ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

# Ancillary information from photometric redshifts for calibrating follow-up observations

Matías Carrasco Kind

Department of Astronomy
University of Illinois

DES collaboration meeting LBNL April 8 - 12, 2013

# DES SV Data

VVDS Deep 02hr (3117)

CDFS (582)

VVDS Wide 14hr (2970)

COSMOS (688)

Secure redshifts $3 \leq \mathrm{ZFLAG} \leq 4$

$0.01 \leq Z \leq 1.5$

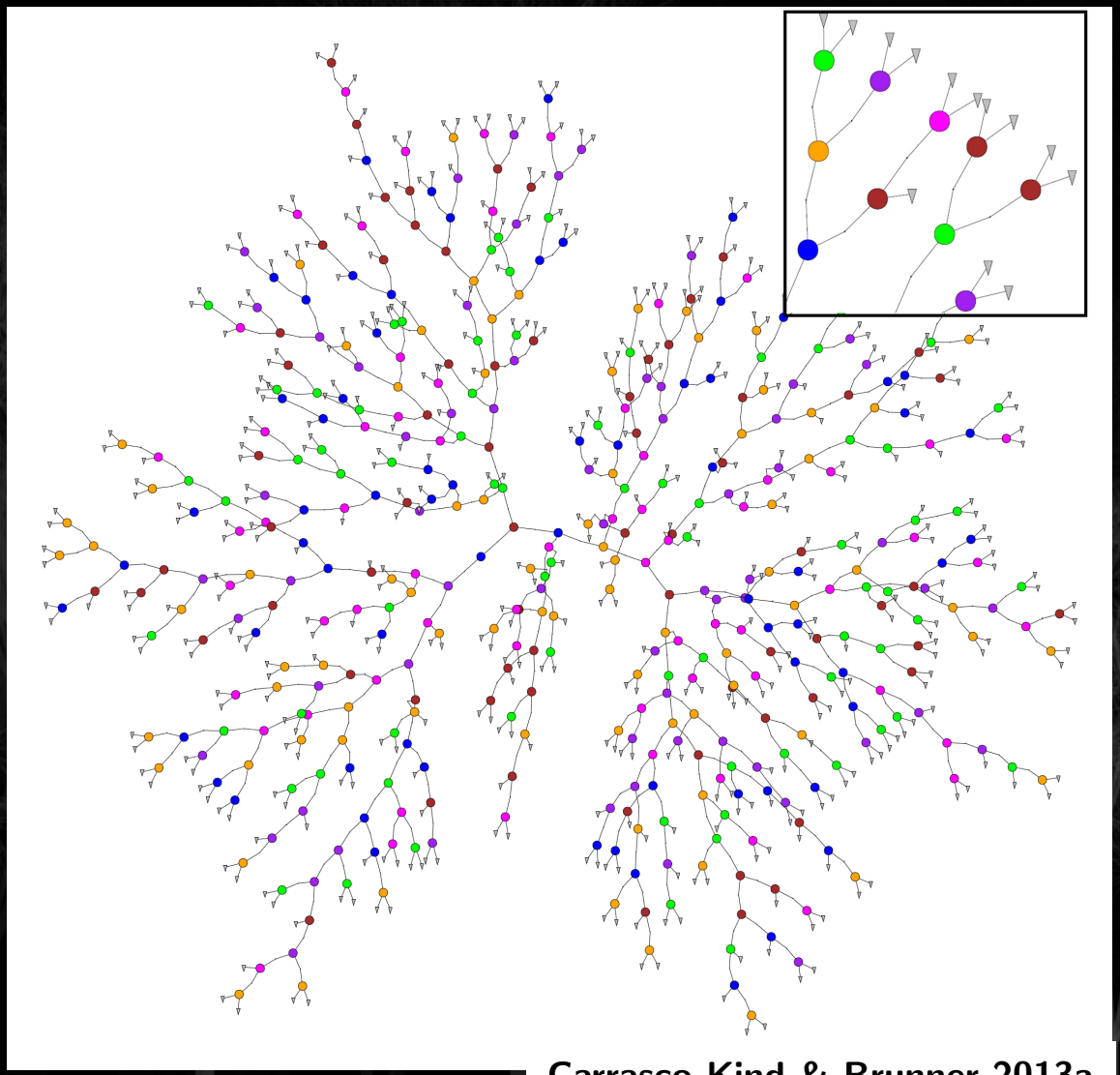
Use all 7357 galaxies

Use 5 bands from MAG_AUTO and MAG_DETMODEL

Use colors (8)

# TPZ: Trees for Photo-Z

- Provides photo-$z$ PDF and confidence values
- Deals with missing data
- Includes measurements errors
- Provides useful ancillary information
- *Out-of-Bag* data for unbiased errors
- No need for validation set



**Carrasco Kind & Brunner 2013a**

http://lcdm.astro.illinois.edu/research/TPZ.html

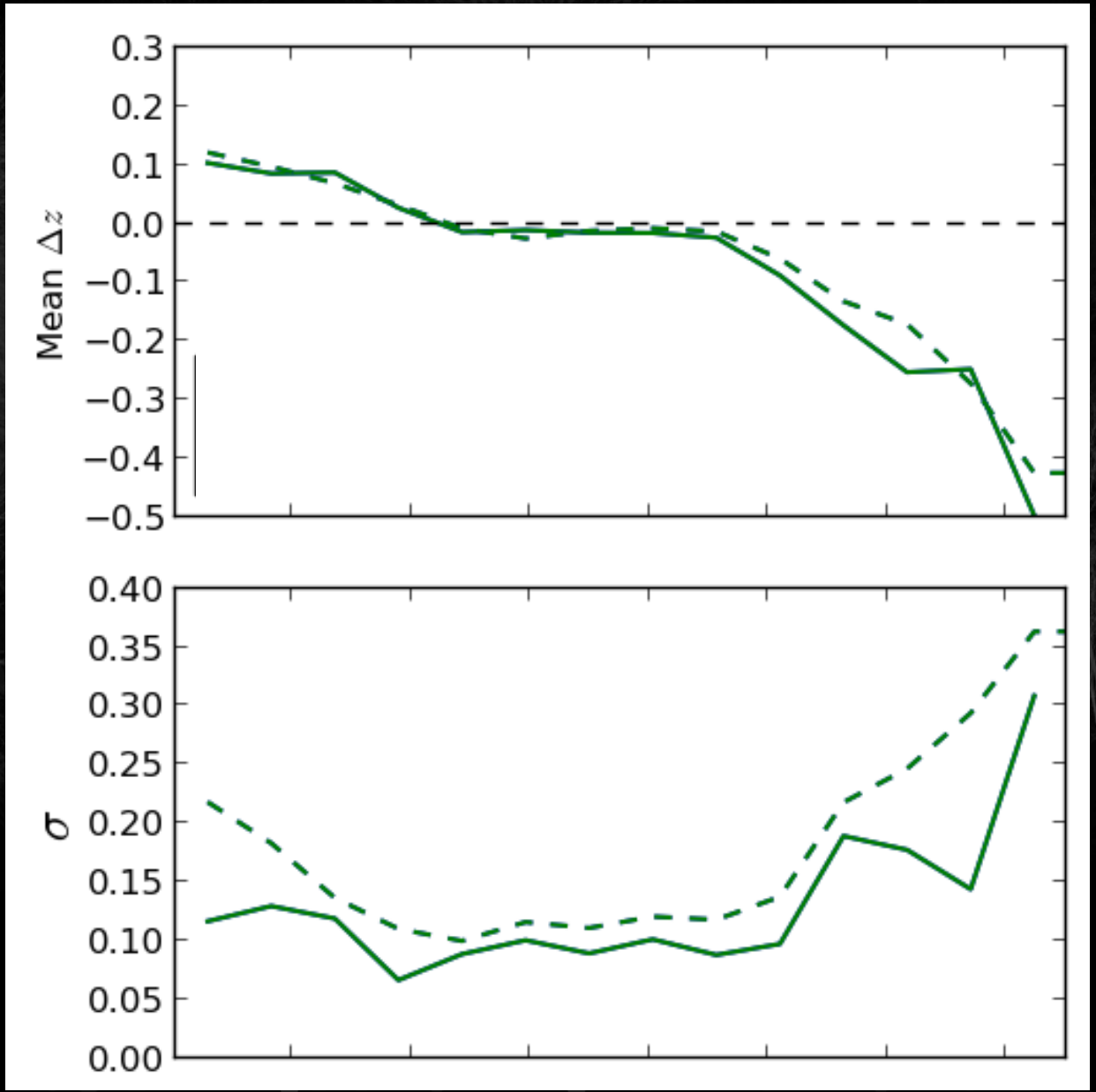
# TPZ: Ancillary information - *prior error* -

Using *Out-of-Bag* data TPZ provides useful extra information

No need of a validation set, use full training set.

Example from DES SV data with testing

A prior unbiased estimations of errors!



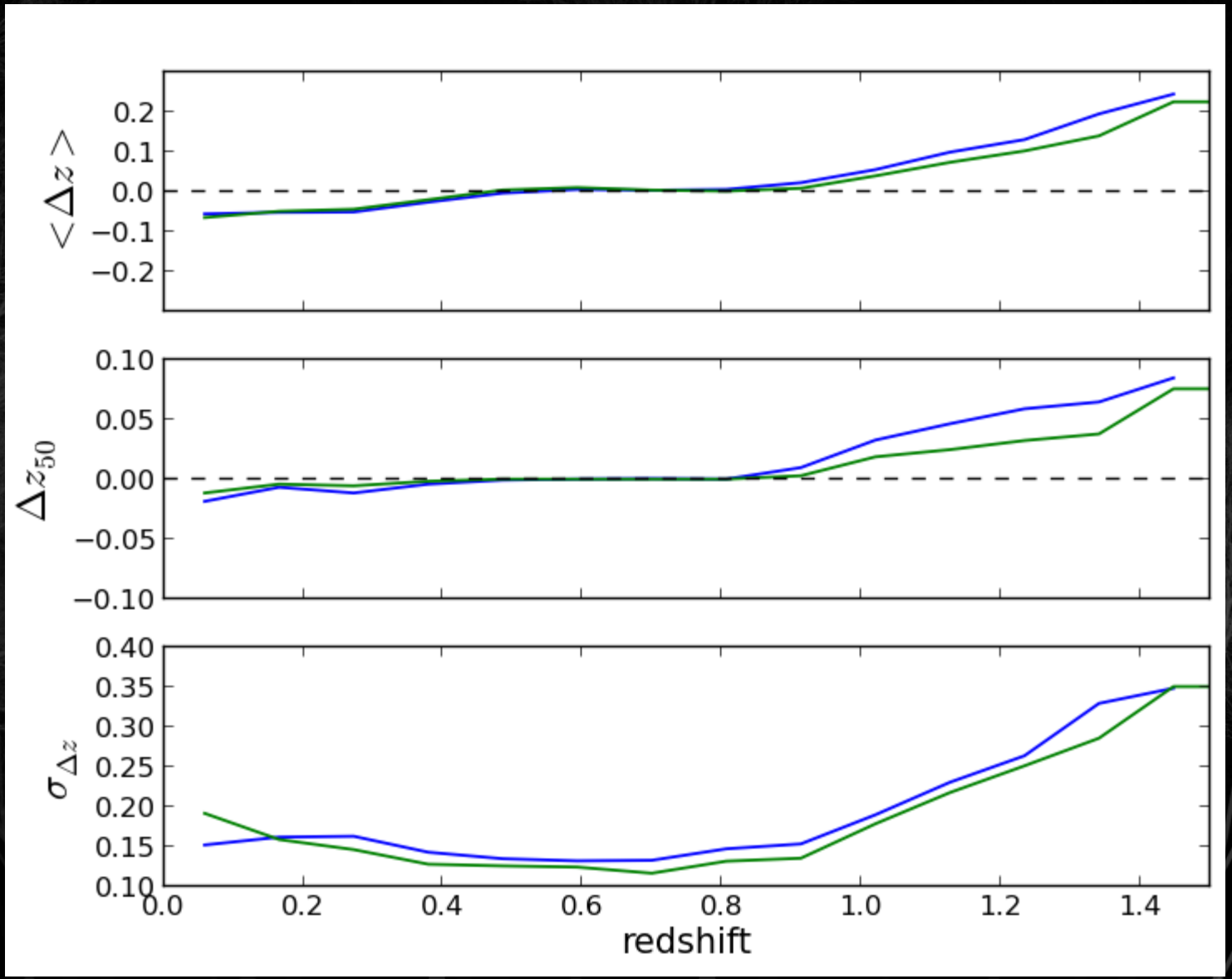
# Application on DES SV data: OOB errors

Prior metrics before real data

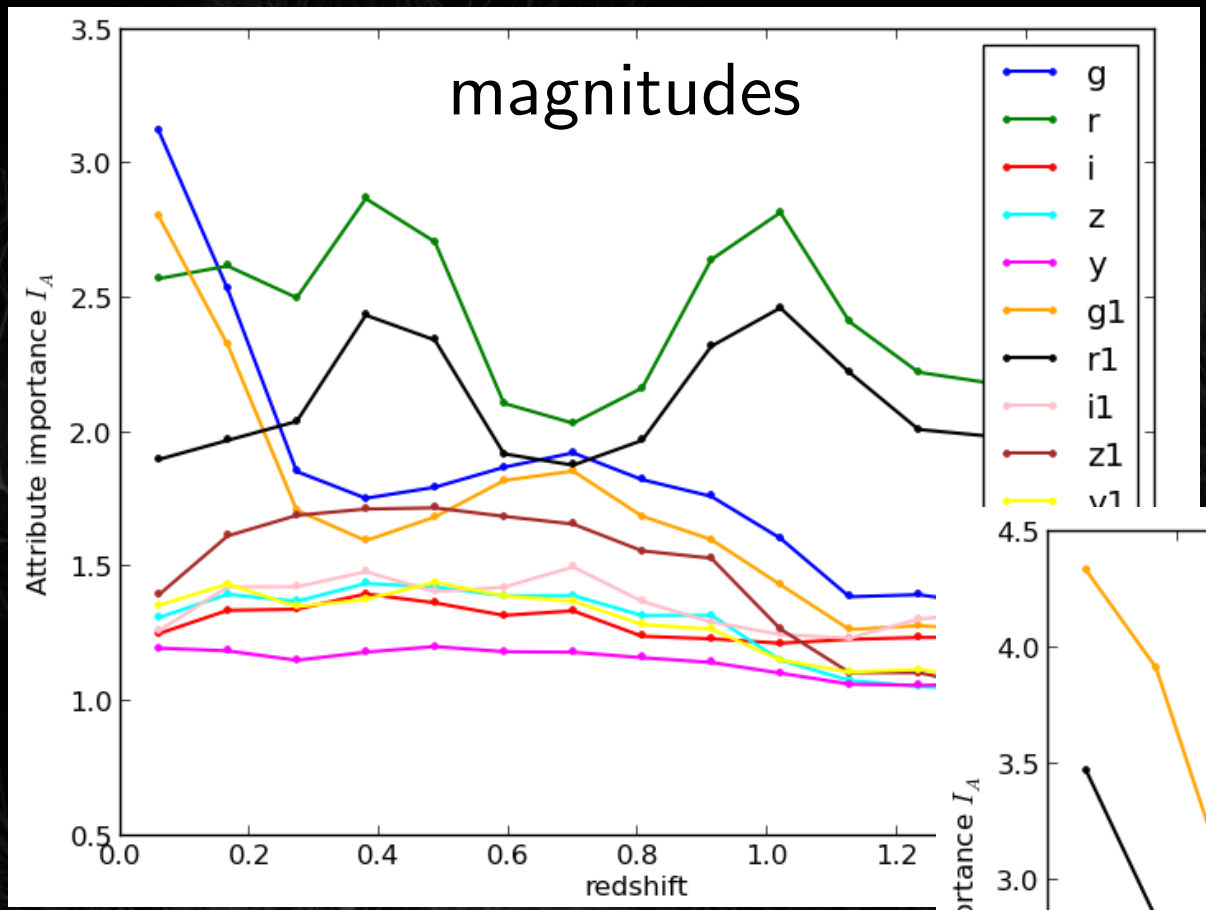
"Upper" limits

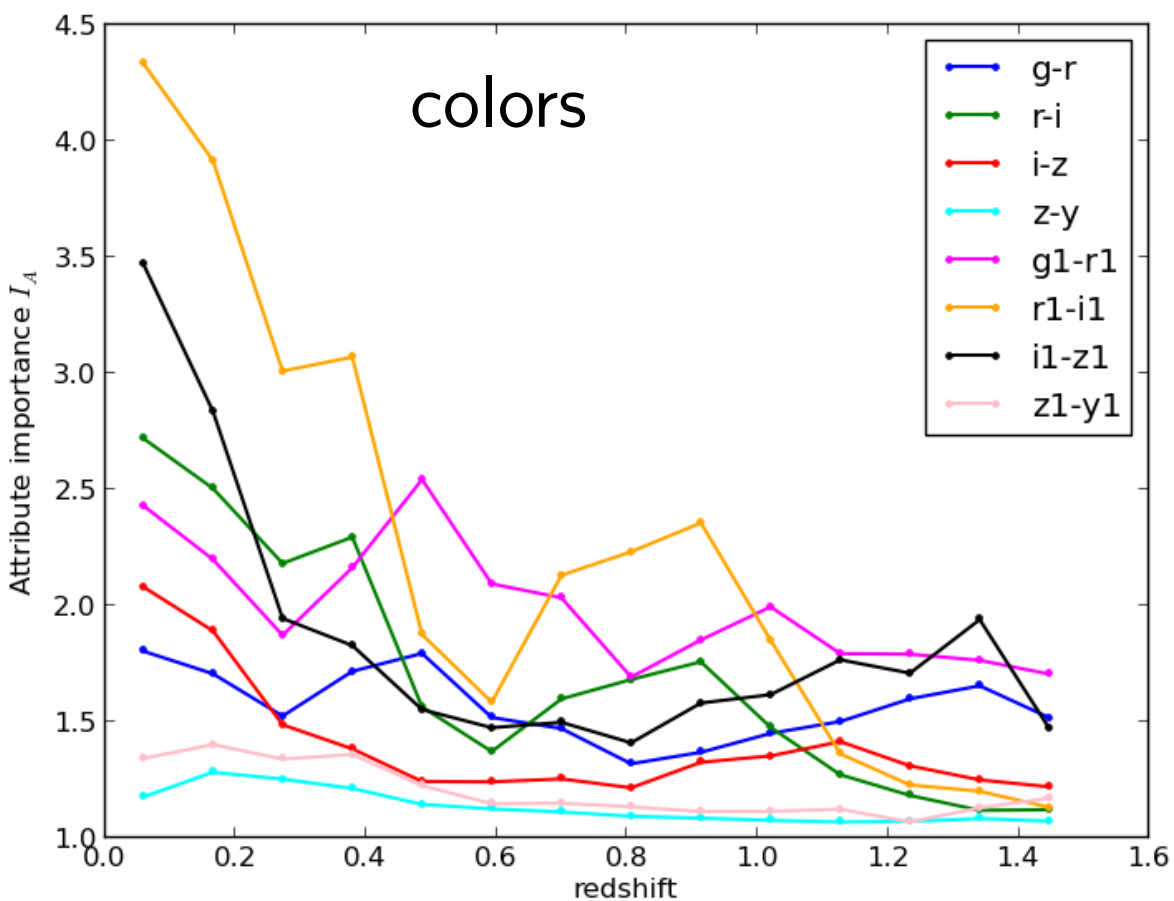
Using ALL available data.

Colors will make better predictions



# Application on DES SV data: Importance

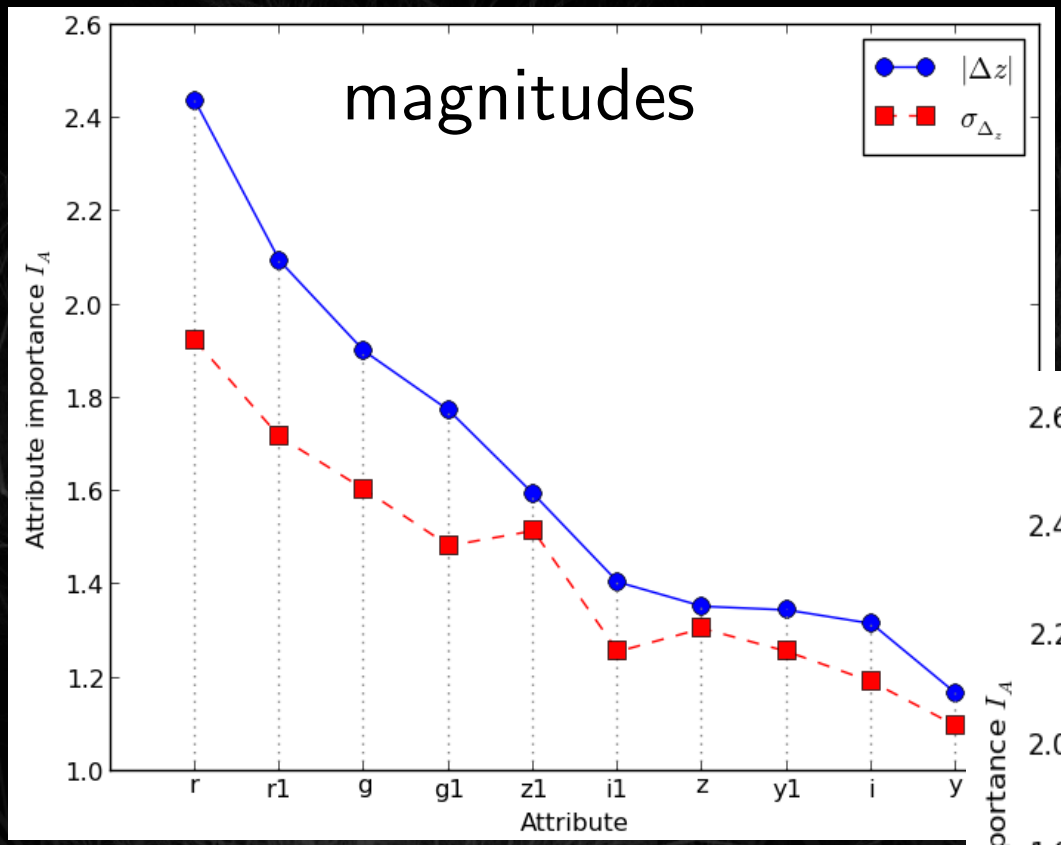


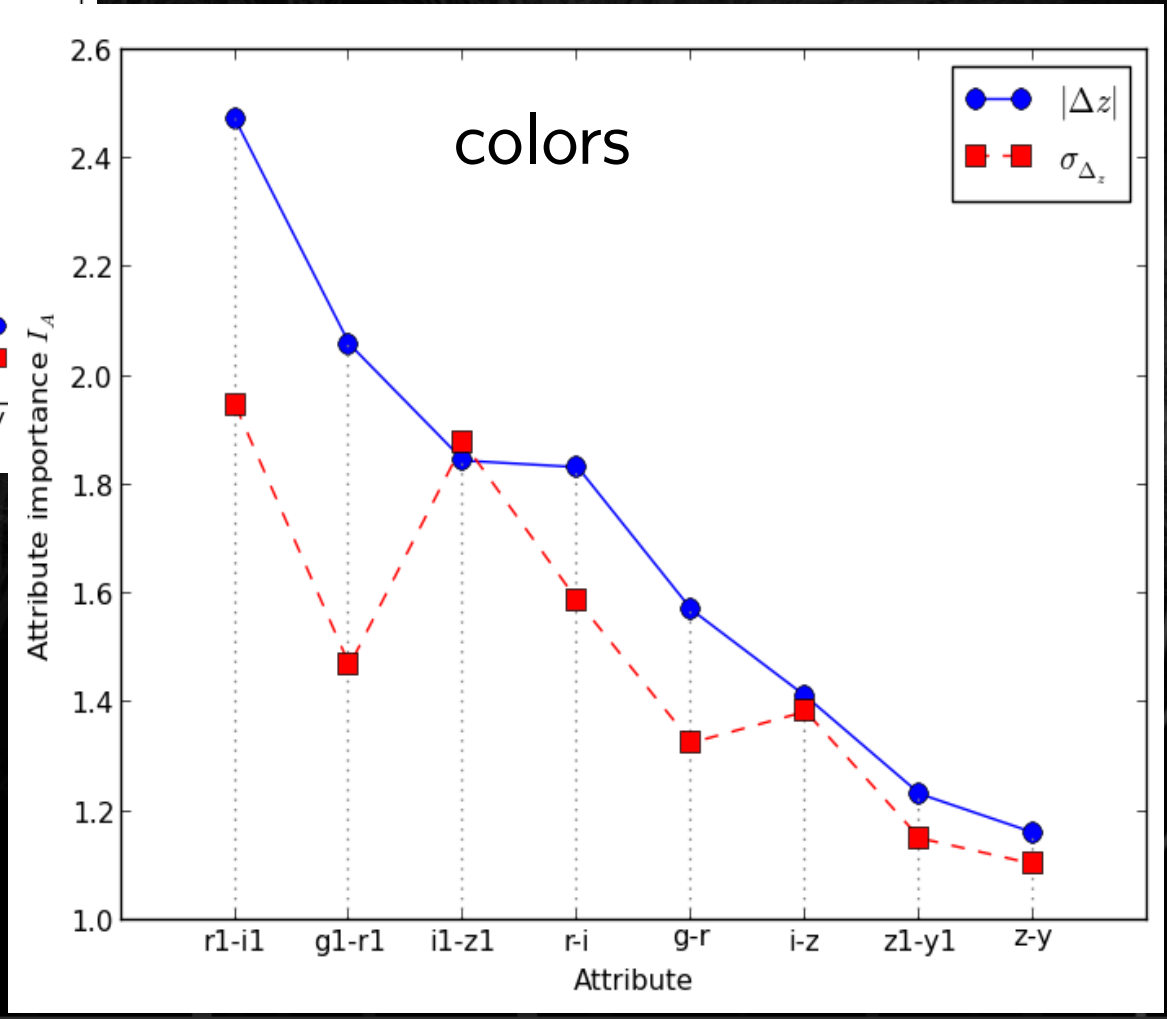

$g, r, i, z, y$ (AUTO)

$g1, r1, i1, z1, y1$ (DET_MODEL)

# Application on DES SV data: Importance

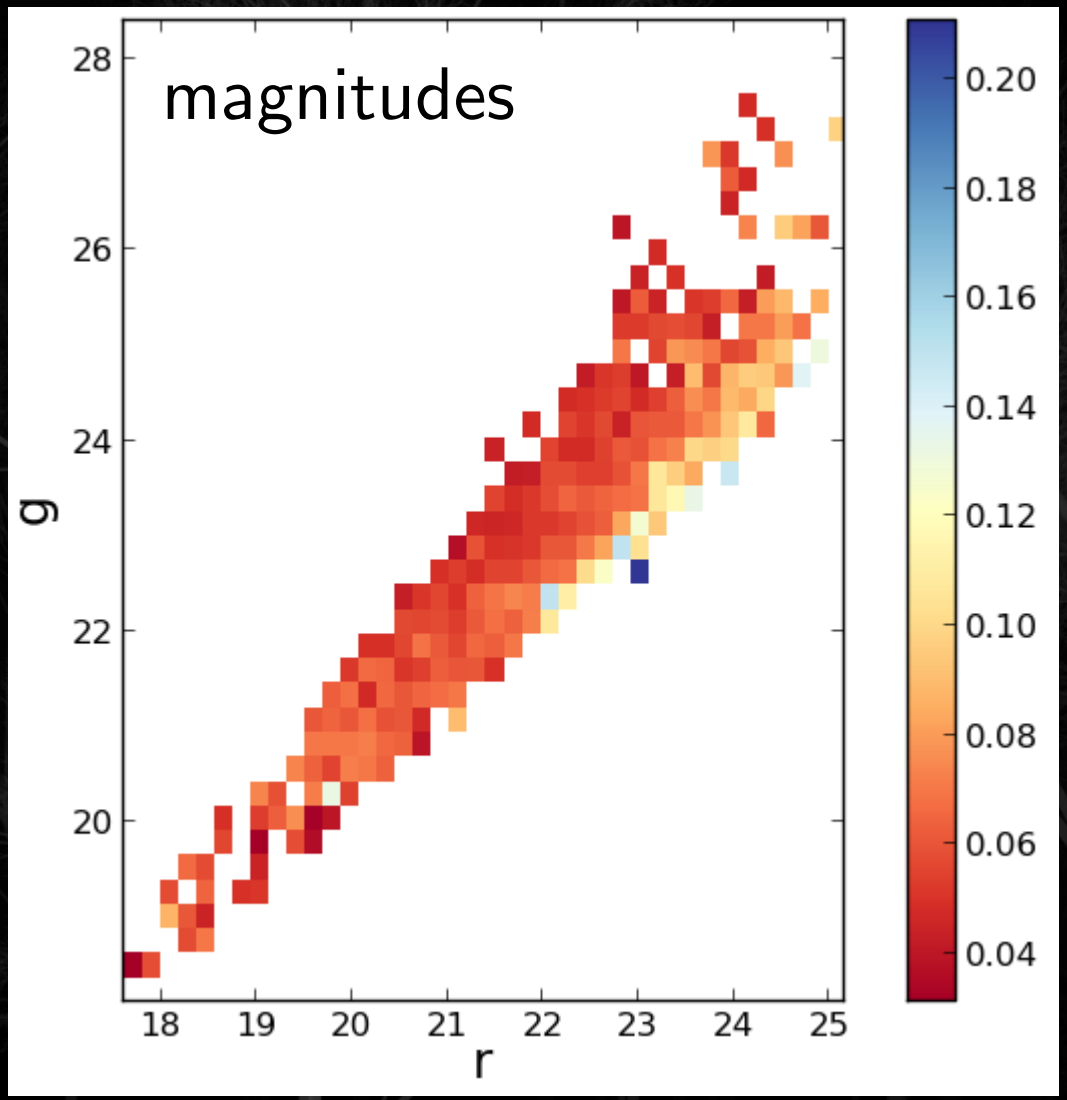


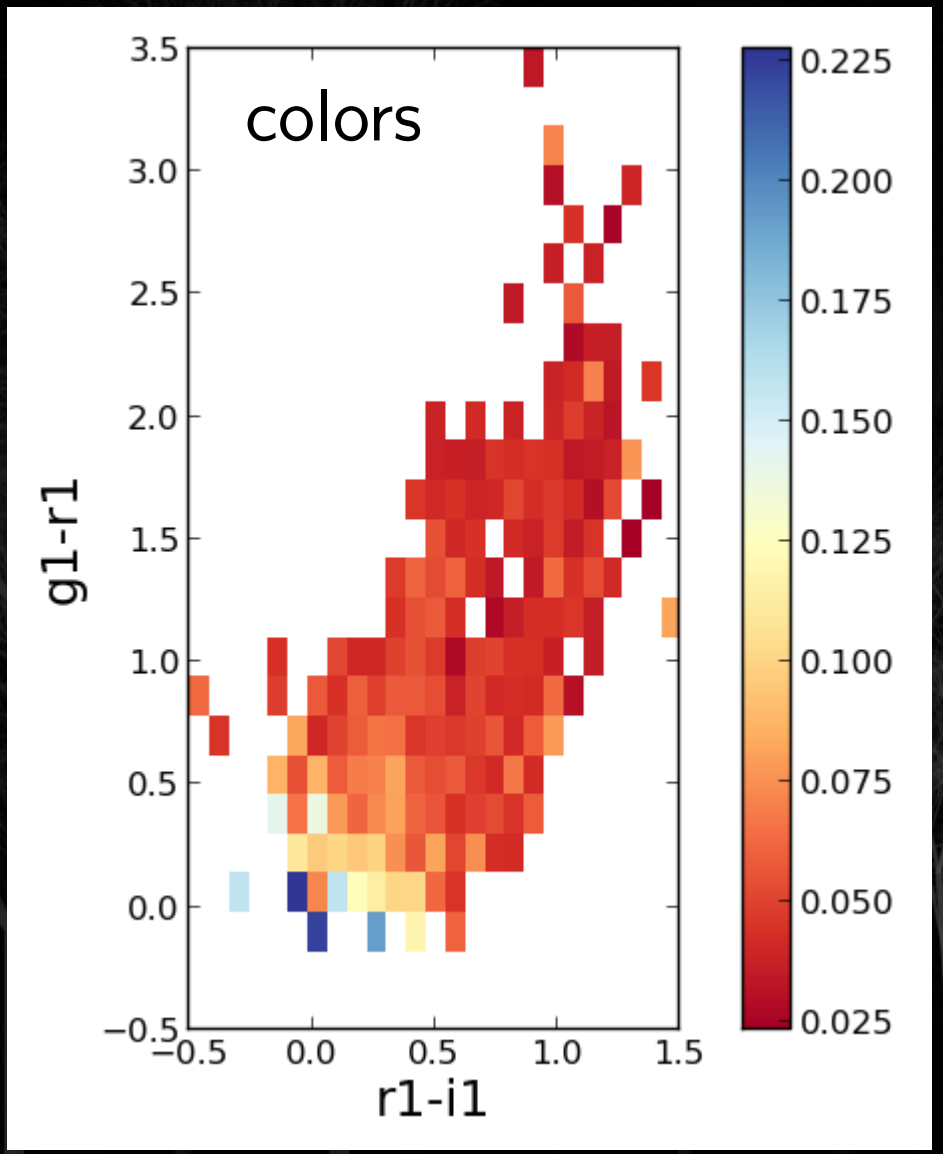

Most important attributes to construct impotance map




Useful for removing unimportant variables reducing the noise

# Application on DES SV data: Performance maps

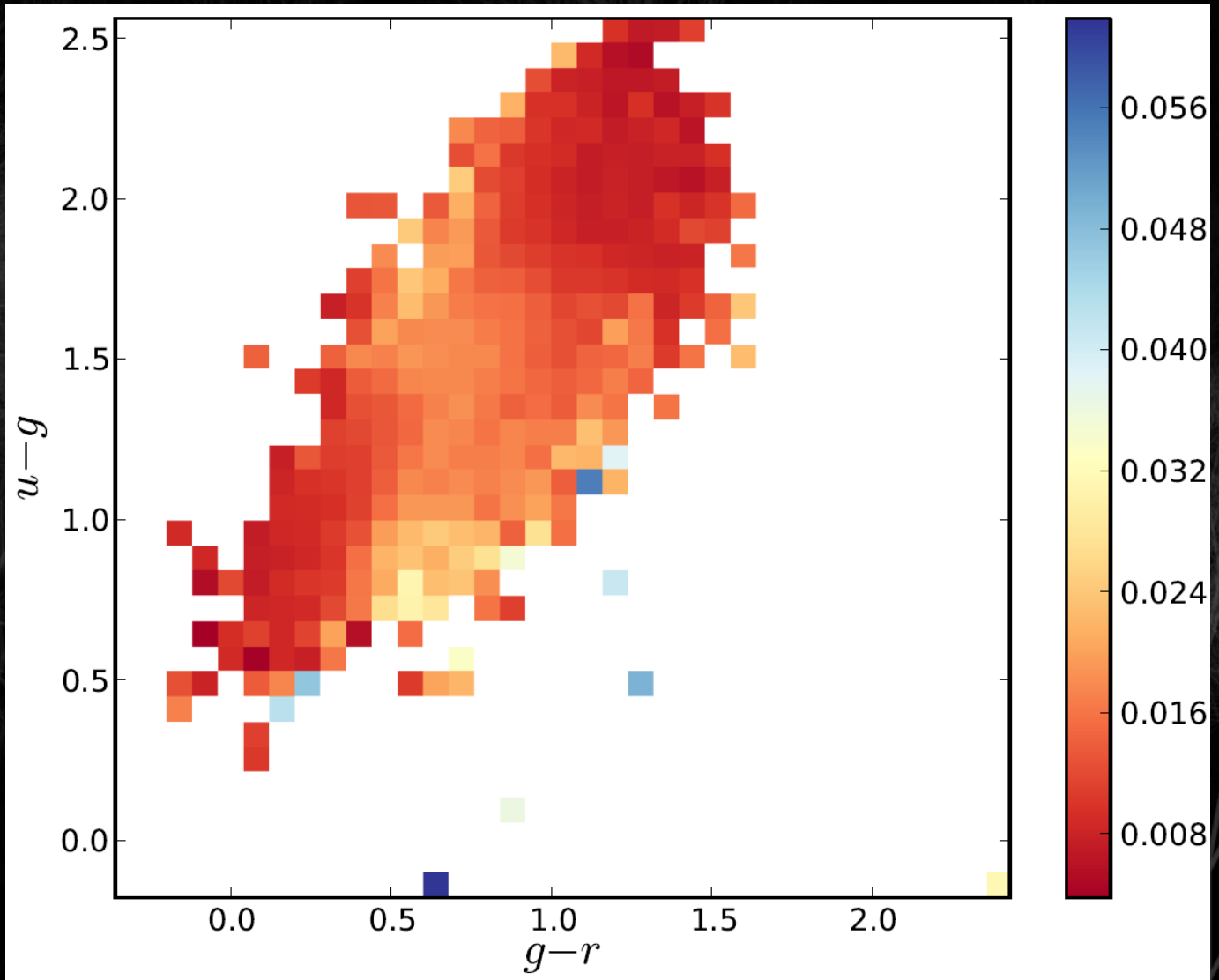



The redder the better, some areas need more attetion

# TPZ: Ancillary information - *Poor area identification* -

- Map of performance using two most important colors
- The redder the better
- Bimodality of SDSS galaxies
- Narrow follow up observations




**Carrasco Kind & Brunner 2013a**

# Conclusions

Still work in progress

These information might be useful fro proposal and targeting follow up observations

What's the cost of these proposed areas of observations?

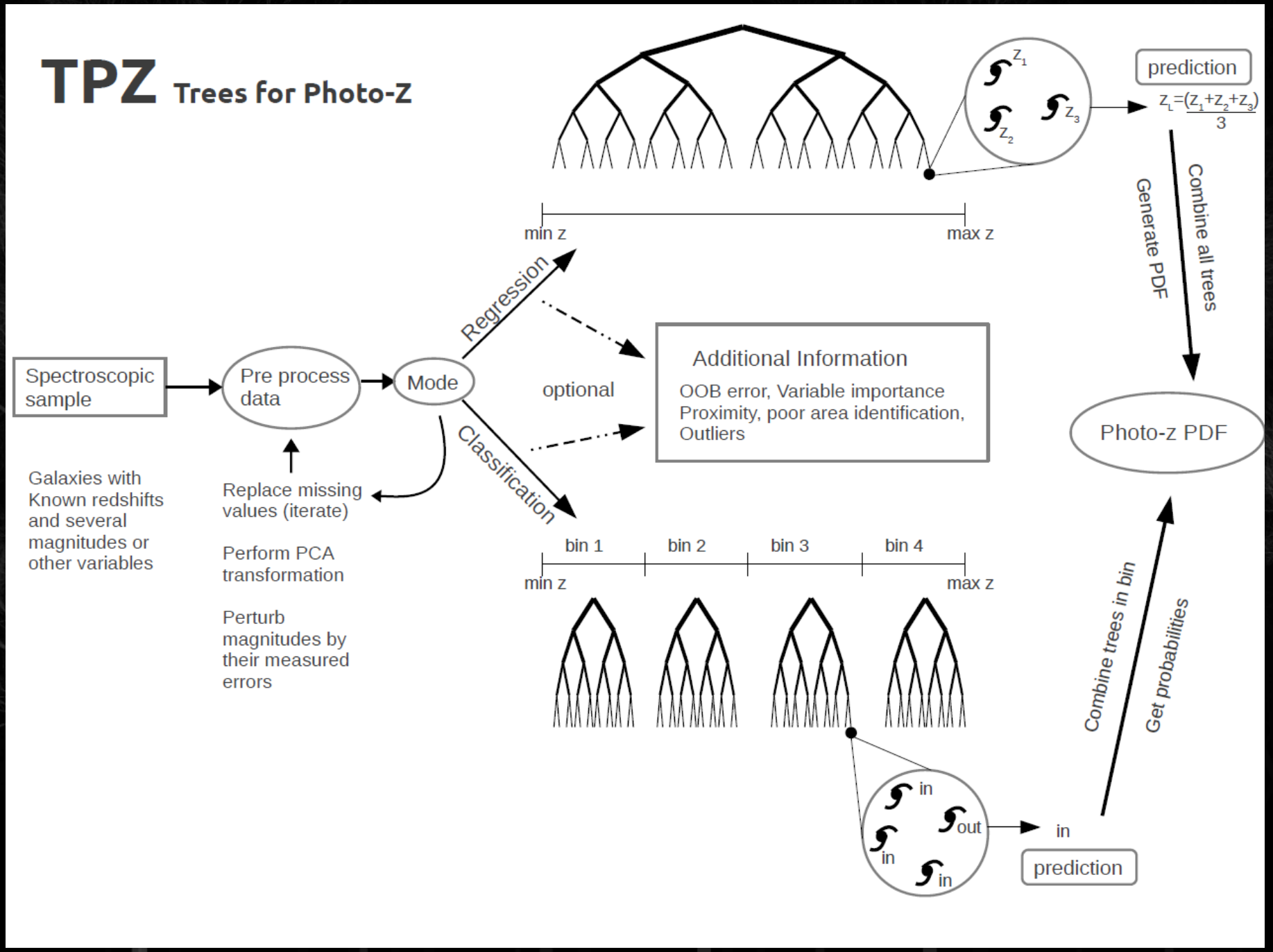
TPZ deals with missing data by "predicting those". We will try this on areas beyond the $z_{\rm spec}$ training set

# Thanks!

## Questions?

Matias Carrasco Kind
University of Illinois
mcarras2@illinois.edu

# TPZ scheme


TPZ Trees for Photo-Z
prediction
$z_L=\frac{(z_1+z_2+z_3)}{3}$
$z_1$
$z_2$
$z_3$
min z
max z
Generate PDF
Combine all trees
Spectroscopic sample
Pre process data
Mode
Regression
Classification
optional
Additional Information
OOB error, Variable importance
Proximity, poor area identification,
Outliers
Photo-z PDF
Galaxies with Known redshifts and several magnitudes or other variables
Replace missing values (iterate)
Perform PCA transformation
Perturb magnitudes by their measured errors
bin 1
bin 2
bin 3
bin 4
min z
max z
Combine trees in bin
Get probabilities
in
out
in
in
in
prediction

# Performance tests