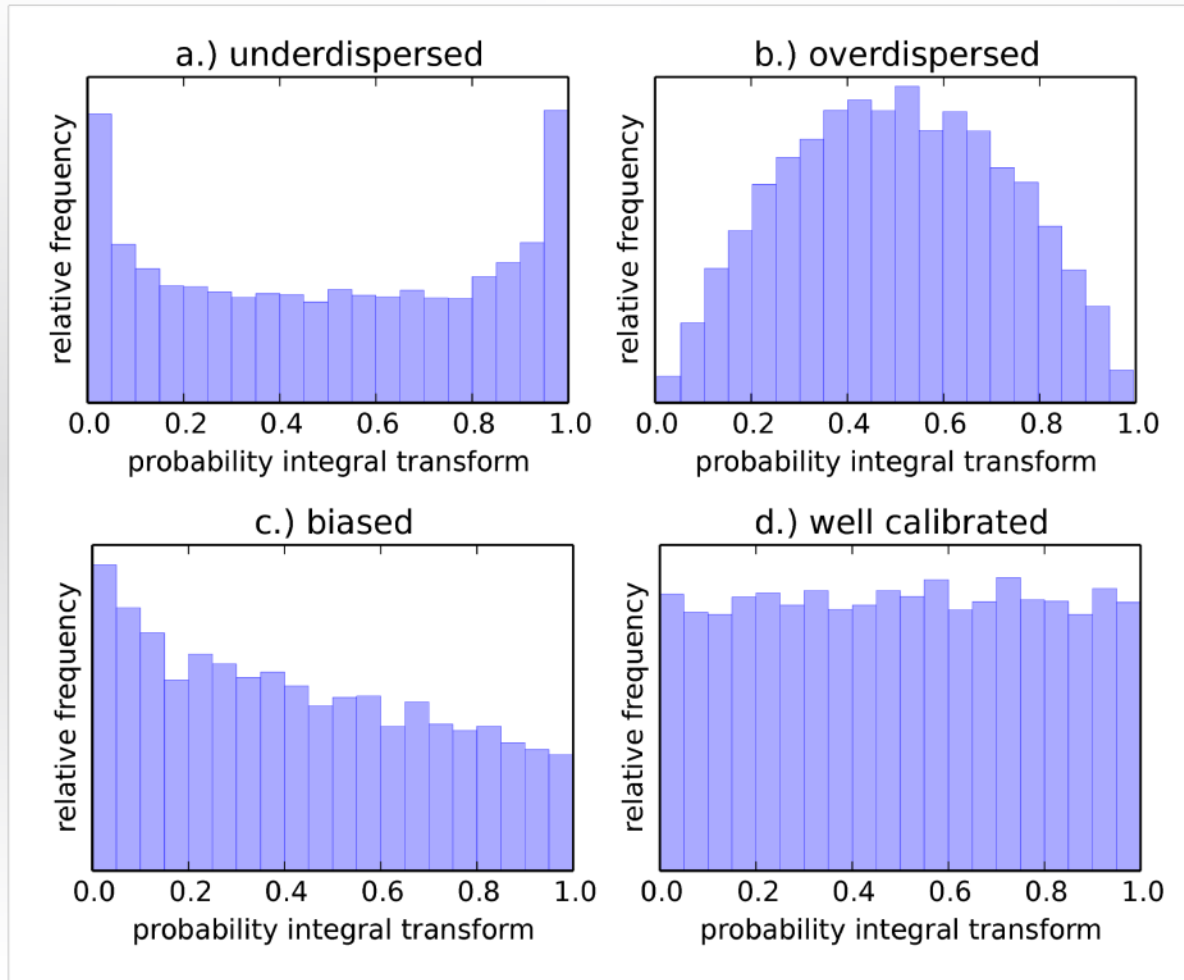
with $crps(CDF_t, z_t) = \int_{-\infty}^{+\infty} [CDF_t(z) - CDF_{z_t}(z)]^2 dz$



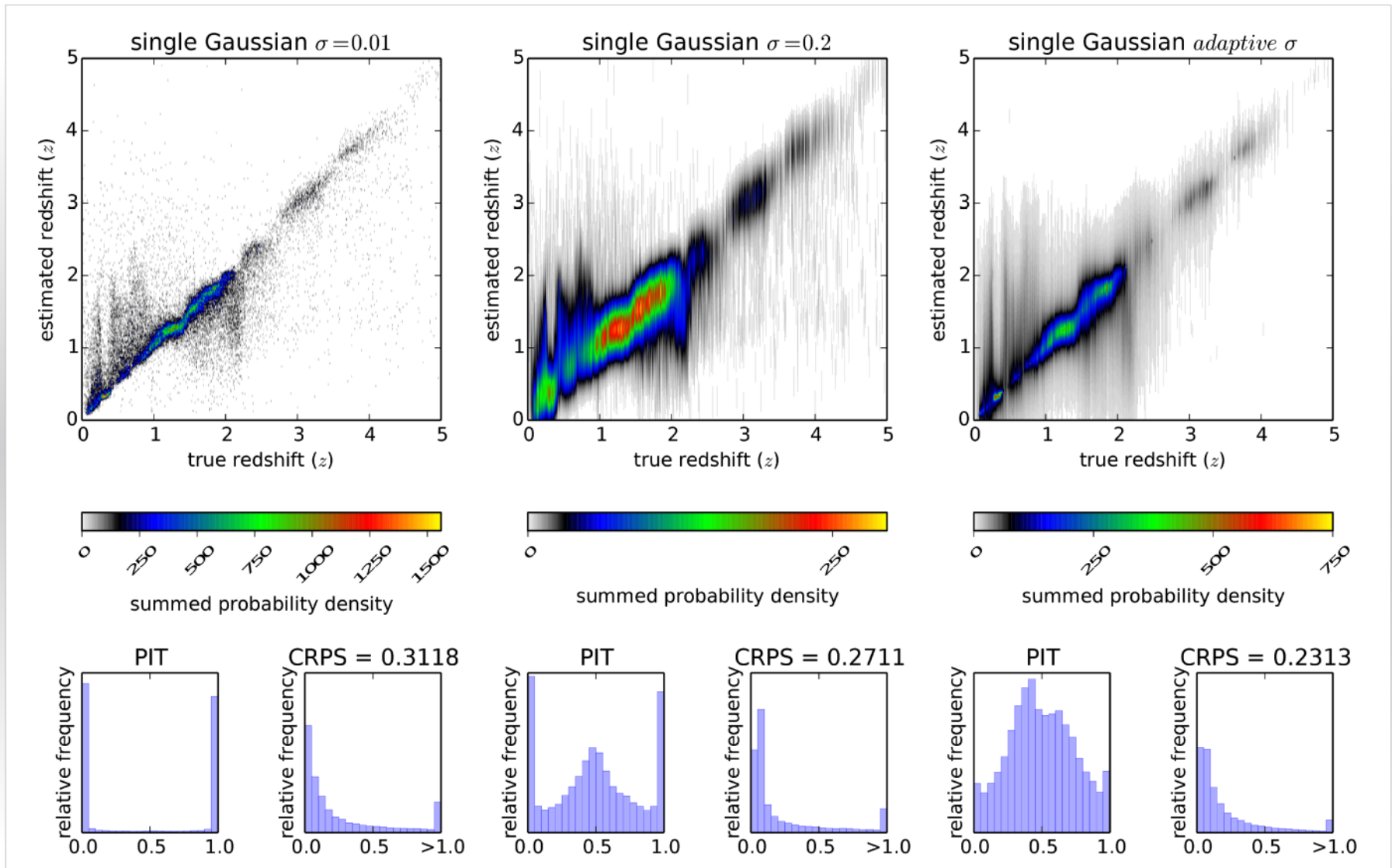
Proper Evaluation Tools / PIT



probability integral transform



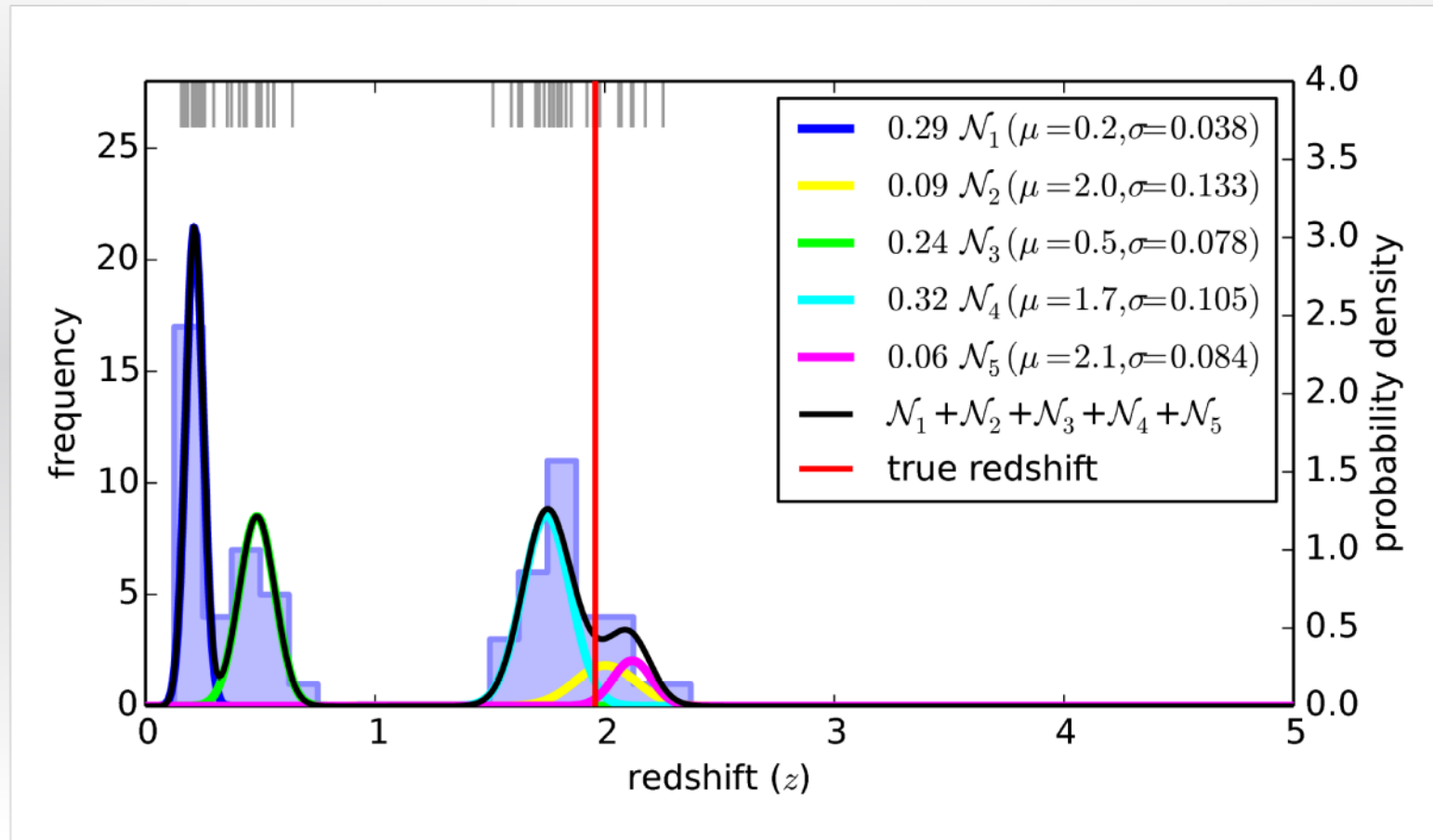
Results



Multi-Modalities



128 nearest neighbors inspected

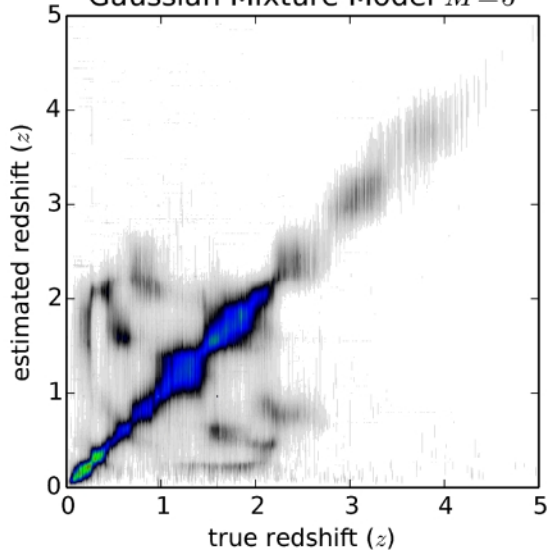


Results



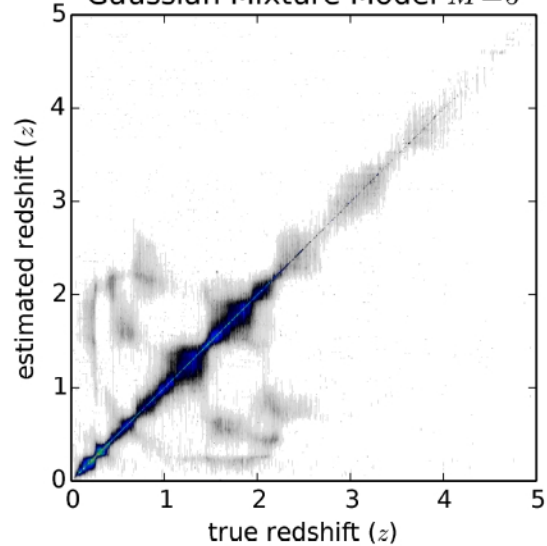
Nearest Neighbors

Gaussian Mixture Model $M=5$



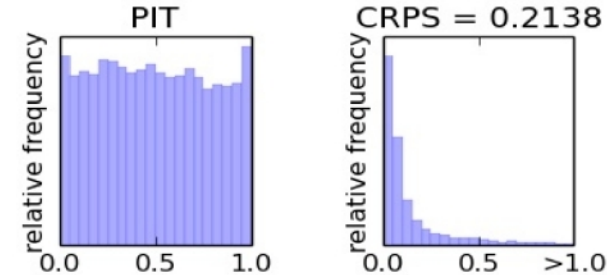
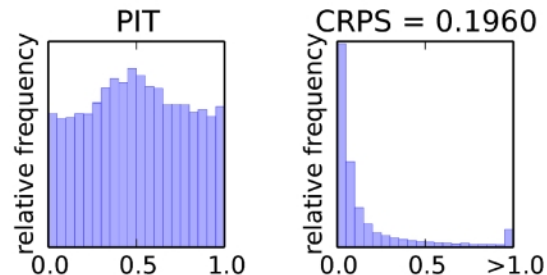
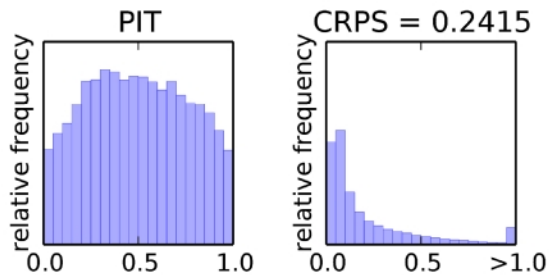
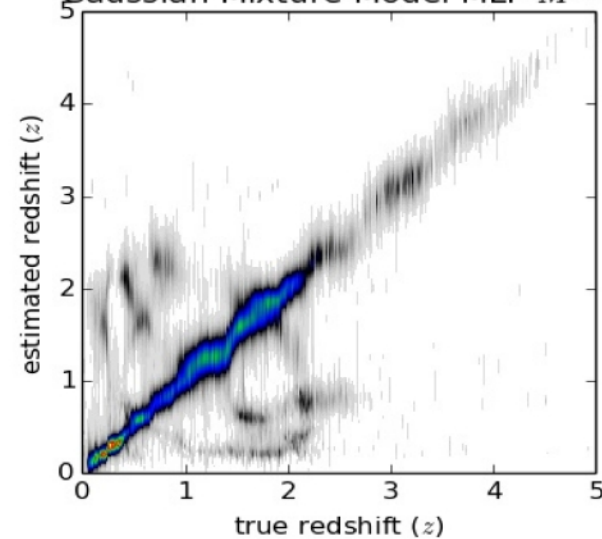
Random Forest

Gaussian Mixture Model $M=5$



Mixture Density Network

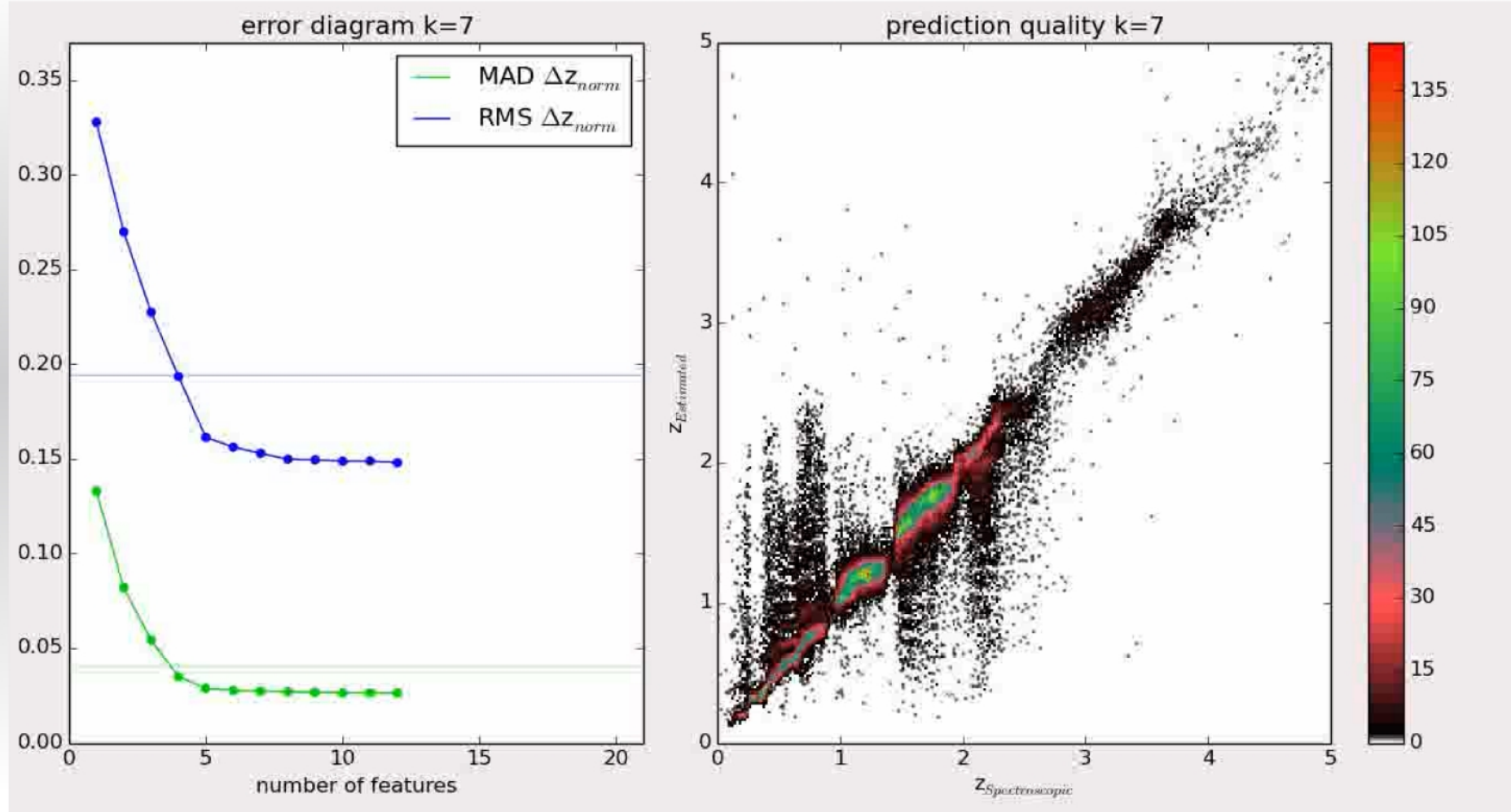
Gaussian Mixture Model MLP $M=5$



Forward Selection



apply greedy forward selection

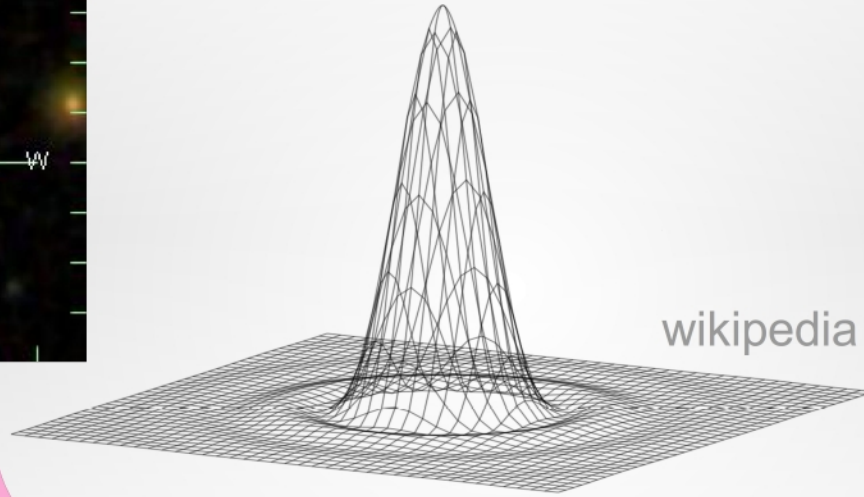
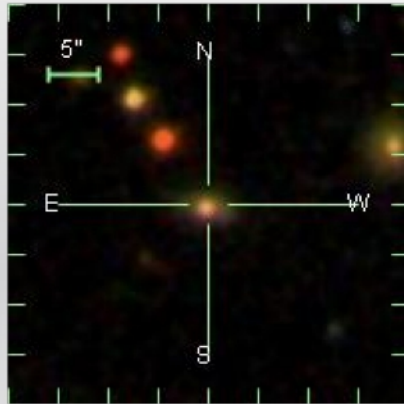


Lessons Learned



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

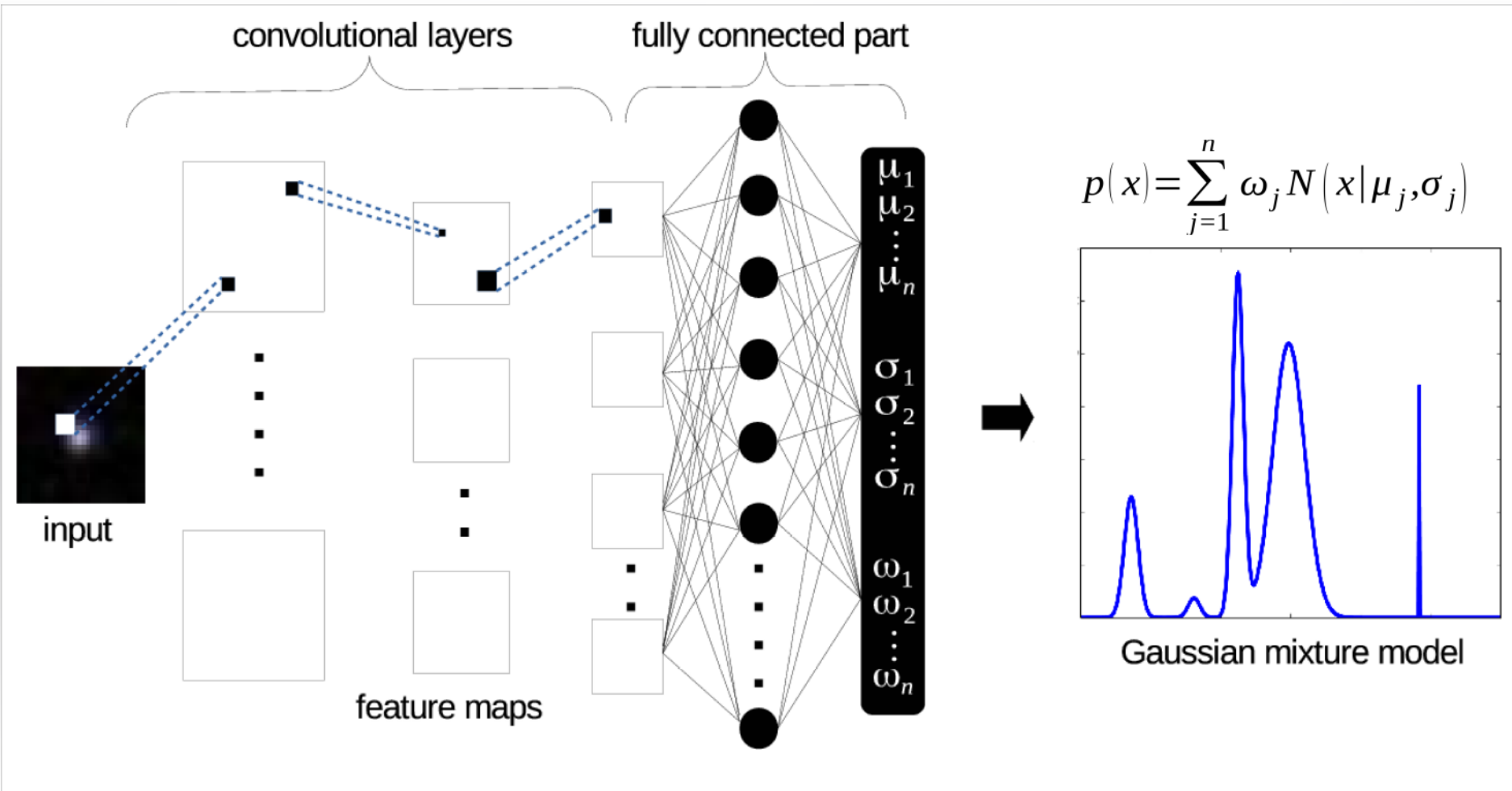
learning
machine



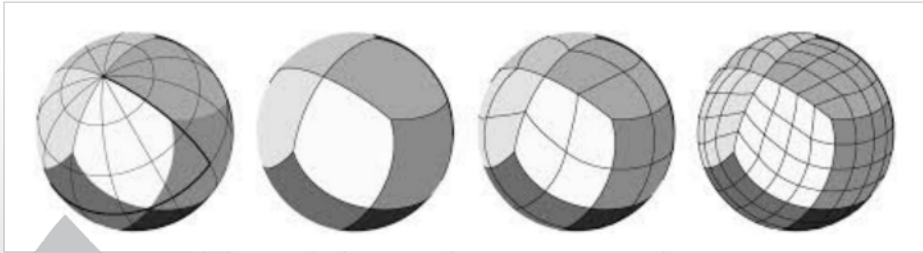
magnitudes
in (u, g, r, i, z)



Deep Convolutional Mixture Density network



Healpix / HiPS / IVOA



J2000 23 31 7.416 -54 54 31.56

Base image layer

WTF

Color map:
nati Reverse

Overlay layers

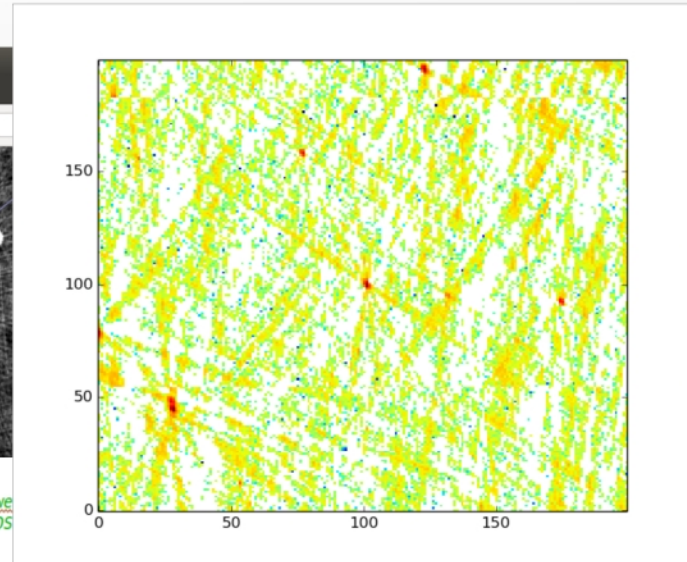
Reticle
 HEALPix grid

Tools

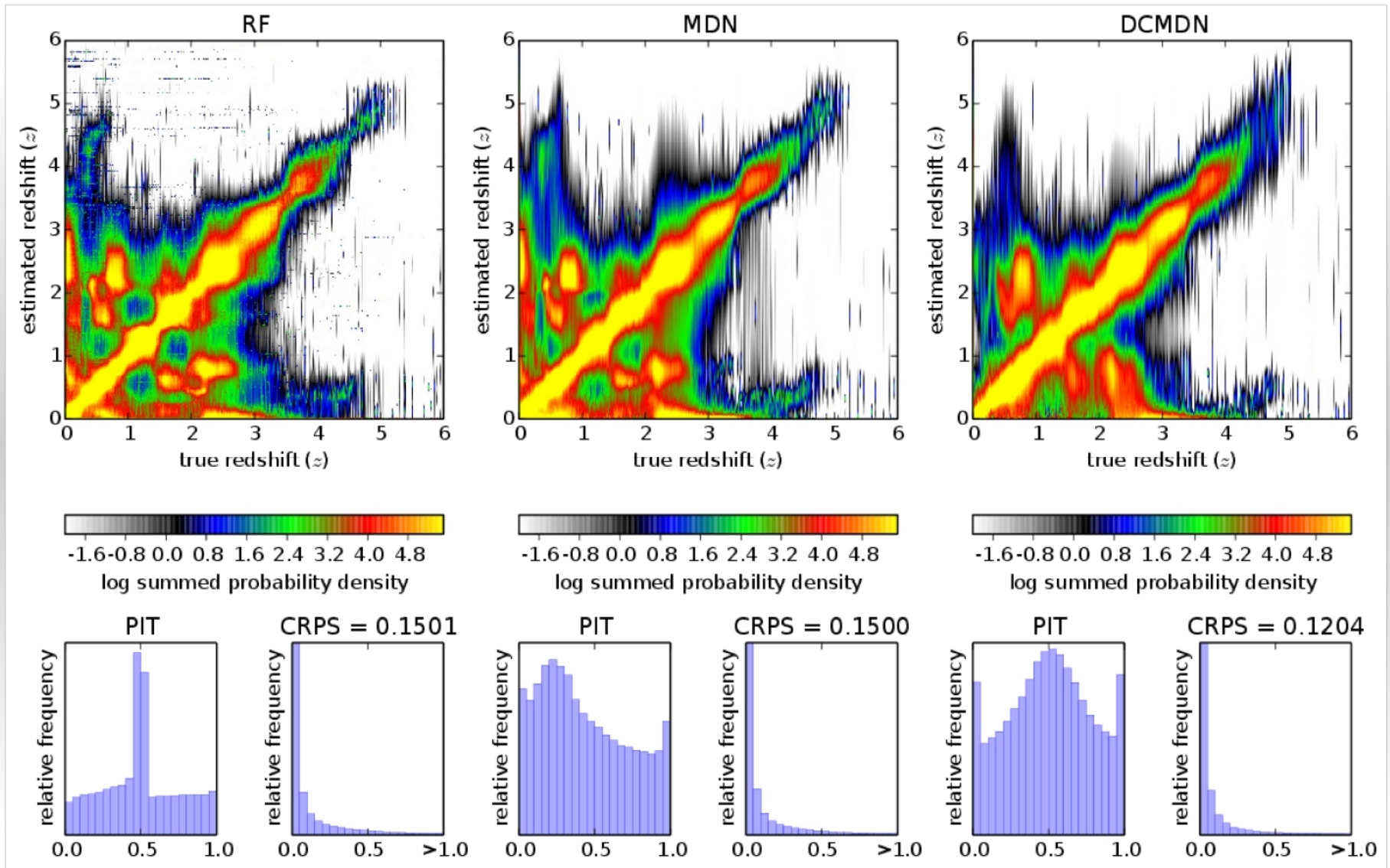
Export view as PNG

```
230 if __name__ == "__main__":  
231     ...surveyAddress = "atlas-spt-hips-2.s3-we  
232     ...surveyAddress = "alasky.u-strasbg.fr/DS  
233     ...  
234     ...coordinate = [350.86, -55.225]  
235     ...size = [200, 200]  
236     ...resolution = 0.002  
237     ...  
238     ...myHiPSfs = HiPSfs(surveyAddress) # create access!  
239     ...  
240     ...image = myHiPSfs.extractCoordinate(coordinate, size, resolution, nested=True) # extract data array!  
241     ...  
242     ...pyplot.figure()  
243     ...pyplot.imshow(image, aspect='auto', interpolation="nearest")  
244     ...pyplot.gca().invert_yaxis()  
245     ...pyplot.show()
```

9/2954221 9/2954222
9/2954220
9/2954214 9/2954217
FoV: 29.96'



Results / Classification + Regression



Conclusion



uncertainty
is very
important to further
improve
the science quality

we must start to use
proper
tools to
evaluate the performance

Thank you for your attention!

Volker
Gaibler

Nikos
Gianniotis

Dorotea
Dudas

Dennis
Kügler

Antonio
D'Isanto

+ Erica Hopkins
+ Jenny Wagner



@AstroInformatix