Statistical challenges for large-scale structure in the era of LSST

IfA Edinburgh

Photometric redshifts by machine learning: Don’t underestimate your scatter & bias

Chieh-An Lin (Linc)

April 19th, 2018

University Oxford, UK

On behalf of

the COIN collaboration

Target model

↓ Training set — With spec-z

Parameters

↓ Testing set — With spec-z

Performance
(scatter & bias)

↓ Data — Without spec-z

Prediction

Photometric redshifts
by machine learning

Target model

Training set — Bright

Parameters

Testing set — Bright

Performance (scatter & bias)

Data — Faint

Prediction

Photometric redshifts
by machine learning

Question

Will different error properties between the bright & faint samples bias the estimates?

Target model

Training set

Set A - SDSS DR12 spec

Parameters

Testing set

Set B - SDSS DR12 spec

Set C - Intermediate case

Set D - SDSS DR12 photo

Performance (scatter & bias)

Data

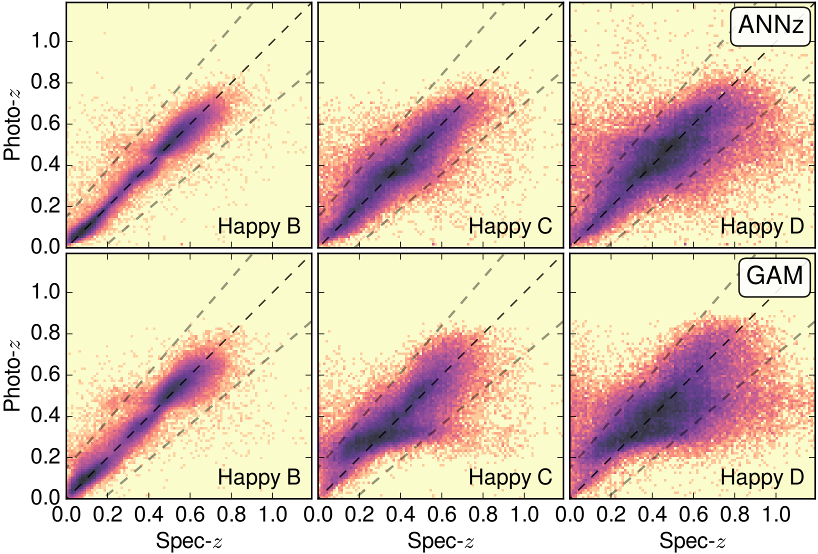
Prediction

Let's design some tests:

The Happy catalogue

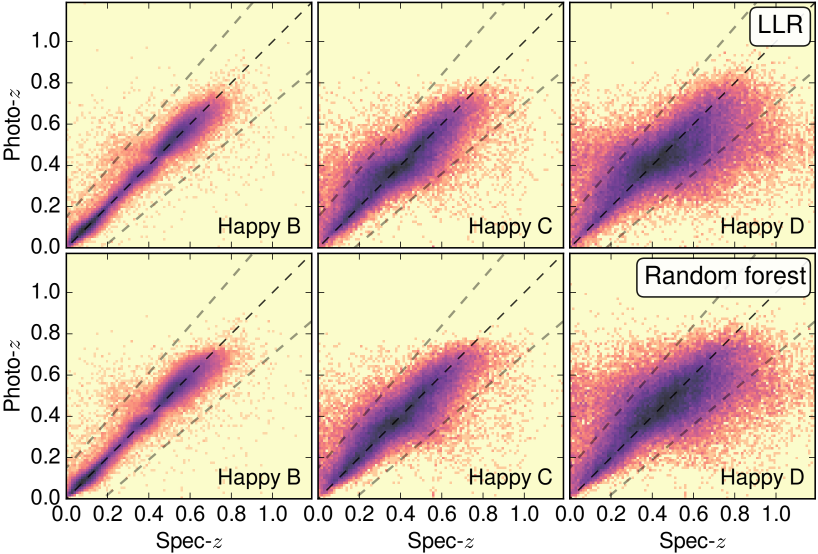
Beck, Lin, et al. (2017)

Oops

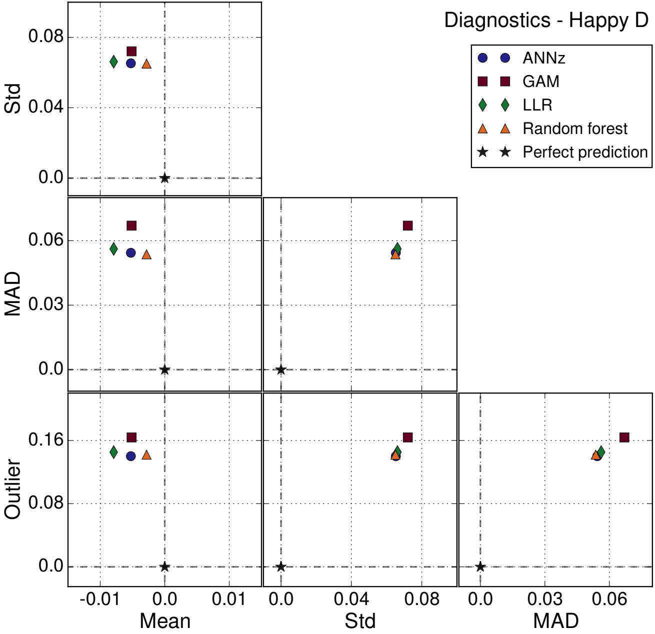


Beck, Lin, et al. (2017)

Ouch



Diagnostics



$$b \equiv \frac{z_{\rm photo} - z_{\rm spec}}{1 + z_{\rm spec}}$$

MAD $= \mathrm{median}(|b|)$

Outliers: $|b| > 0.15$

Beck, Lin, et al. (2017)

Don’t underestimate your scatter & bias

COIN

Cosmostatistics Initiative

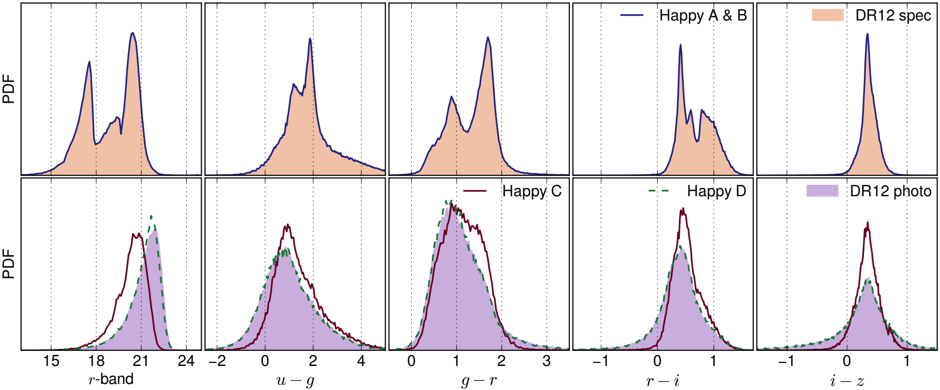
Catalogues available on GitHub

https://github.com/COINtoolbox/photoz_catalogues

Backup slides

Beck, Lin, et al. (2017)

Color distributions



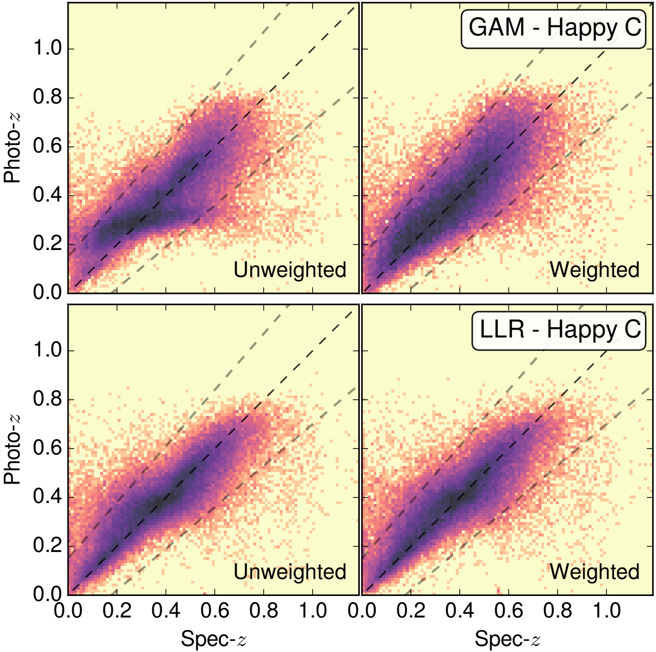
Diagnostics

Method	Set	Mean	Std	MAD	Outliers
ANNz	B	0.04	2.87	1.49	0.99
	C	0.16	5.41	3.60	5.59
	D	-0.52	6.53	5.44	14.01
GAM	B	0.09	3.50	1.95	1.36
	C	0.86	6.34	4.84	7.37
	D	-0.51	7.21	6.70	16.38
LLR	B	0.13	2.81	1.39	1.11
	C	0.52	5.45	3.59	6.07
	D	-0.79	6.62	5.62	14.52
Random forest	B	0.05	2.82	1.41	1.02
	C	0.34	5.39	3.51	5.58
	D	-0.28	6.51	5.36	14.2

Units: 10^{-2}

Extended spectroscopic samples

Survey	References	Number of matches
2dF	Colless et al. (2001, 2003)	770
6dF	Jones et al. (2004, 2009)	765
DEEP2	Davis et al. (2003); Newman et al. (2013)	7456
GAMA	Driver et al. (2011); Baldry et al. (2014)	53373
PRIMUS	Coil et al. (2011); Cool et al. (2013)	32459
VIPERS	Garilli et al. (2014); Guzzo et al. (2014)	18967
VVDS	Le Fèvre et al. (2004); Garilli et al. (2008)	8381
WiggleZ	Drinkwater et al. (2010); Parkinson et al. (2012)	43874
zCOSMOS	Lilly et al. (2007, 2009)	2789
Total		168834



Reweighting does not cure