Star – galaxy classification update

Nacho Sevilla* (CIEMAT, visiting UIUC)

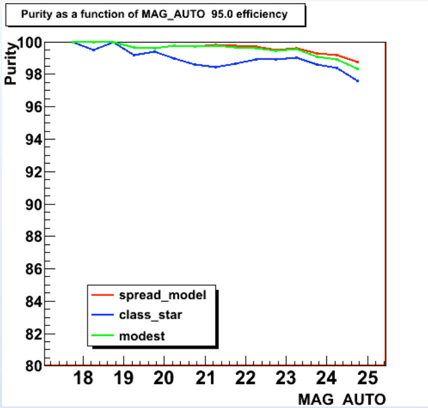
Christopher Bonnett, Robert Brunner*, Alex Drlica-Wagner*, Edward Kim*,
Eli Rykoff, Maayane Soumagnac

Stars confused as galaxies

* Chicagoland players

We are running Star-galaxy separation challenge with two goals.

1. Provide an improved classifier output for SVA1 catalog (eventually Year X) grounded on firmly based plots and tests (from Modest → Proud)



w/ plots to estimate missclassification bias and/or datasets to estimate them

2. Study the behavior of several SG classifiers (training fields, varying conditions, input vectors, Machine-Learning vs Template)

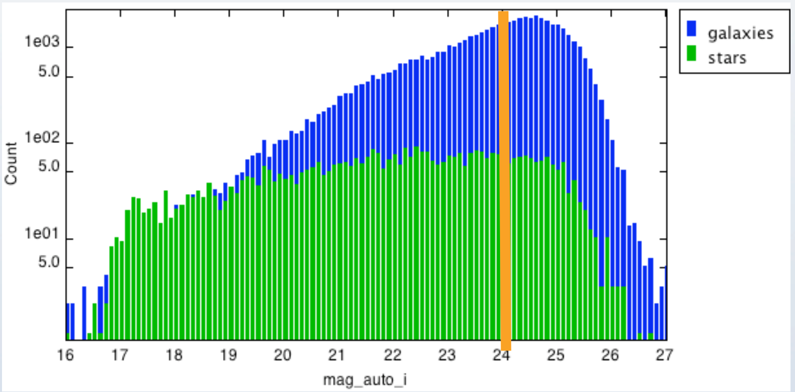
https://cdcvs.fnal.gov/redmine/projects/des-sci-verification/wiki/SG_separation_challenge
https://cdcvs.fnal.gov/redmine/projects/des-sci-verification/wiki/SG_separation_challenge_details

Round X sets are divided into three

Sets are divided into training (60%), test (20%) and blind (20%) subsamples.

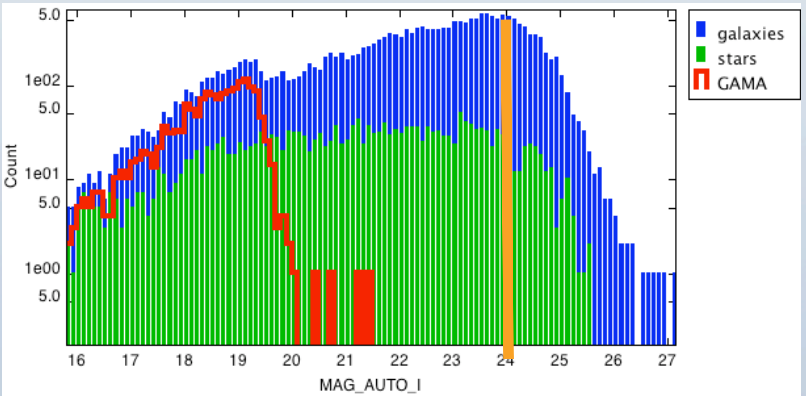
Results are submitted on the blind subsample for which no truth is available to code testers.

Round 2



7% of stars
COSMOS is 93%
dN/dmag peaks at 24.5

Round 3



9% of stars
COSMOS is 63%
dN/dmag peaks at 23.75

Besides the SExtractor outputs, several Machine Learning codes

- Modest classifier (Eli et al.). SVA1 Gold baseline.

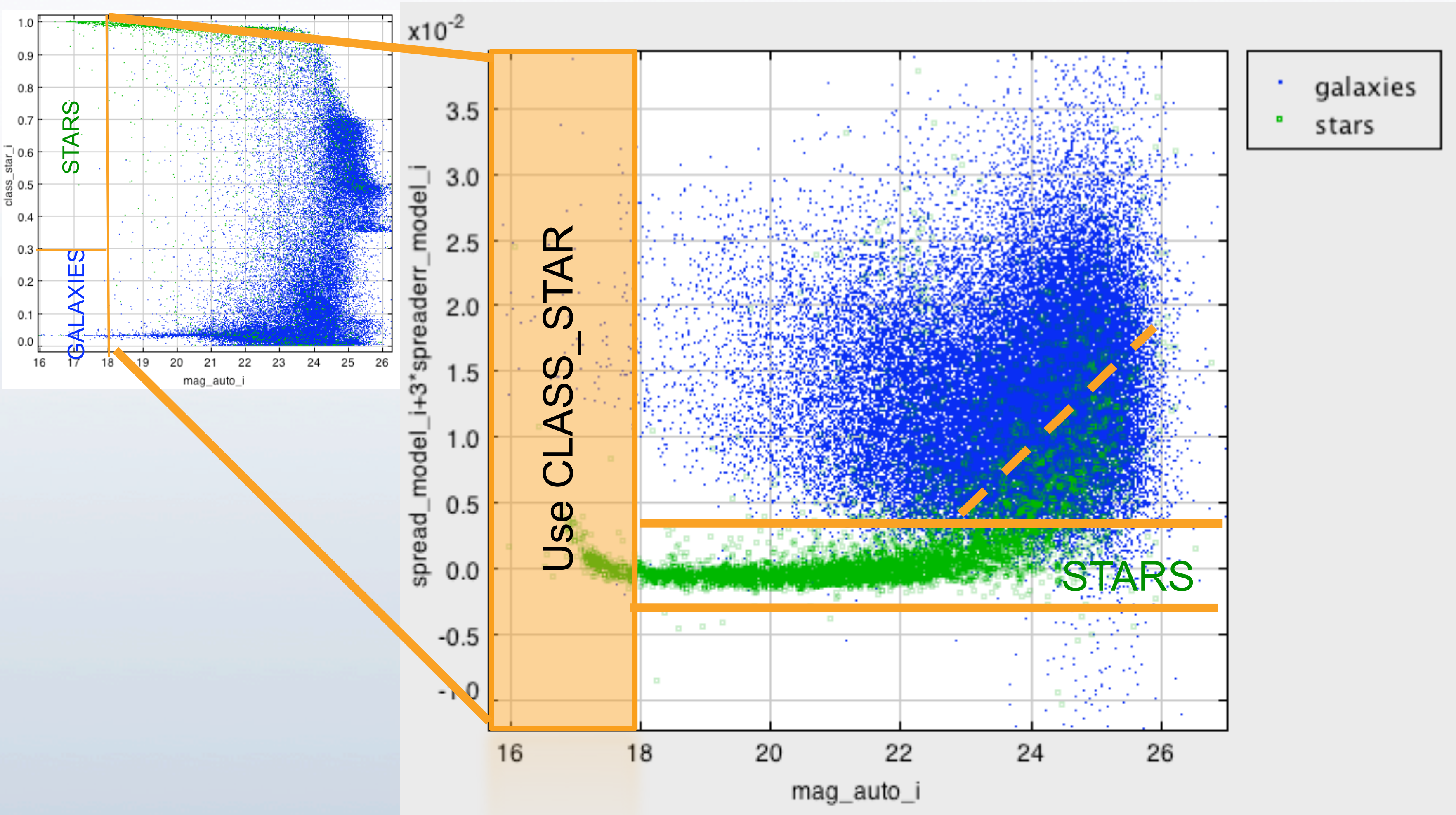
- Boosted Decision Trees (Drlica-Wagner) *(see P.Etayo-Sotos, I.S. 2012)*

- Multiclass (Soumagnac) *(see M.Soumagnac et al. 2013)*

- TPZ (Kim) *(see M.Carrasco-Kind, R.Brunner 2013)*

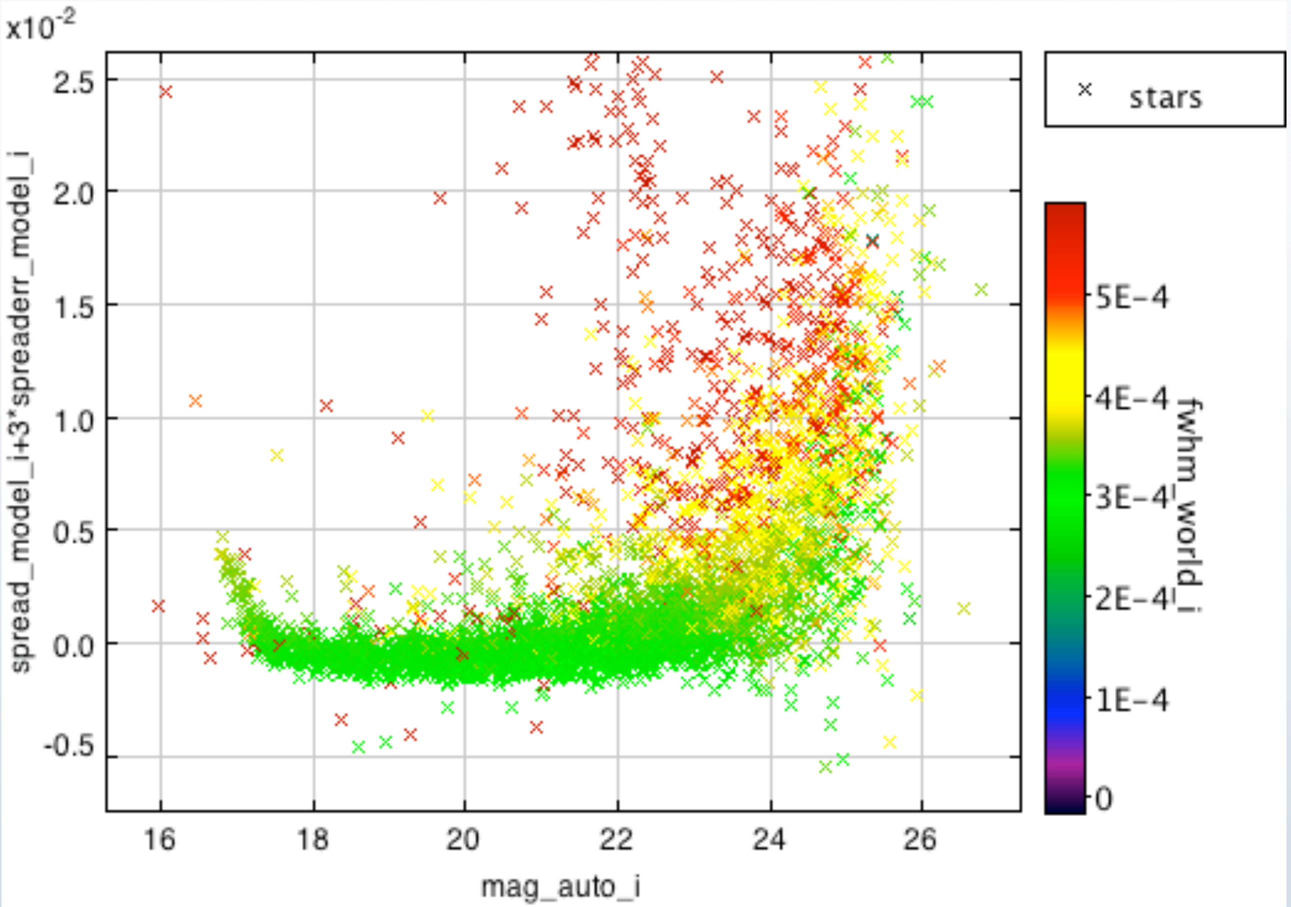
- Bonnett's Menagerie (Skynet, Support Vector Machines, other flavors of decision trees and random forests) *(see C.Bonnett 2013, Graff et al. 2013)*

MODEST classifier



Uses `CLASS_STAR` at bright end (m_auto<18) and `SPREAD_MODEL`*3*err fainter than that

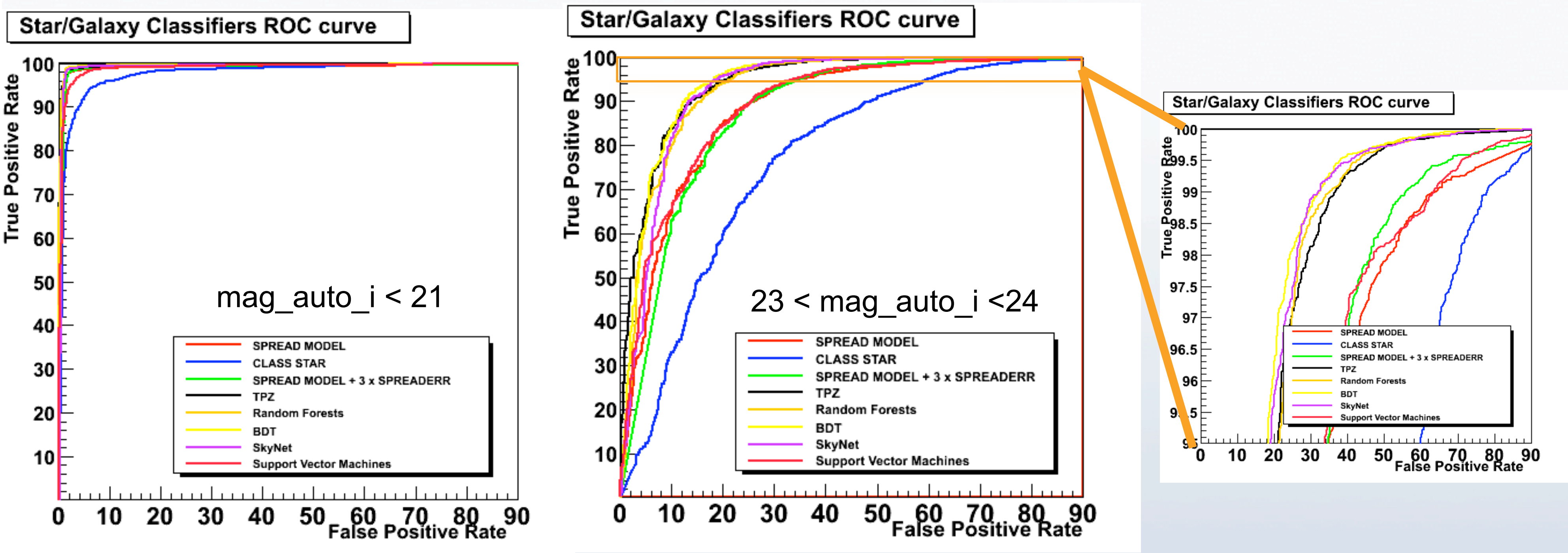
MODEST classifier



Large seeing tends to push stars to high `SPREAD_MODEL`, as expected.

DARK ENERGY SURVEY

Truth table results: ROC

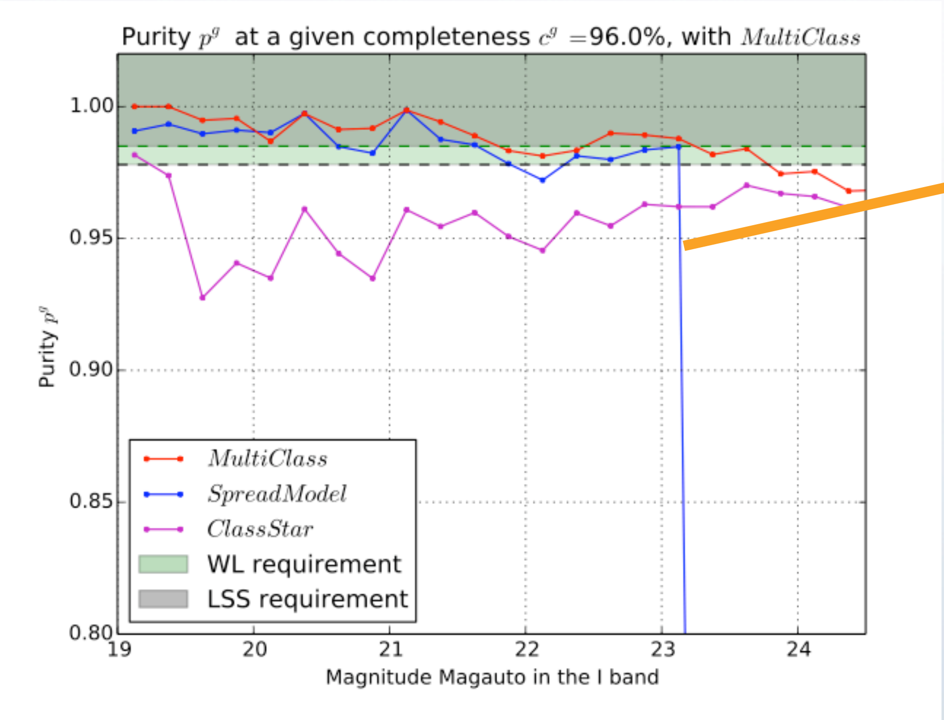


ML methods > MODEST > SPREAD_MODEL > CLASS_STAR

Differences stand out in the faint end

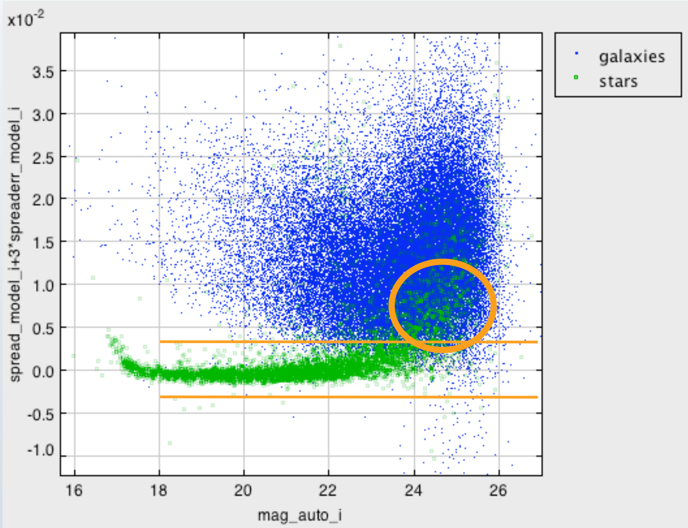
Overall performance on set very similar to training

Truth table results: Completeness & Purity



M. Soumagnac

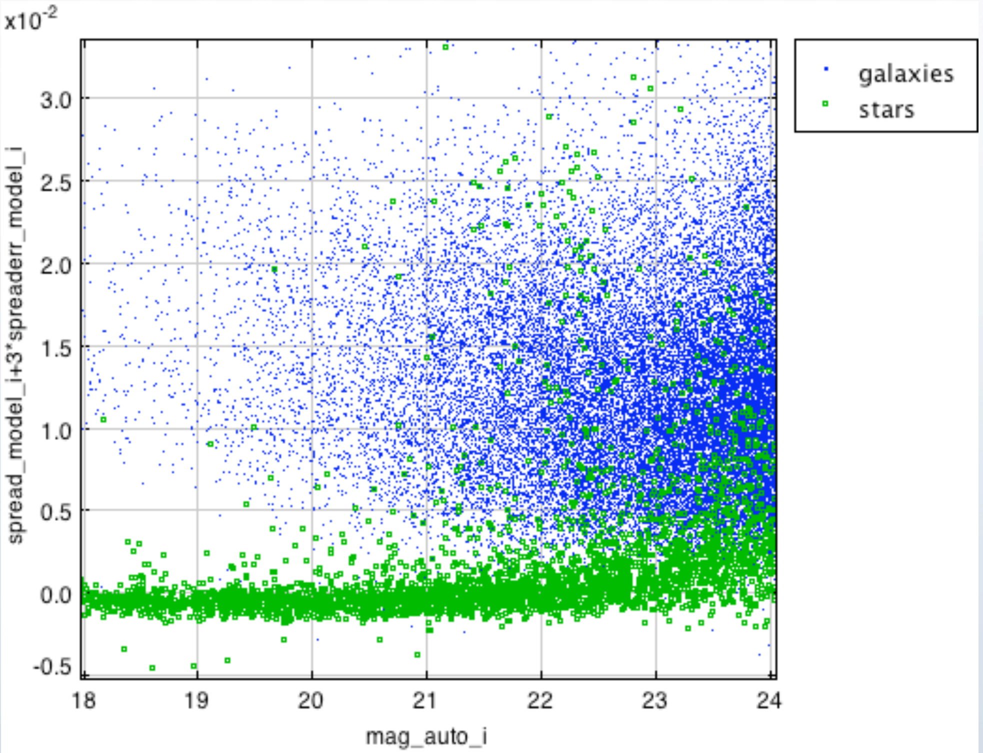
Spread_model by itself gives
good galaxy purity, but this drops drastically
at faint magnitudes (for fixed completeness)

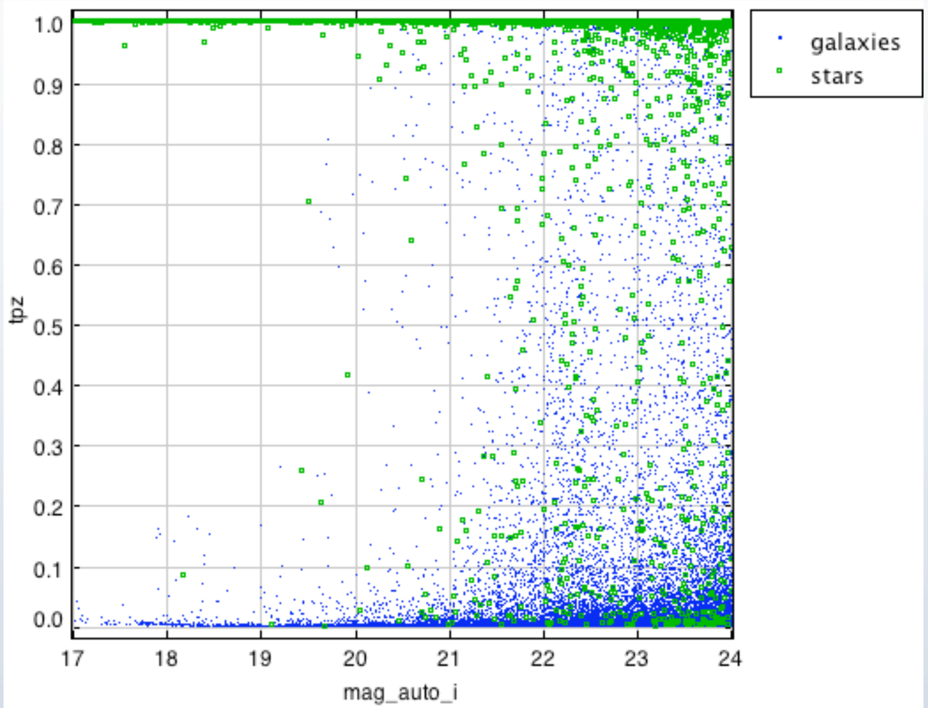


Modest behaves likewise

Multi-epoch SPREAD_MODEL (Bauer, Yanny) or T-SIZE (Sheldon) quantities will improve Modest performance.

Truth table results: Completeness & Purity

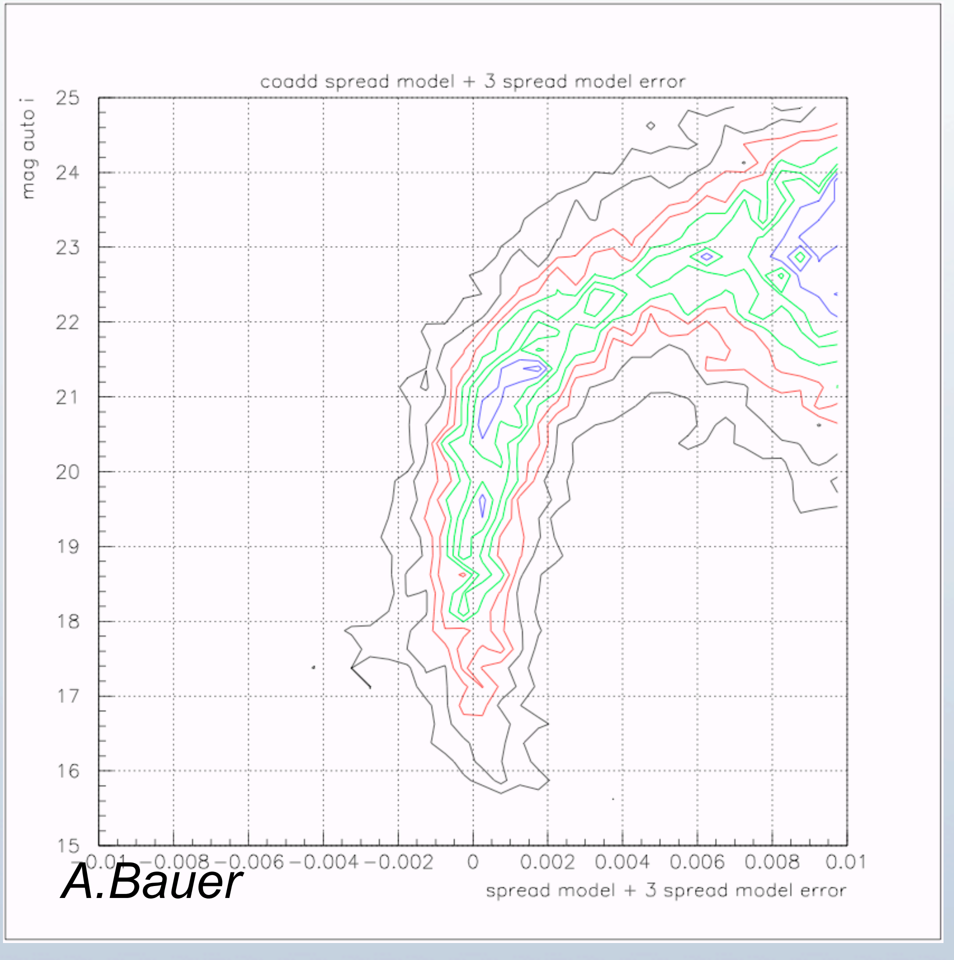


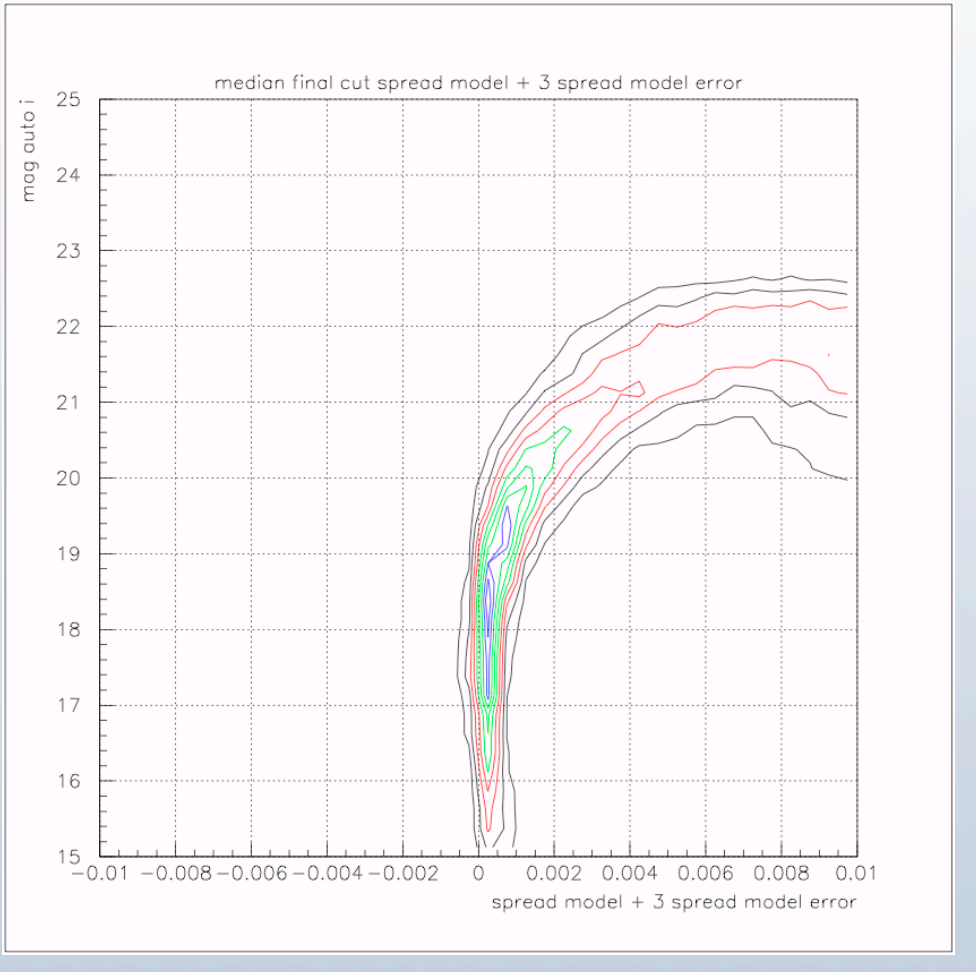


Improvements from roughly 97% to 99% purity for galaxies at faint magnitudes from modest to ML (see tables in wiki).

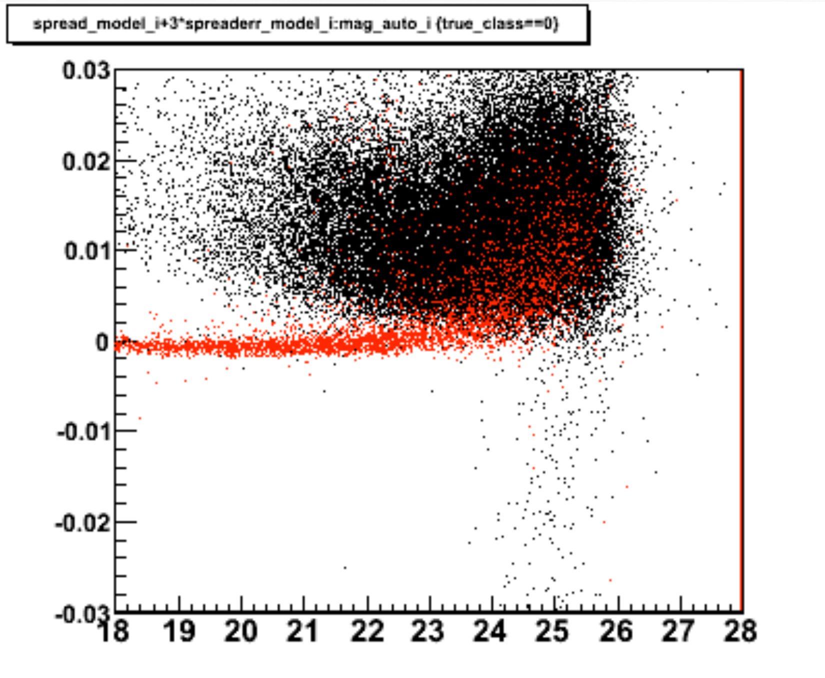
Some new approaches to improve spread_model (and modest) are being explored.

Use median/average/best `SPREAD_MODEL` from multi-epoch data instead of coadd (Bauer, Yanny)

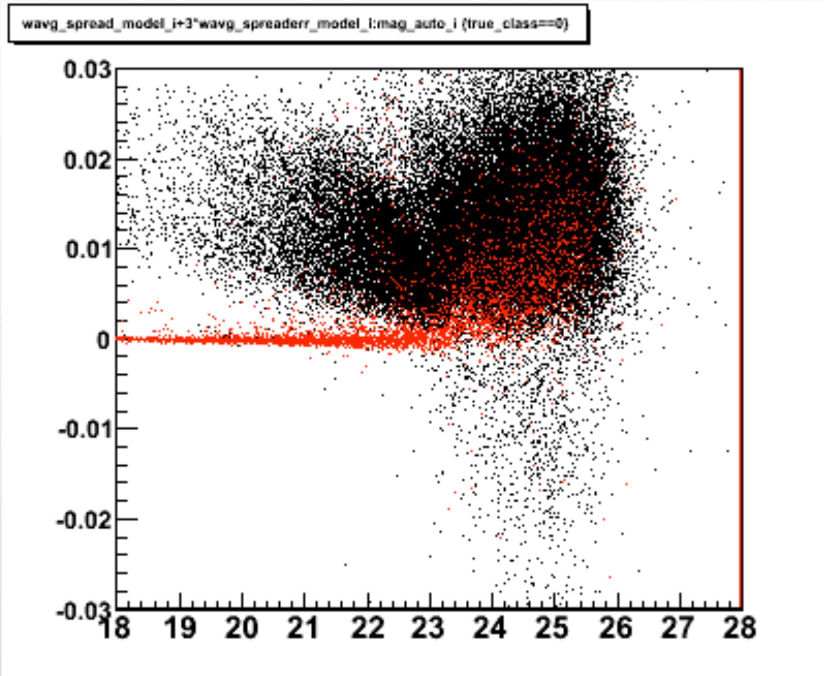




Truth table results: Completeness & Purity



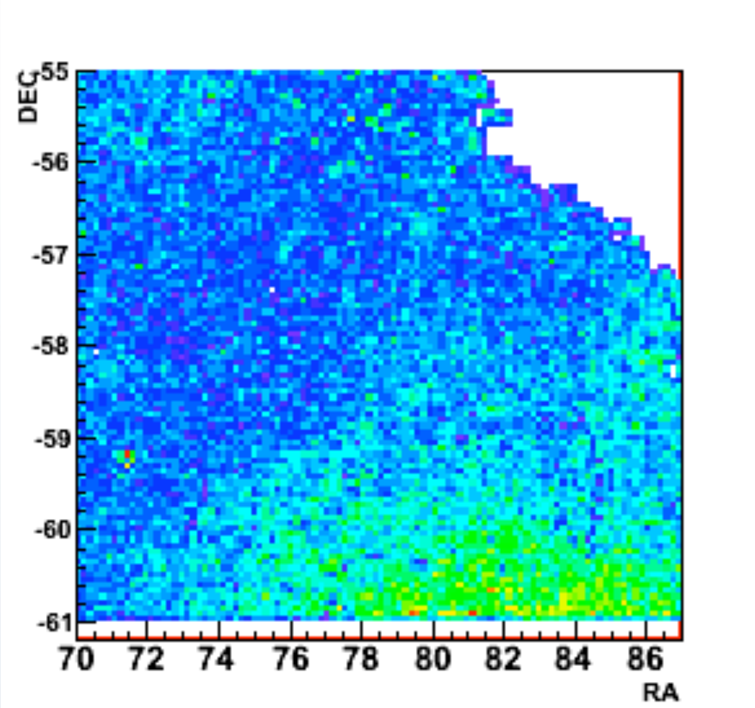
Spread Model + 3*err performance



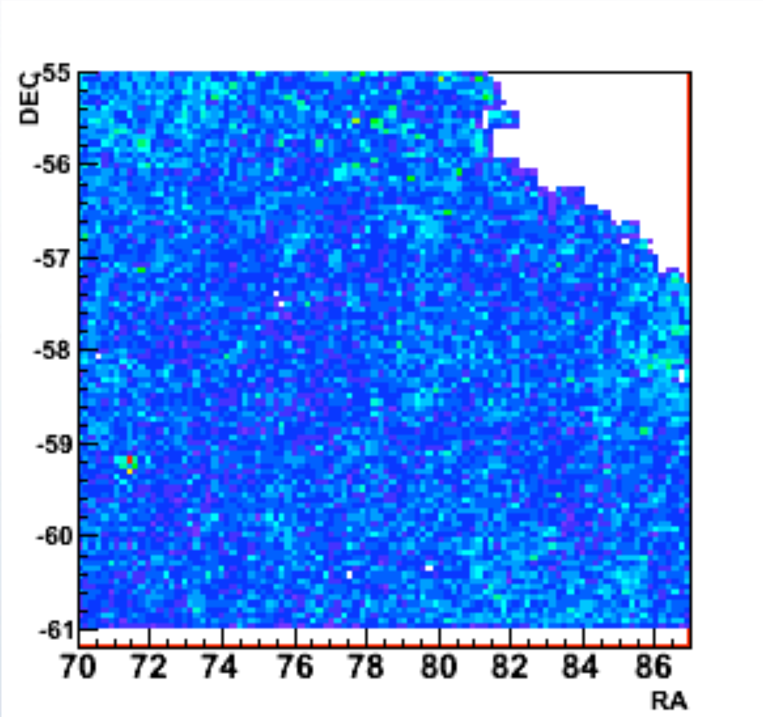
Weighted Spread Model + 3*err performance

In the calibration (COSMOS + SN) fields, the weighted average is not performing so well.

Some contamination of blue objects can be seen in LMC with Modest



Blue galaxy distribution LMC area w/MODEST



Blue galaxy distribution LMC area w/ harder cuts or new classifiers.

Blue (r-i<0.2) LMC stars get into galaxy sample.
Using TPZ_SG cut < 0.01.
Change cut inside MODEST from 0.003 → 0.008
Using weighted coadd spread_model > 0.004

Same-ish completeness slightly worse modest

Machine learning methods very effective but need review of calibration

Galaxy sample completeness, fixing at same purity

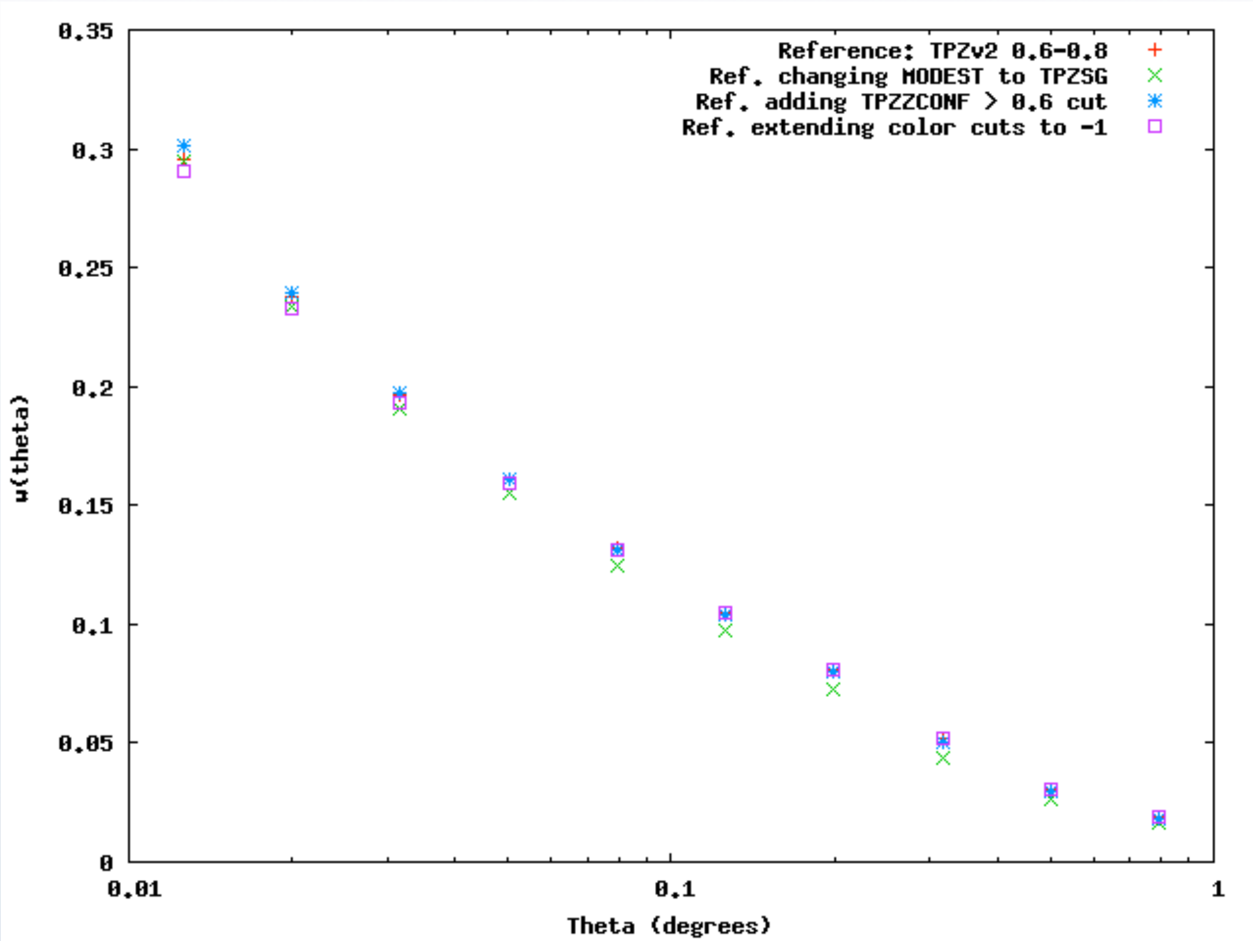
Classifier	COSMOS+SN	SPTE (LMC)	Y1 - Stripe 82
Spread + 3*err	98.6%	Slightly worse	91.7%
weighted avg Spread	95.0%	Good	94.9%
ML	>99.5%	Good	92.3%

(ML methods trained on different deep COSMOS+SN data, no S82 training)

DO NOT compare numbers across columns: different star/galaxy ratios.

Probably revisit the COSMOS calibration when new reductions are done.
Or just use stripe 82.

We have started to use correlation functions as diagnostics too



Small impact on current SVA1 studies.

Will be important for larger scales and precise determinations.

Lot of good work here from SG team, Bauer, Cawthon, J. Sanchez, Sobreira

Also star-galaxy cross-correlations.

Summary

Several star-galaxy classifiers are being tested with SVA1 and Y1A1 stripe 82 data.

Calibration fields: Machine Learning codes perform better than Modest. Slightly better than tuned and weighted average spread_model.

Stripe 82: Weighted average more robust currently. Room for improvement with ML codes.

Current Modest classifier could be made a little tighter for galaxy studies.

Keep updated at des-sci-release@fnal.gov list and telecons.

Infrastructure Bonus Points Screen

Bayesian methods under-represented in this work. Combination methods in the works.

Probabilistic output.

Study dependency with training features, fields.

Other fields:

- QSO classification

- Artifact identification

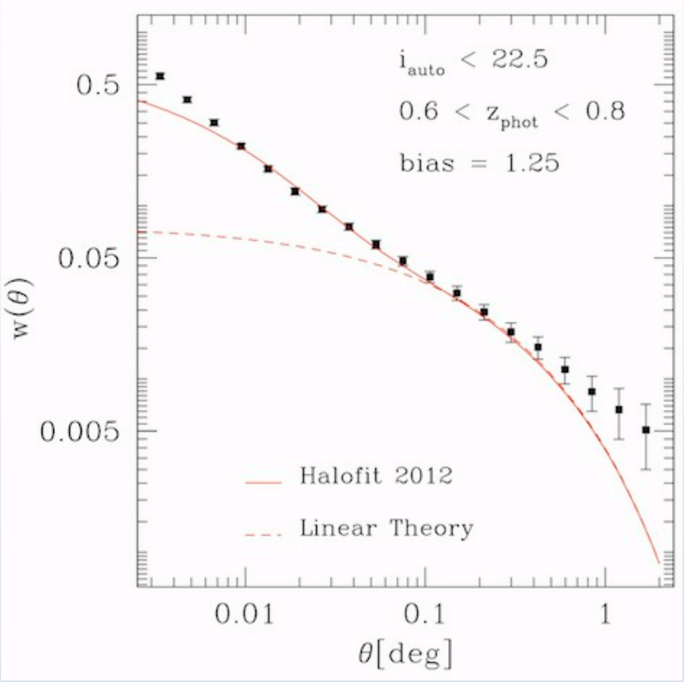
Some tables to play around:

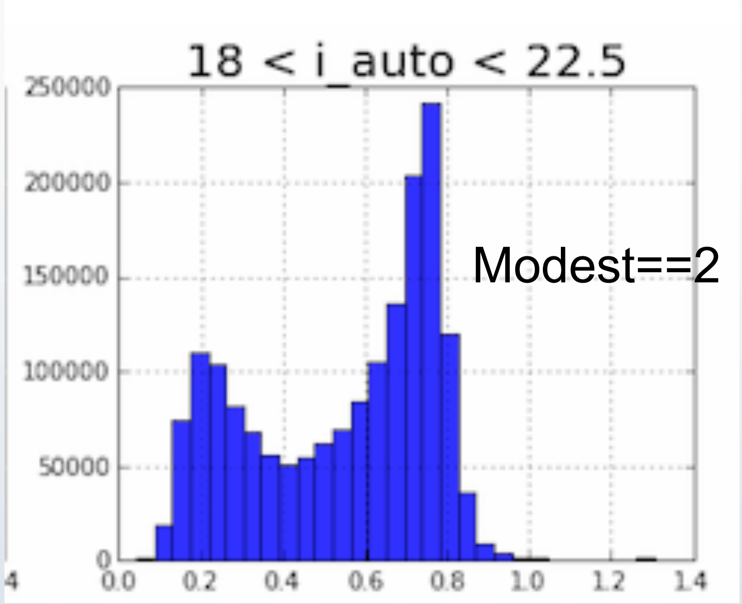
MCARRAS2.TPZ_SVA1_GOLD (courtesy M.Carrasco-Kind, latest TPZ, experimental TPZ-SG)
NSEVILLA.S82_TPZ_SG (courtesy E.Kim, TPZ-SG on stripe 82, just for testing!)
NSEVILLA.Y1_STRIPE82_MATCHES (matches to SDSS dered mags and true spectroscopic class)

- **https://cdcvs.fnal.gov/redmine/projects/des-y1/wiki/Y1A1_Stripe_82**

Backup slides start here

An example of why:





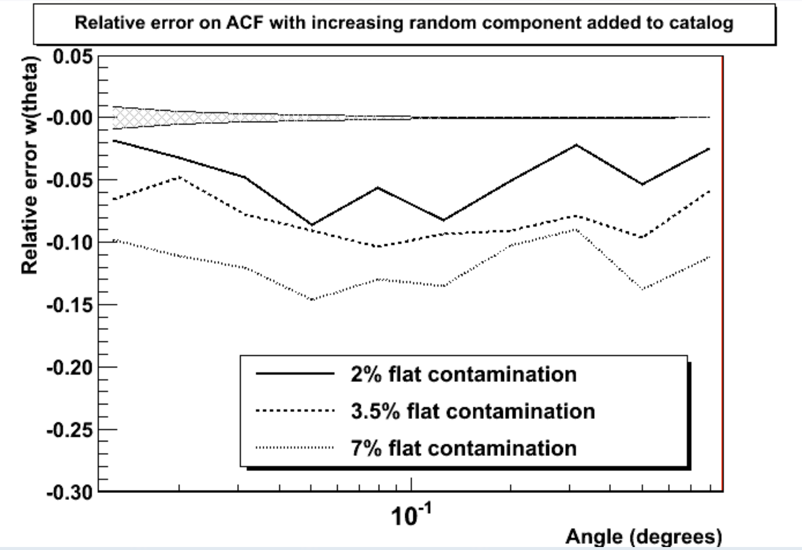
M.Crocce, C.Bonnett et al.

Excess in 'expected' clustering could be related to higher rate of stars at certain redshift bins.

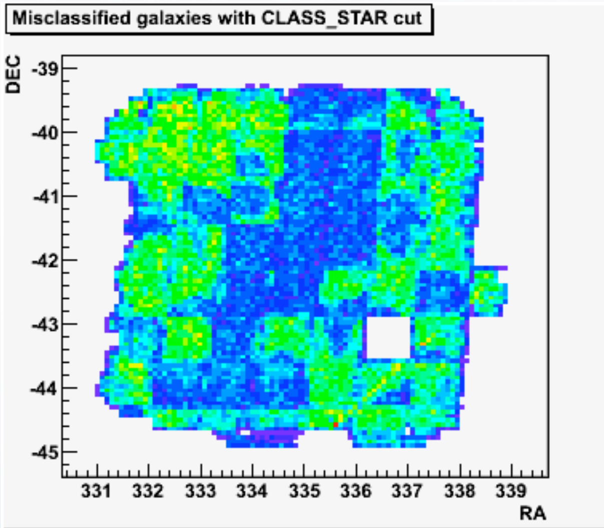
Impact on determination of clustering amplitude (1-f)**2 (f=impurity of galaxy sample)

Milky Way science only wants stars! Cluster composition purity.

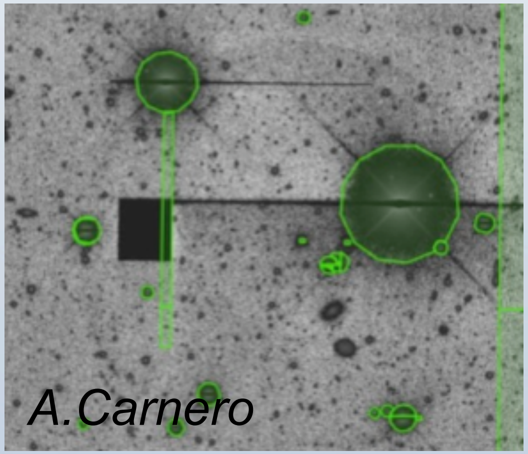
There are several problems associated with SG confusion and bright stars



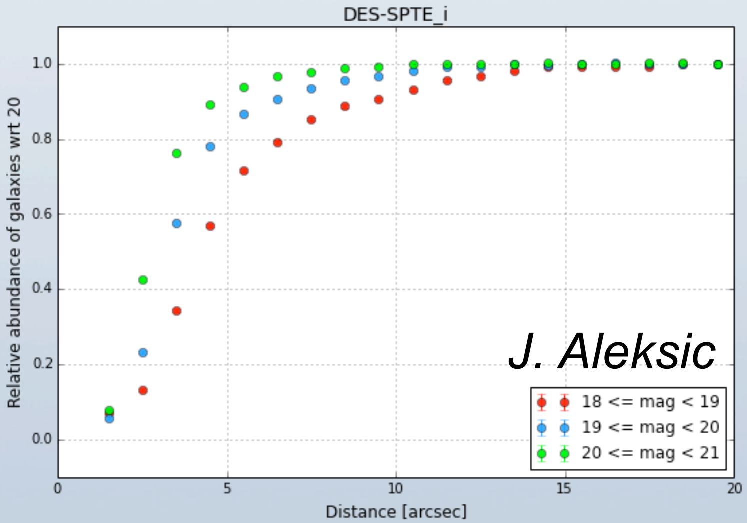
Overall contamination of sample



Spatially dependent contamination



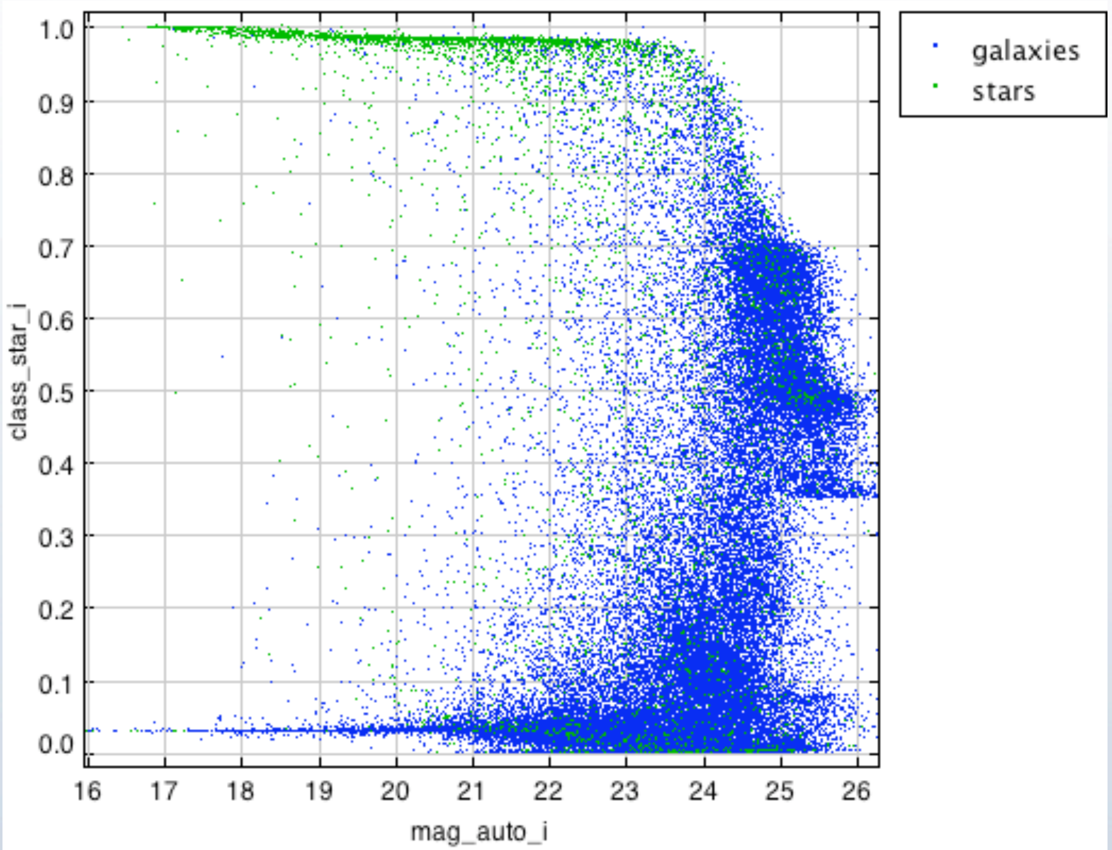
Saturated stars



Stellar obscuration

CLASS_STAR classifier

- SExtractor's CLASS_STAR is the output of multi-layered Perceptron (1 hidden layer) trained on simulated ground -based images
- Derived from a classifier that would originally operate on photographic scans (Bertin 1994).
  - Isophotal areas
- One of the inputs acts as a "tuning button" set to the current PSF FWHM ("seeing")
  - Isophotal areas are expressed in units of the (local) PSF $FWHM^2$



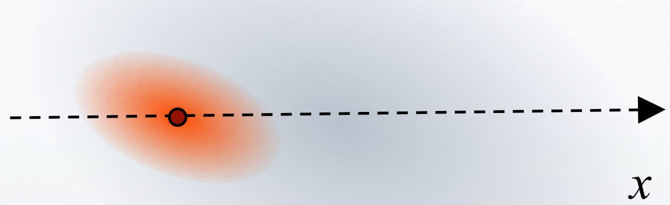
E.Bertin

SPREAD_MODEL classifier

- The new **SPREAD_MODEL** compares the object to both the local PSF and a barely resolved, PSF-convolved exponential model (linear discriminant analysis):

PSF PSF x exp

$$x = \Sigma^{-1}(\Phi - \overline{G}).I$$ pixel map



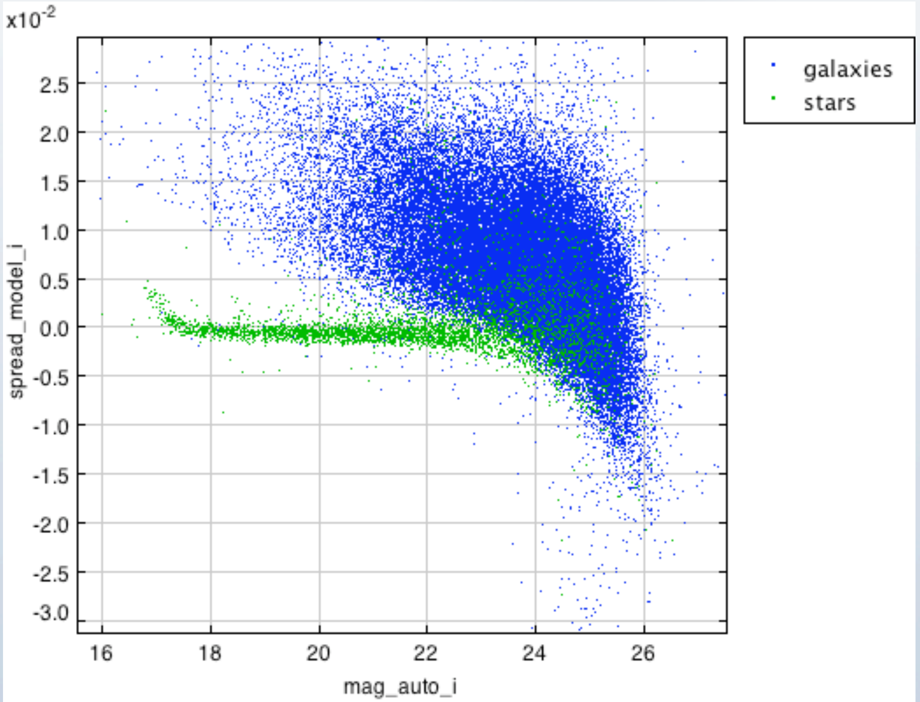
- We normalize with respect to the local PSF and galaxy model:

$$\textbf{SPREAD\_MODEL} = \frac{\sum_i \phi_i I_i}{\sum_i \phi_i^2} - \frac{\sum_i \overline{G}_i I_i}{\sum_i \overline{G}_i \phi_i}$$

- ***G*** is the convolution of the local PSF with a circular exponential profile with r_h= FWHM/16

- **SPREADERR_MODEL** can be used to define the decision boundary with respect to the stellar locus

- The stellar locus itself can be used to monitor things such as the accuracy/stability of the PSF model and linearity of the data



E.Bertin

MODEST classifier

Classification proposal put together by *E.Rykoff* after a long exchange and tests and using past experience (with E.Bertin, D.Capozzi, B.Santiago, N.S., W.Wester)

Galaxies:

(FLAGS_I <=3)
AND NOT
(((CLASS_STAR_I > 0.3) AND (MAG_AUTO_I < 18.0))
OR ((SPREAD_MODEL_I + 3*SPREADERR_MODEL_I) < 0.003)
OR ((MAG_PSF_I > 30.0) AND (MAG_AUTO_I < 21.0))
)

Stars:

(FLAGS_I <=3)
AND
(((CLASS_STAR_I > 0.3) AND (MAG_AUTO_I < 18.0))
AND (MAG_PSF_I < 30.0)
OR (((SPREAD_MODEL_I + 3*SPREADERR_MODEL_I) < 0.003)
AND ((SPREAD_MODEL_I +3*SPREADERR_MODEL_I) > -0.003)))

Uses CLASS_STAR at bright end (m_auto<18) and SPREAD_MODEL*3*err fainter than that

Two rounds became three

Round 1: COSMOS deep imaging with HST data for training and testing.

Goal 0: to get things running (Jan 2014)
Participants: BDTs, TPZ, Bayesian, Skynet, class_star, spread_model, modest_class

Round 2: COSMOS deep imaging and SN fields (Gold) with HST data and ground-based spectra.

Goal 1: get a catalog for SVA1 (March-May 2014)
Participants: BDTs, TPZ, Random Forests, Skynet, SVM, class_star, spread_model, modest_class

Round 3: COSMOS shallower imaging (Rykoff) and SN fields (Gold) with HST data and more ground-based spectra (Kim).

Goal 1 and 2: get a catalog for SVA1 (October-November 2014)
Participants: BDTs, TPZ, Random Forests, class_star, spread_model, modest_class

We use mainly the COSMOS photometric field and various spectroscopic datasets

RA, DEC
MAG_MODEL_GRIZ and MAGERR errors
MAG_DETMODEL_GRIZ and MAGERR errors
MAG_AUTO_GRIZ and MAGERR errors
MAG_PSF_GRIZ and MAGERR errors
SPREAD_MODEL_GRIZ and SPREADERR errors
CLASS_STAR_GRIZ (not in Gold, taken from SVA1_COADD_OBJECTS)
FWHM_WORLD_GRIZ (not in Gold, taken from SVA1_COADD_OBJECTS)
MODEST_CLASS

To be added:
A_IMAGE
B_IMAGE
KRON_RADIUS

Catalog	Field	Area	Type	Nb. of good quality stars	Magnitude range	Associated paper	Comments
ACS COSMOS	COSMOS	~1 deg^2	Space imaging	~15000	mag_auto_i<25	Leauthaud07	
VVDS-DEEP-02	SN-X	0.6 deg^2	Spectroscopy	~600	17.5<mag_auto_i<24.75	LeFèvre13	Includes ultra-deep
VVDS-CDFS	SN-C		Spectroscopy	~100	17.5<mag_auto_i<24		
ACES	SN-C		Spectroscopy	~300	18<mag_auto_i<23	Cooper12	
ACS GOODS-S	SN-C	~160 arcmin^2	Space imaging	~300	mag_auto_z < 27	Giavalisco04	Conservative cut in mag_auto<24

Catalog	Field	Area	Type	Nb. of good quality stars	Magnitude range	Associated paper	Comments
SDSS-DR10	SN-X	30 deg^2	Spectroscopy	~1000	14-22.5	Ahn13	
VIPERS	SN-X	TBD	Spectroscopy	~250	17-23	Guzzo13	
GAMA	SN-X	10 deg^2	Spectroscopy	~140	14-20	Liske14	
DES-AAOmega	SN-X, SN-C, SN-E	TBD	Spectroscopy	~620	15-22	TBD	
UDS	SN-X	TBD	Spectroscopy	~20	18-26	TBD	

We compare classifiers with a blind sample, using truth values, and general distributions , over whole SVA1

Against truth table

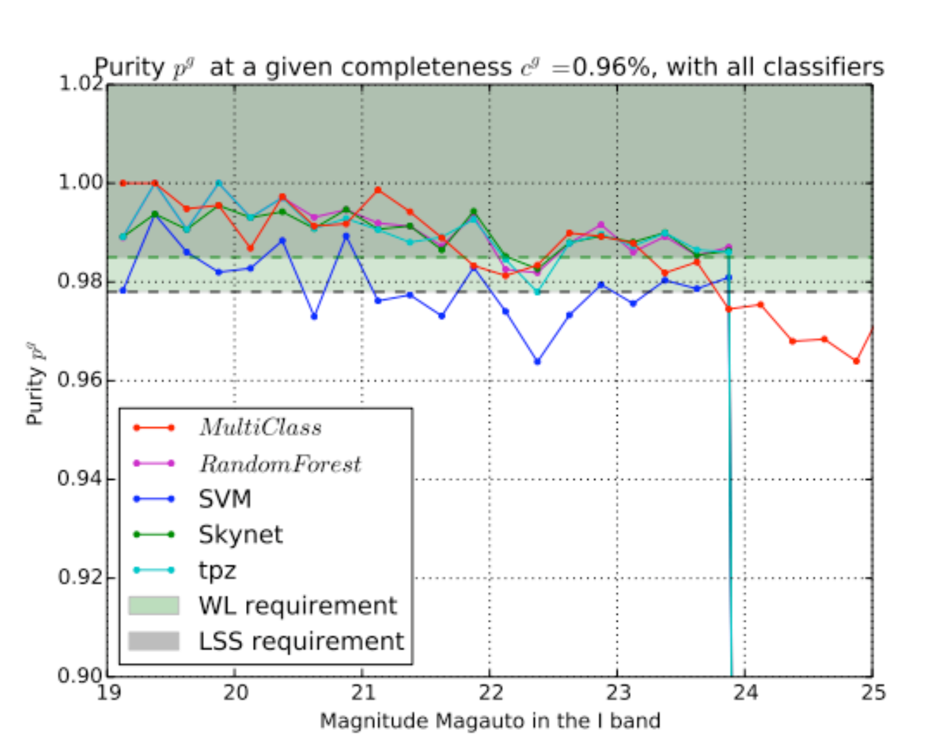
• True positive rate (aka completeness): true galaxies correctly identified as galaxies over all true galaxies.
• False positive rate: true stars incorrectly identified as galaxies over all true galaxies.
• Positive predictive value (aka purity): true galaxies in overall galaxy identifications.

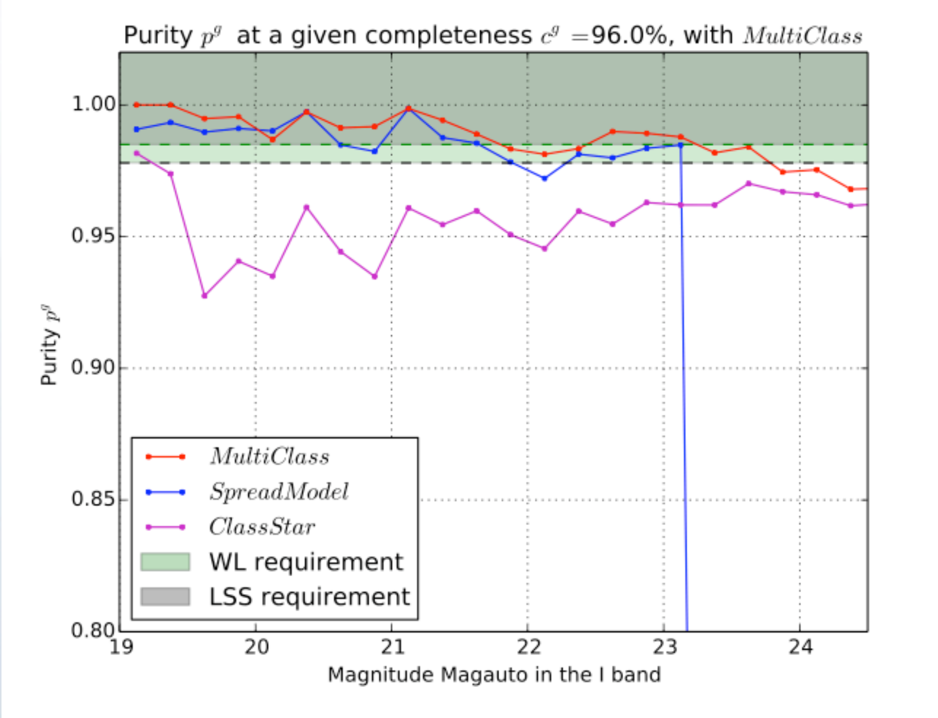
ROC curve vs magnitude: best performance the larger the area under the curve.
Completeness and purity vs magnitude: compare vs requirements.
Purity vs photo-z: relevant to LSS benchmark testing.

General distributions

Classifier outputs
N(m) fors stars and galaxies
N(photo-z) for stars and galaxies
Stellar loci
Star-galaxy ratios, star and galaxy densities
Correlations

Truth table results: Completeness & Purity

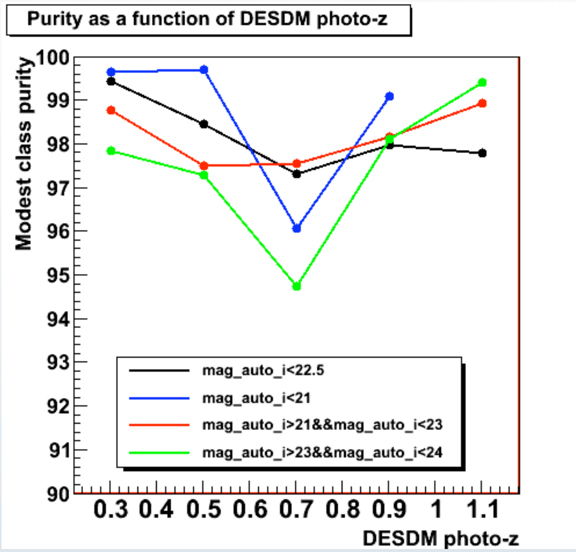


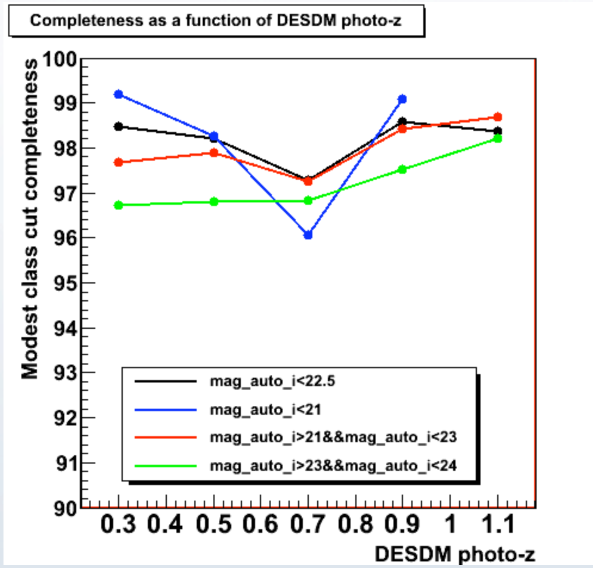


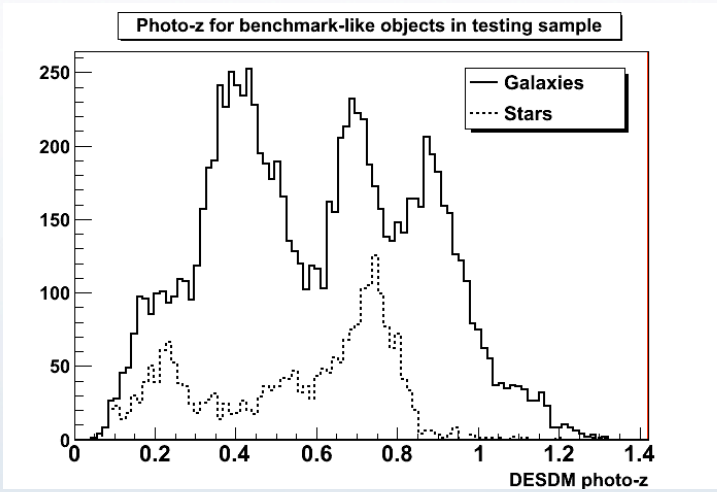
M. Soumagnac

All machine learning methods perform equally well in testing sample.

Truth table results: photo-z







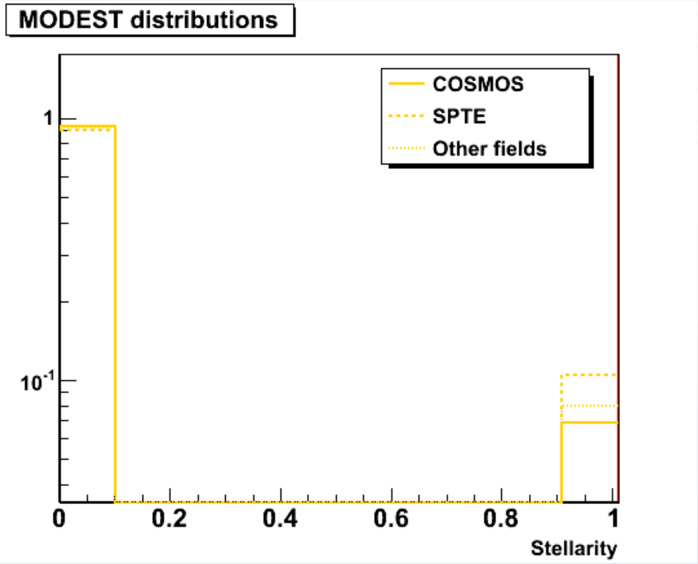
(black line in plots on left)

As expected, purity depends also on true star/galaxy ratio (varies with redshift/magnitude)

To derive purity for application, need ‘similar’ star/galaxy ratio in testing area.

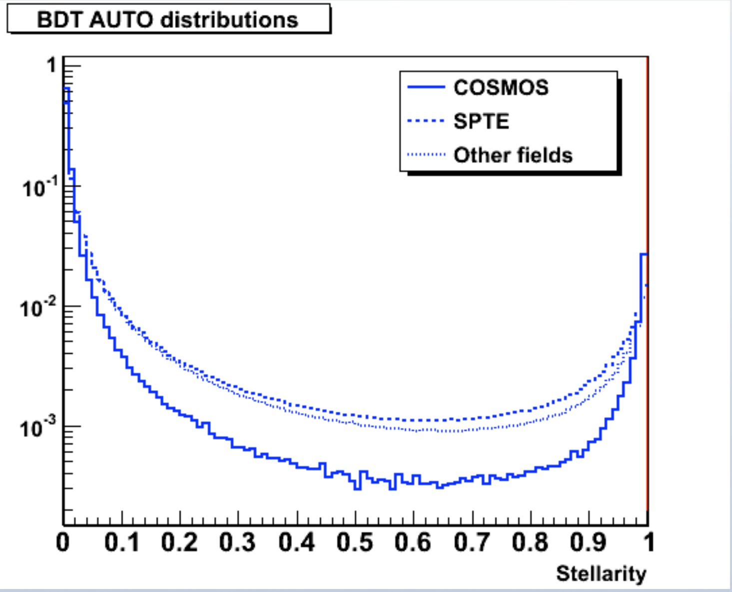
Currently only COSMOS, VVDS. In Y1 add COMBO-17@CDFS, Stripe 82 (but shallow)
Also can use simulations, reweighting, upper limits.

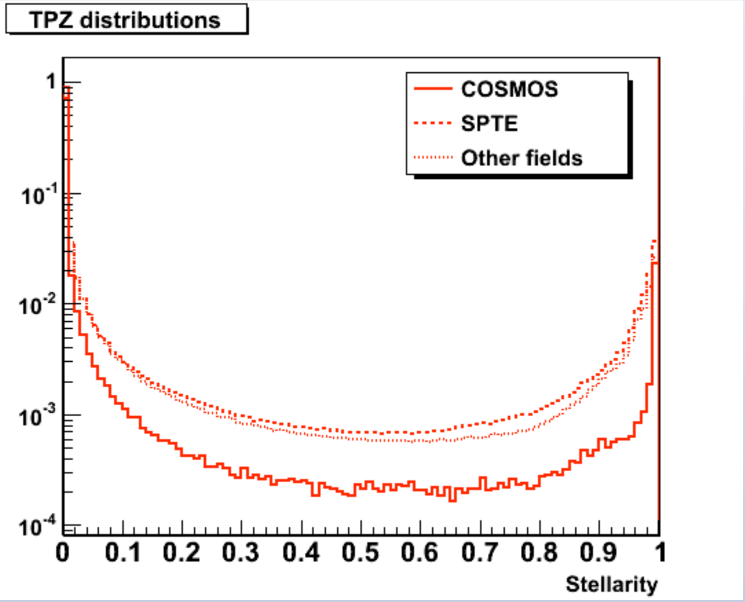
General distributions: classifier outputs



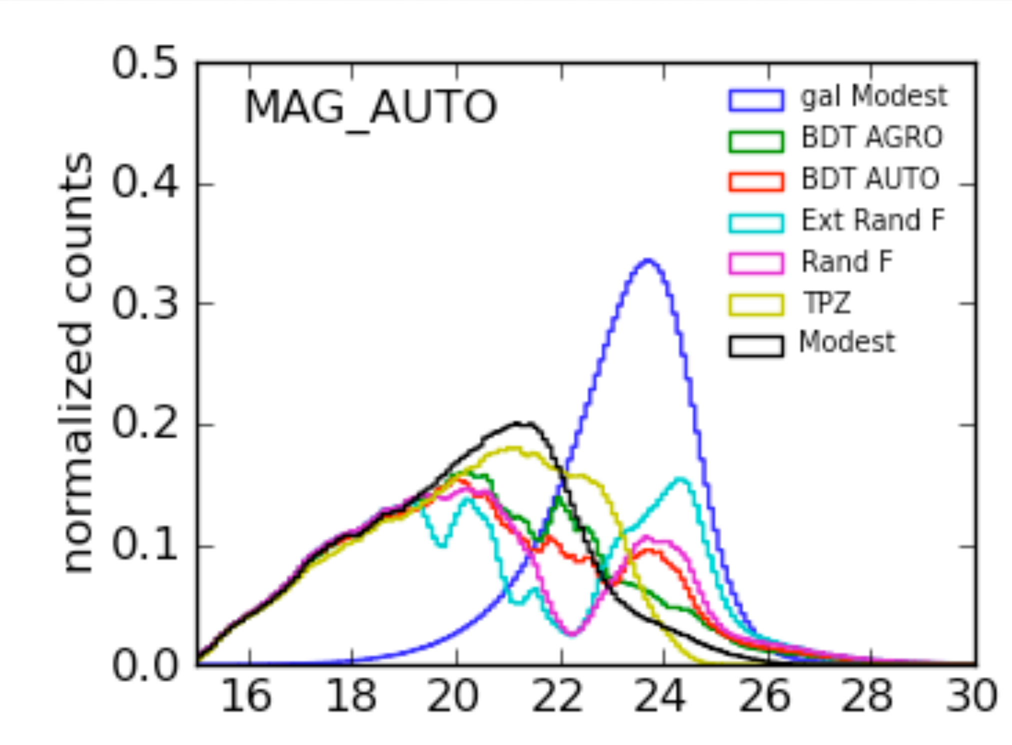
Ratio of stars to galaxies similar.

Higher uncertainty when generalizing to the whole SVA1 area.





General distributions: N(m)



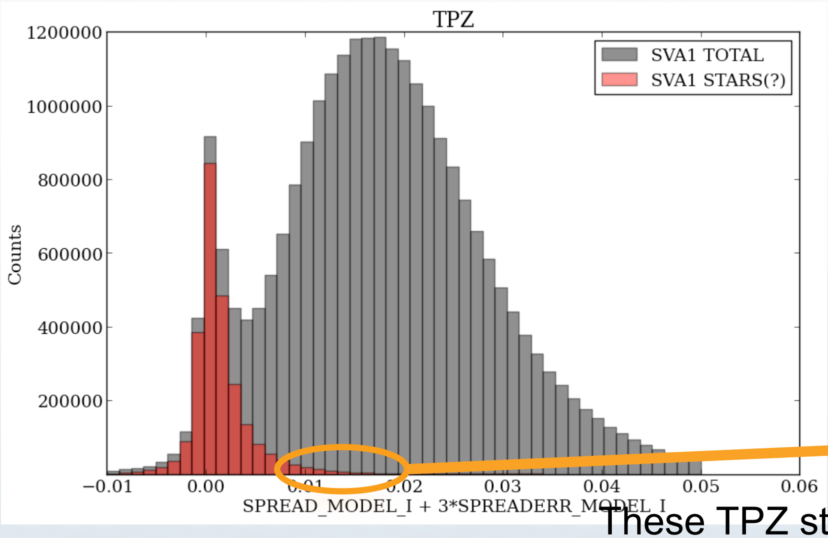
Look at N(m) for the whole SVA1 area.

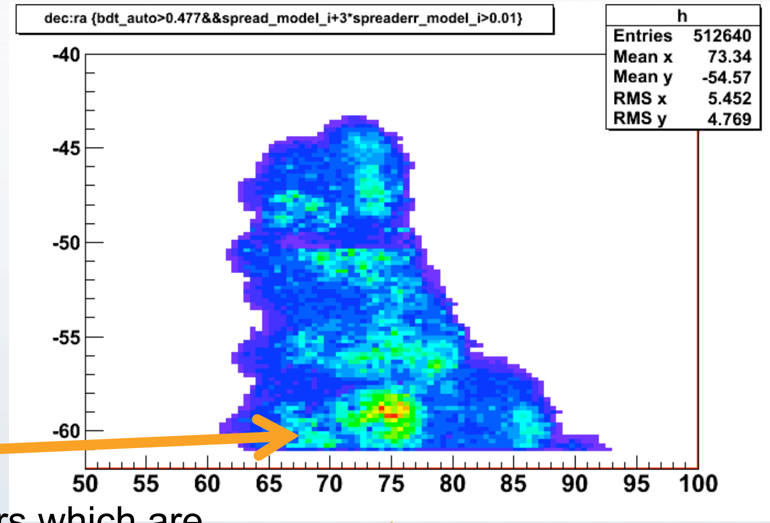
TPZ seems to generalize for ‘stars’ better than BDT or Random Forests (but slightly different behavior than modest).

‘Galaxy’ distributions are similar.

F. Sobreira, A.Drlica-Wagner

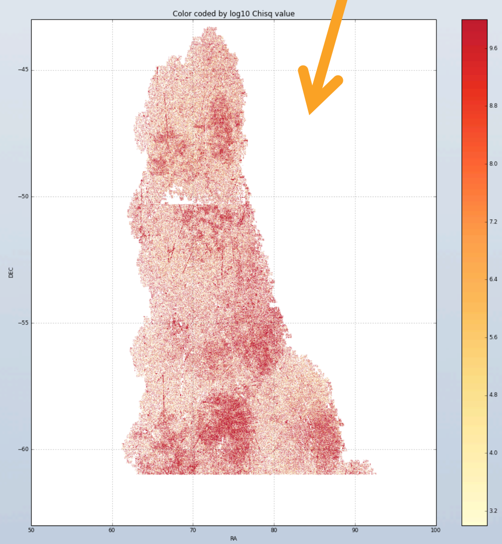
General distributions: modest outliers





A.Drlica-Wagner

These TPZ stars which are MODEST galaxies cluster like this…

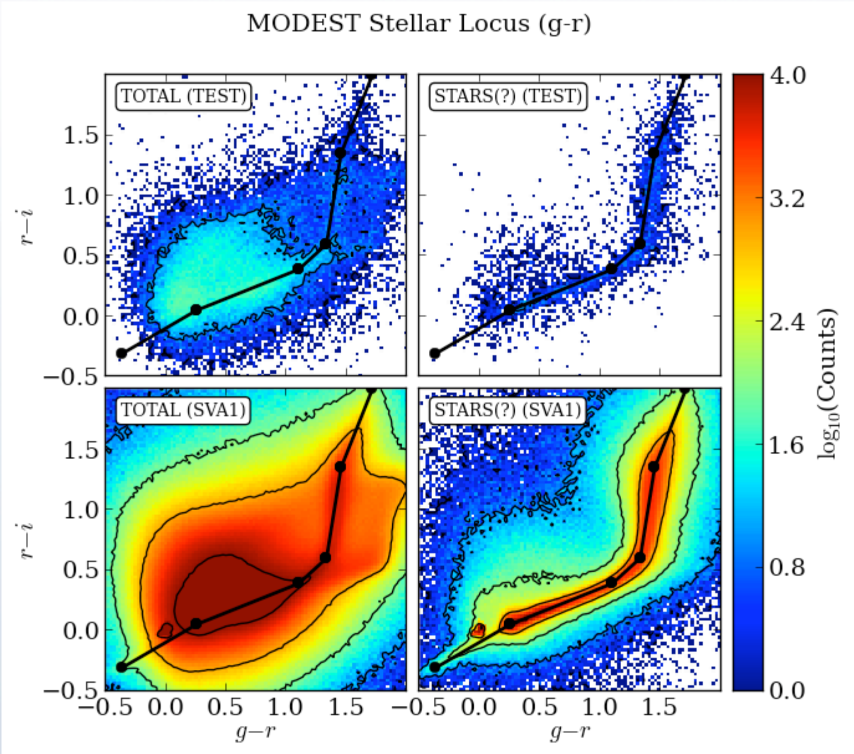


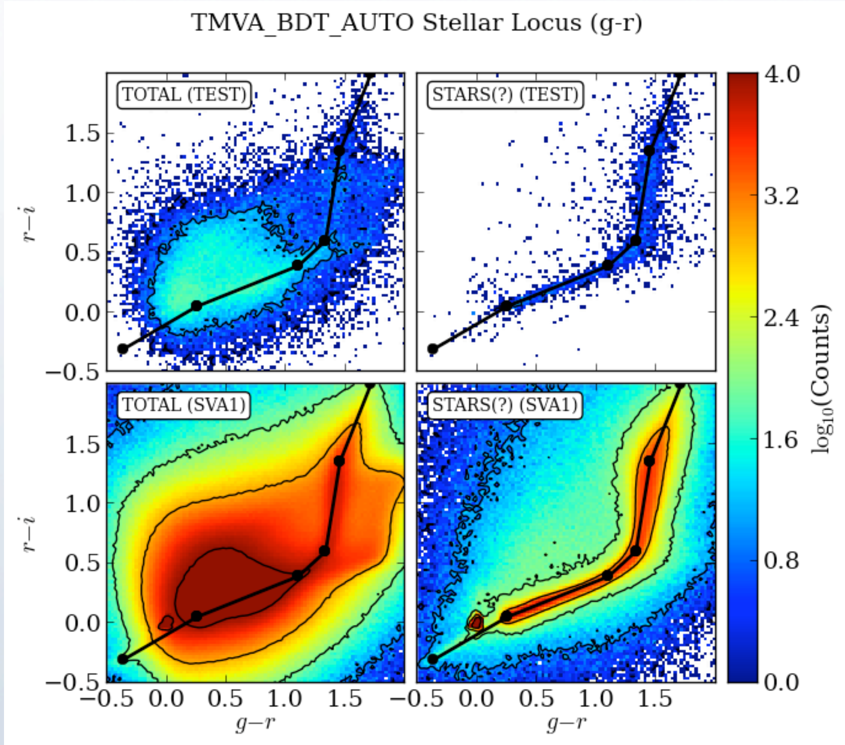
…just like bad fits to LePhare templates.

C.Bonnett, C.Sánchez

Some objects with large `SPREAD_MODEL` values actually are stars for machine learning methods AND have bad fits to LePhare galaxy templates.

General distributions: stellar locus

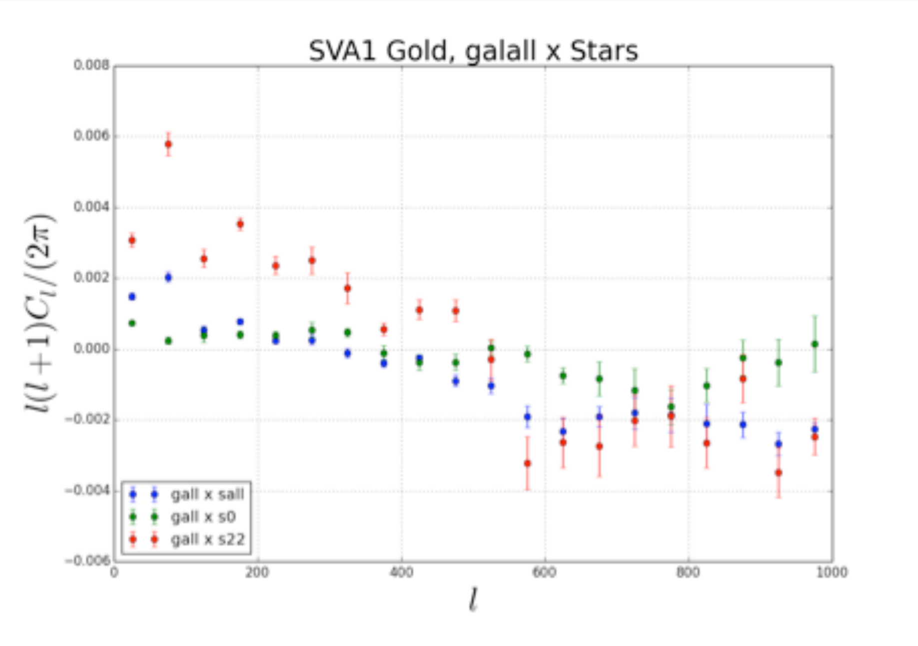


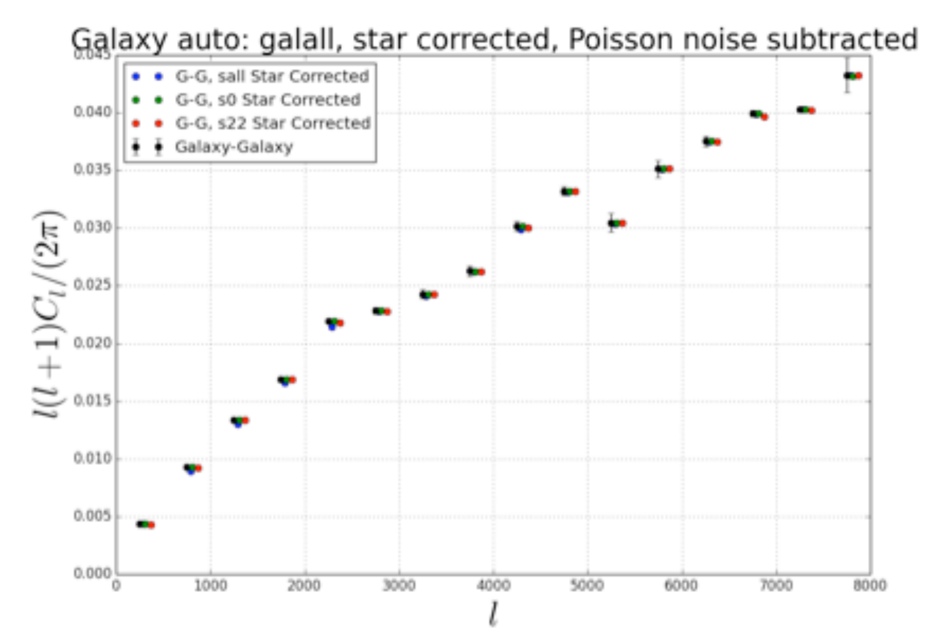


A.Drlica-Wagner

No large indicative differences seen in stellar locus with different classifiers.

Galaxy-star cross-correlations





R.Cawthon

Small but puzzling correlations at small and large scales.

Also work by FJ Sánchez shows negligible corrections for correlation functions.

We expect several improvements for MODEST and inputs for ML methods

•Multi-epoch spread_model, T-size

•QSOs

•Photo-zs as inputs

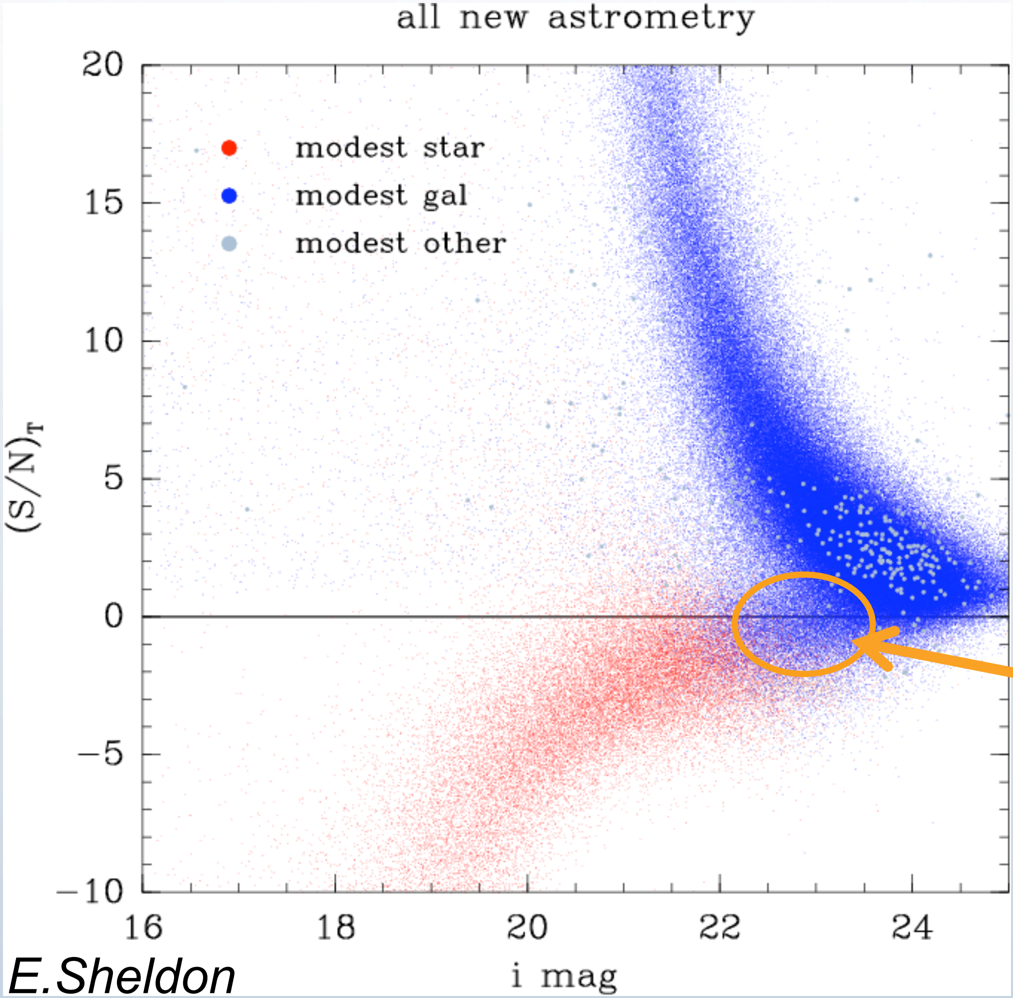
•Chi-square fits to galaxy templates

•White Dwarfs (*A.Drlica-Wagner says that maybe for whole survey area*)

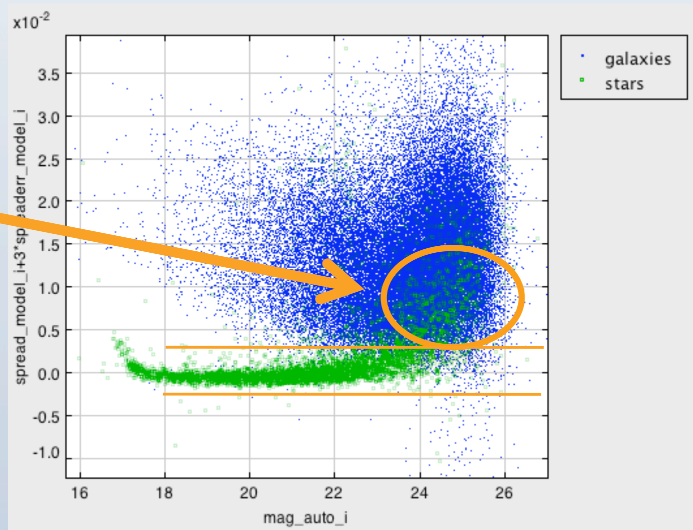
•Stripe82, COMBO-17 (CDFS) fields

•Systematizing procedure for future comparisons (vetting)

Improvements



Use S/N of size estimate of galaxy vs magnitude from single-epoch fits to postage stamp objects.



(note that the depth of this data is different)