STATISTICS
TEXAS A&M UNIVERSITY

Statistical Challenges in Astronomy

September 1, 2015

Classification of Variable Stars

Period Estimation for Variable Stars

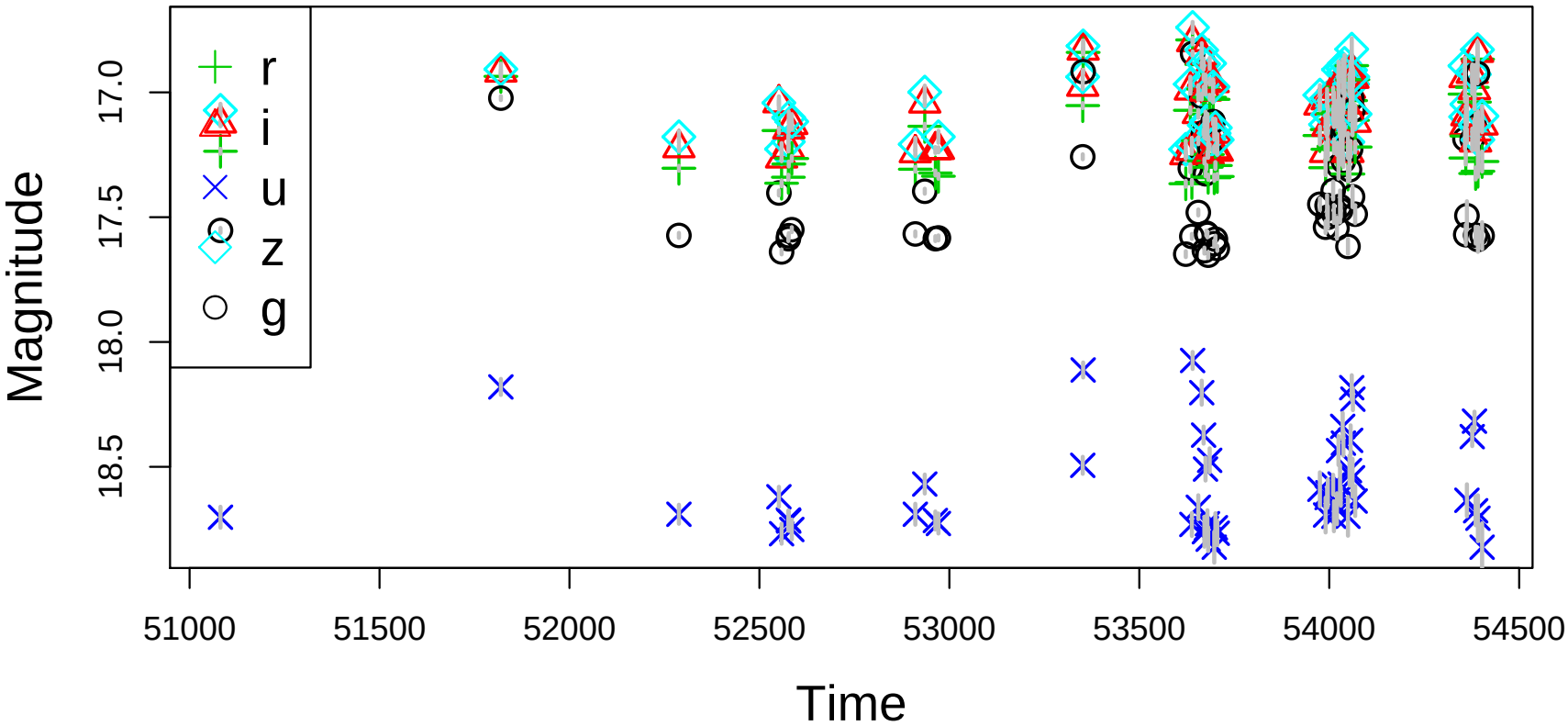
Photometric Redshift Estimation

Large Scale Structure of the Universe

Bayesian Hierarchical Modeling of Supernovae

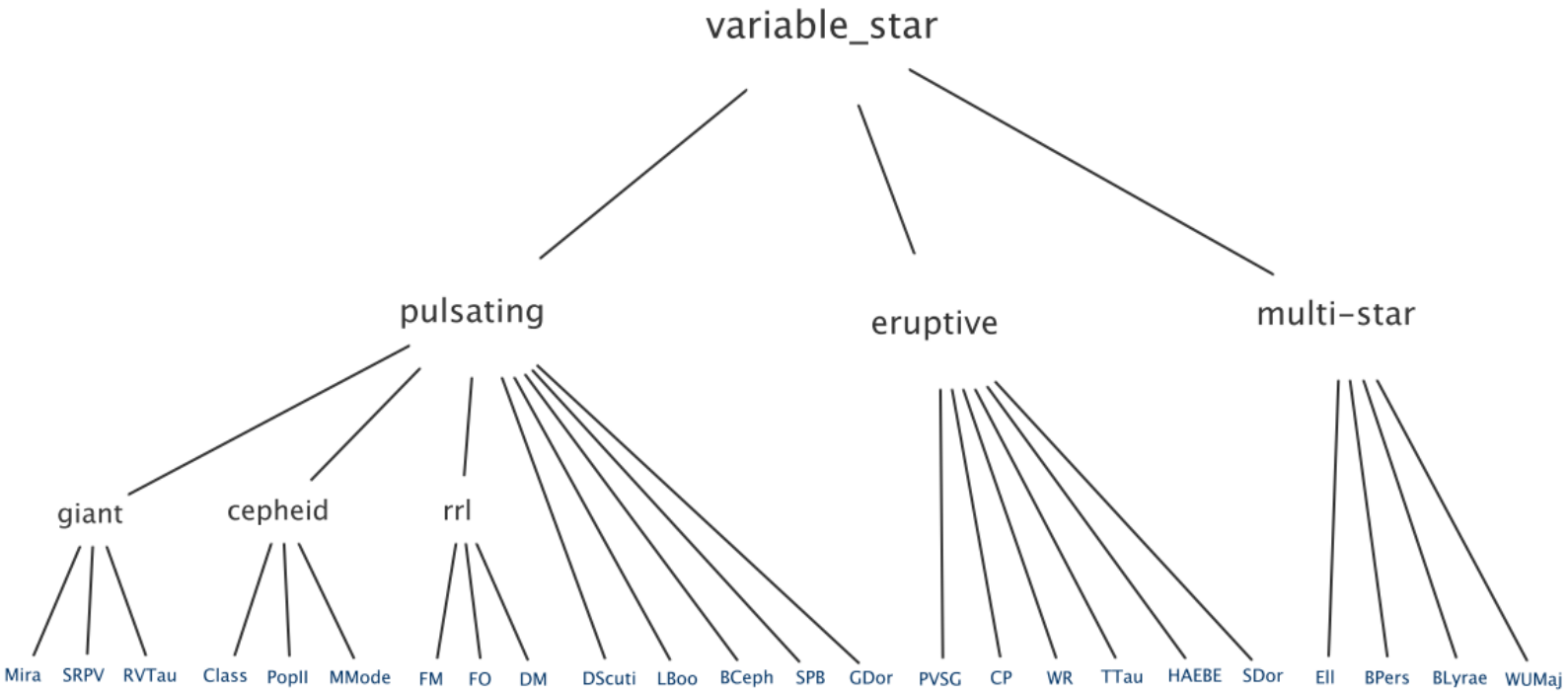
Outline

Stripe 82 Variable Star Light Curve



Surveys collect 100,000s or millions of these.

Classification Structure from [2]

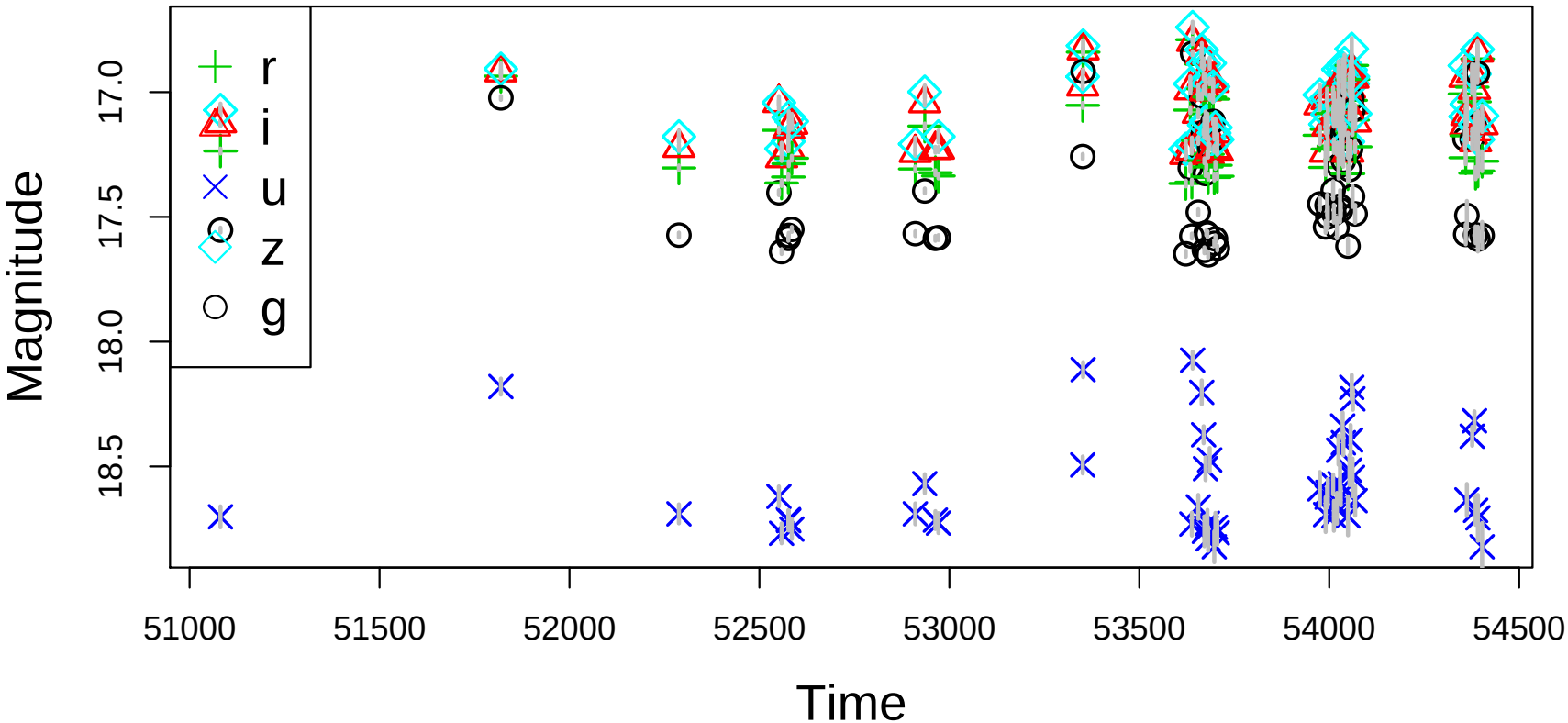


Major Themes

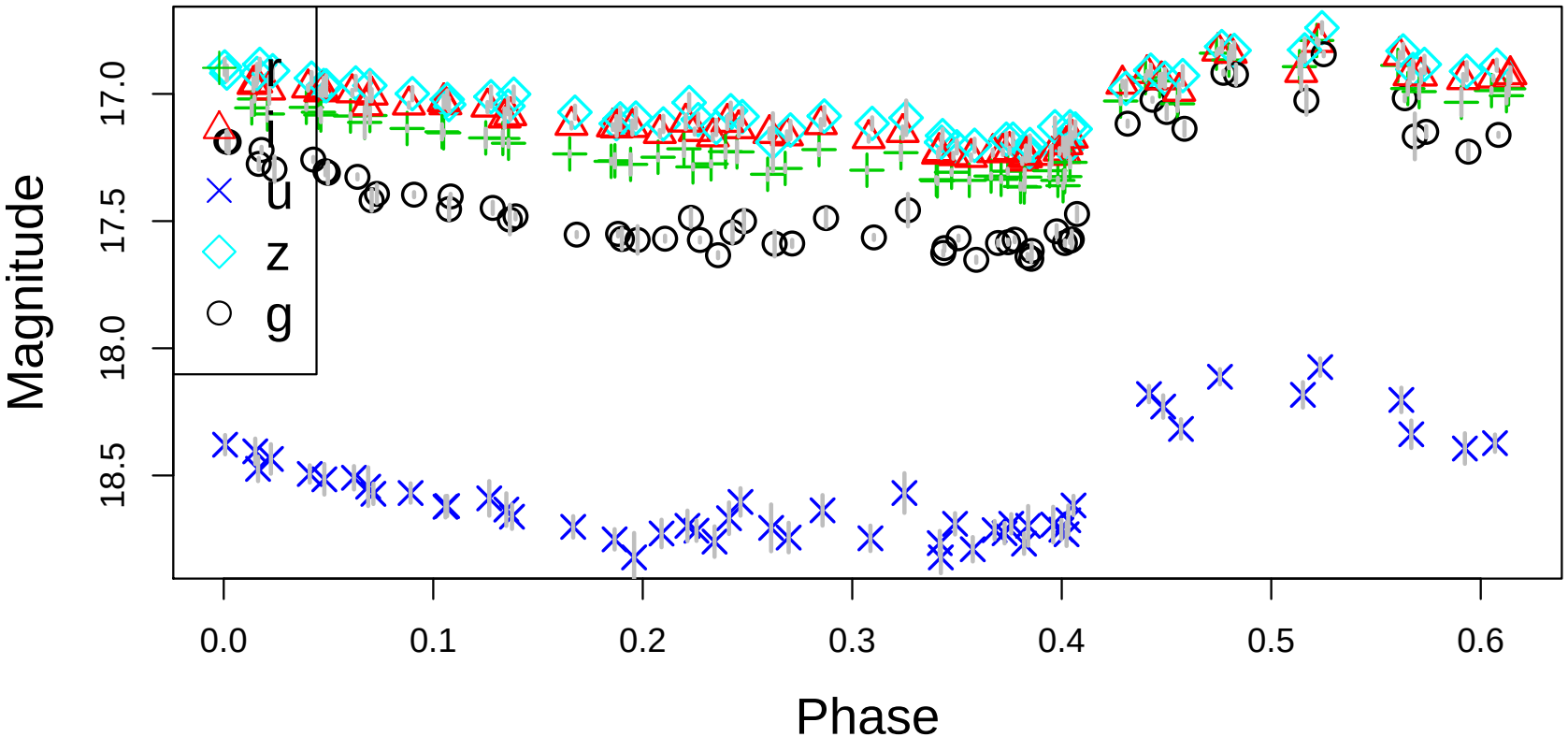
- statistical classifier
- assessing classifier accuracy
- multiclass / hierarchical classification
- classifier using multiband data
- non identically distributed training / test sets
- feature extraction versus modeling
- possible data source: sdss stripe 82, ogle

Outline

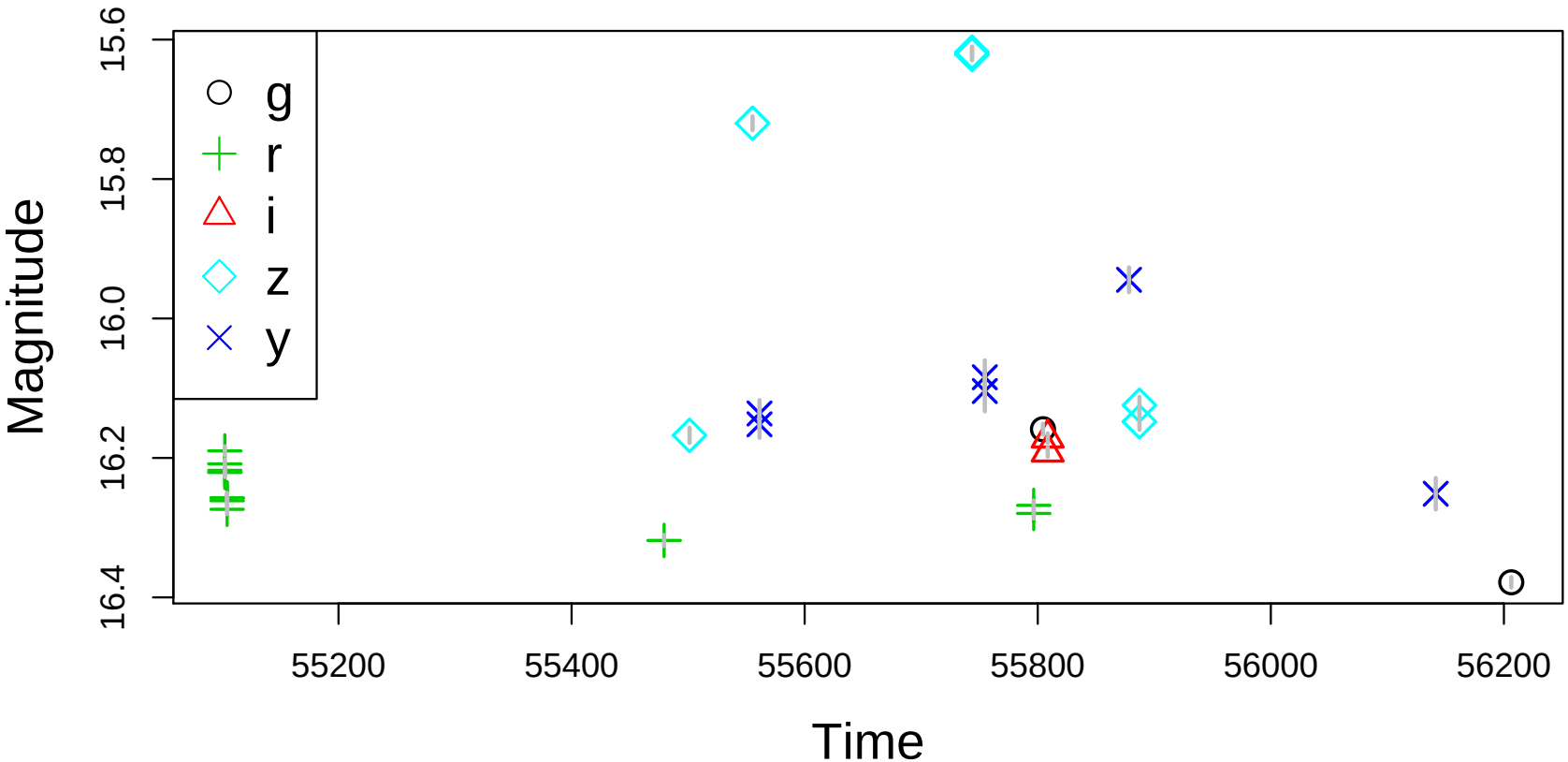
SDSS Stripe 82 RR Lyrae



SDSS Stripe 82 RR Lyrae



PanStarrs RR Lyrae

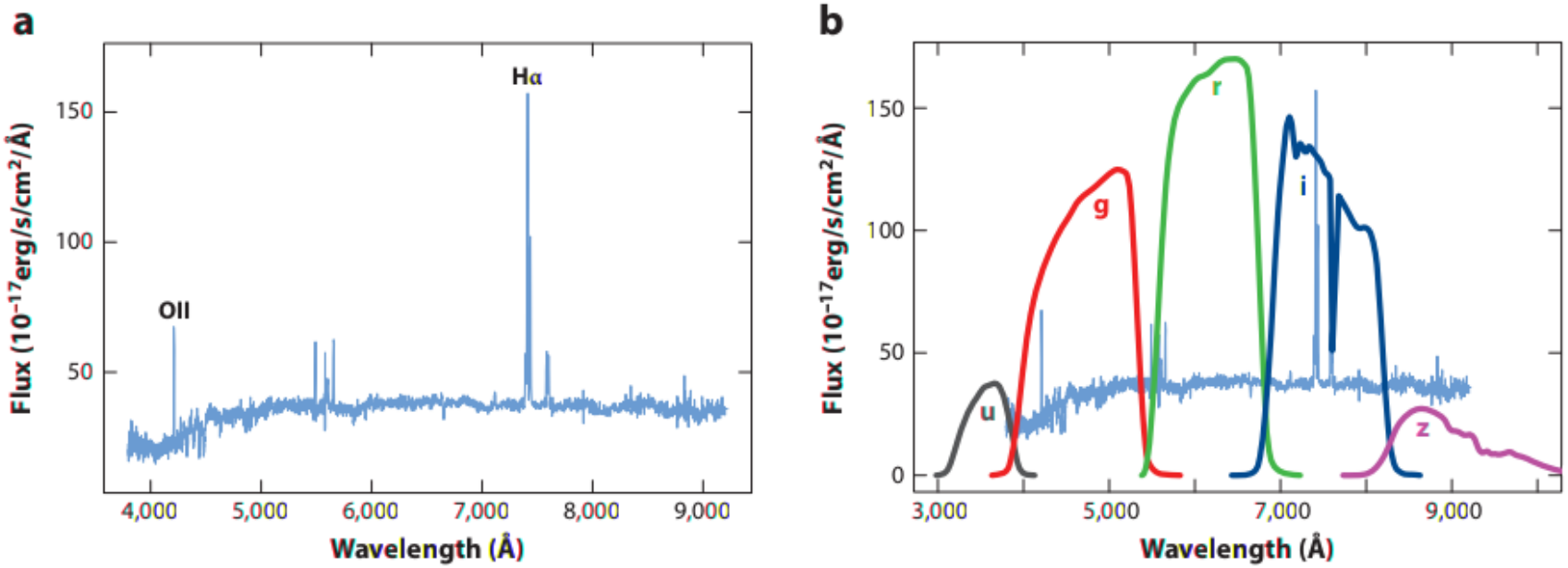


Major Themes

- fourier analysis
- measurement error / model misspecification
- template methods
- model building
- fast algorithms
- possible data source: panstarrs I rr lyrae

Outline

Spectroscopy and Photometry



From [4]. With spectroscopy (left) we can easily calculate redshift. Spectroscopy is expensive so we collect much more photometric data (right). Want to predict redshift only using photometric measurements.

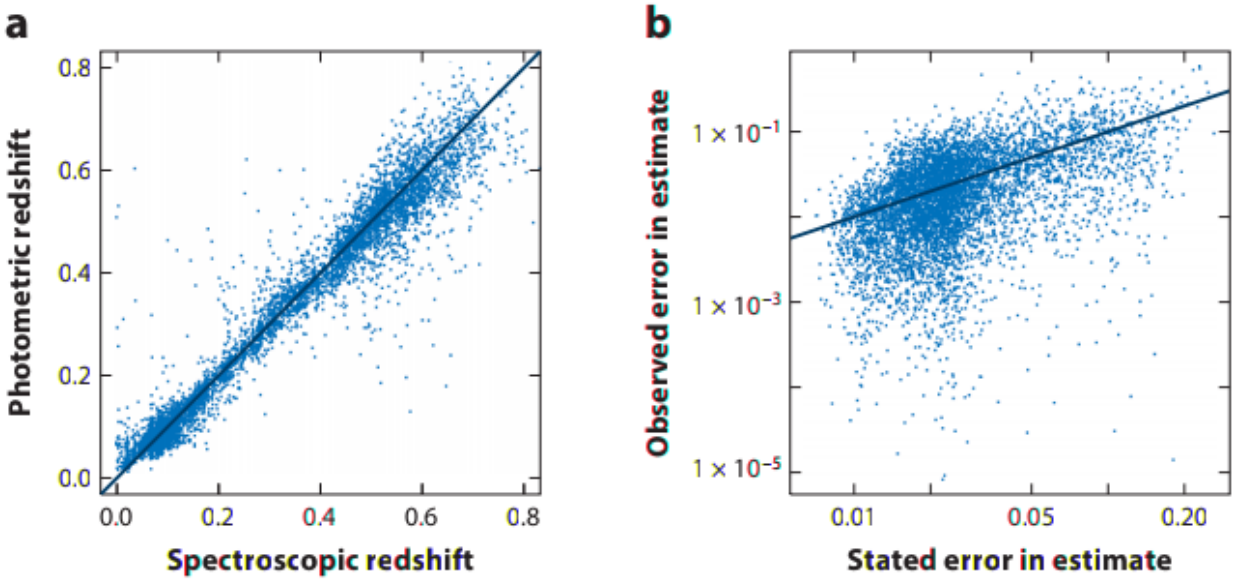
Quasar Photoz Dataset

13 rows out of $\approx 50,000$

SDSS_J	R.A.	Dec.	z	u_mag	sig_u	g_mag	sig_g	r_mag	sig_r	i_mag	sig_i	z_mag	sig_z
000009.26+151754.5	0.038605	15.298476	1.1986	19.921	0.042	19.811	0.036	19.386	0.017	19.165	0.023	19.323	0.069
000009.38+135618.4	0.039088	13.938447	2.2400	19.218	0.026	18.893	0.022	18.445	0.018	18.331	0.024	18.110	0.033
000009.42-102751.9	0.039269	-10.464428	1.8442	19.249	0.036	19.029	0.027	18.980	0.021	18.791	0.018	18.751	0.047
000011.41+145545.6	0.047547	14.929353	0.4596	19.637	0.030	19.466	0.024	19.362	0.022	19.193	0.025	19.005	0.047
000011.96+000225.3	0.049842	0.040372	0.4790	18.237	0.028	17.971	0.020	18.025	0.019	17.956	0.014	17.911	0.029
000013.14+141034.6	0.054786	14.176303	0.9491	19.519	0.034	19.281	0.028	19.115	0.016	19.155	0.024	19.071	0.053
000017.38-085123.7	0.072421	-8.856607	1.2499	19.151	0.034	18.722	0.020	18.263	0.021	18.276	0.036	18.260	0.037
000024.02+152005.4	0.100116	15.334840	0.9885	19.413	0.044	19.183	0.035	18.988	0.015	19.079	0.023	19.133	0.060
000026.29+134604.6	0.109578	13.767970	0.7678	19.345	0.030	18.998	0.023	18.922	0.023	19.010	0.022	18.838	0.042
000028.82-102755.7	0.120086	-10.465496	1.1377	20.461	0.086	19.697	0.030	19.176	0.016	19.143	0.023	19.105	0.061
000035.75-103305.3	0.148966	-10.551496	1.2177	19.404	0.041	19.455	0.029	19.045	0.015	19.006	0.023	19.181	0.066
000038.99-001803.9	0.162498	-0.301102	2.1224	19.204	0.044	19.076	0.022	18.886	0.017	18.894	0.018	18.794	0.044
000042.89+005539.5	0.178746	0.927660	0.9512	18.353	0.037	18.150	0.015	17.941	0.011	17.899	0.021	17.802	0.024

z column is calculated from spectroscopy. Use `filter_mag` and `sig_filter` columns to predict z.

Photometric Redshift Estimation



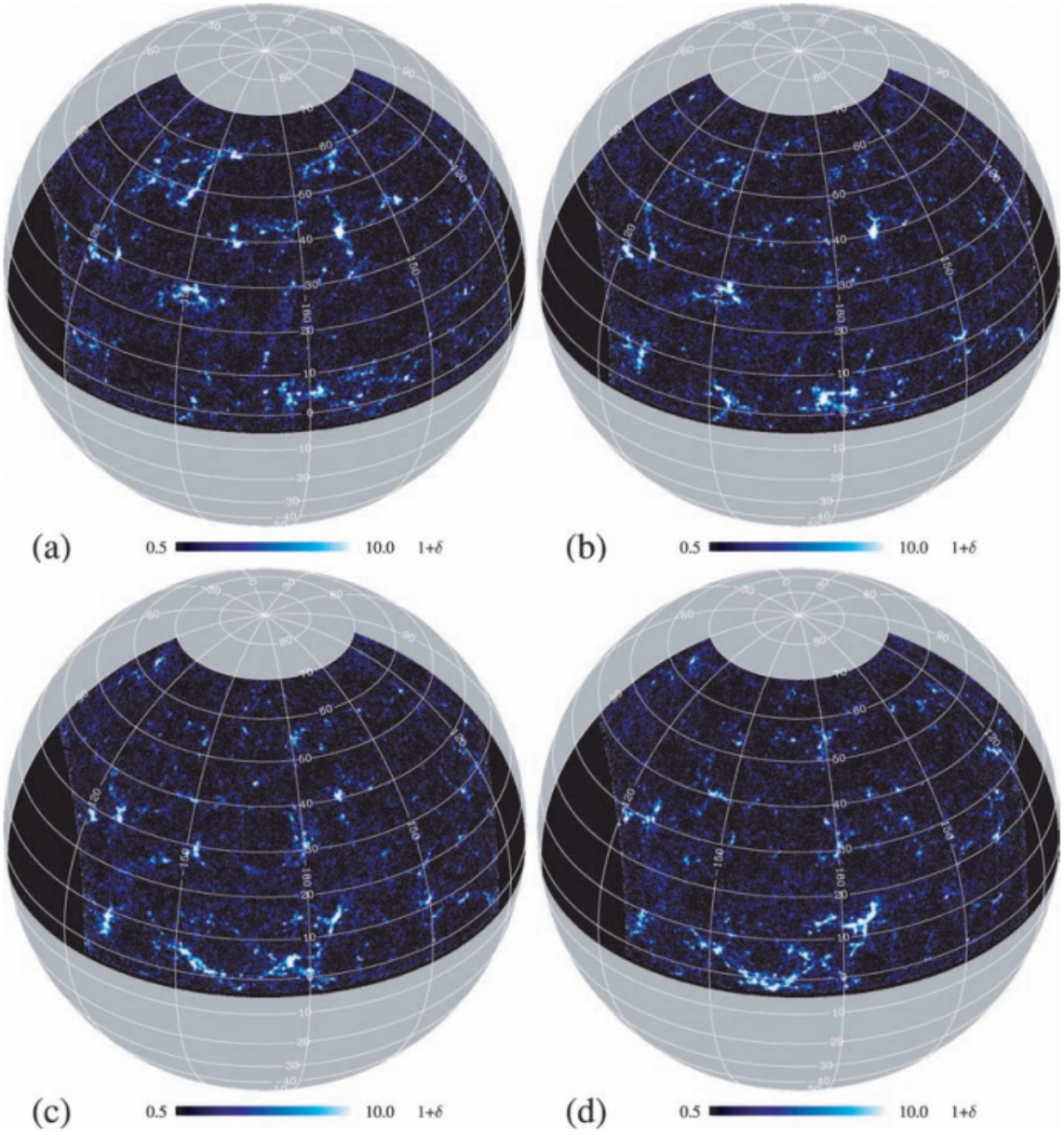
From [4]. a) Spectroscopic redshift (truth) versus redshift predicted by photometry.

Major Themes

- machine learning
- incorporation of measurement error into predictive model
- possible data: sdss or candles

Outline

Sloan Great Wall from [1]

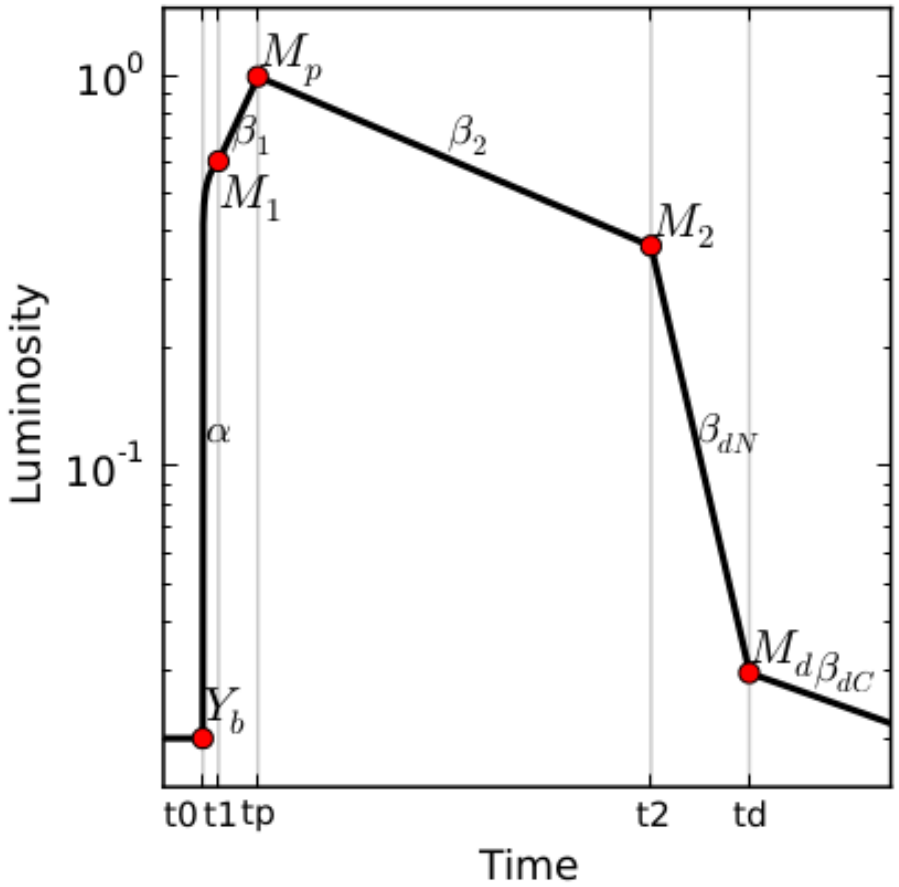


Major Themes

- spatial statistics
- clustering algorithms

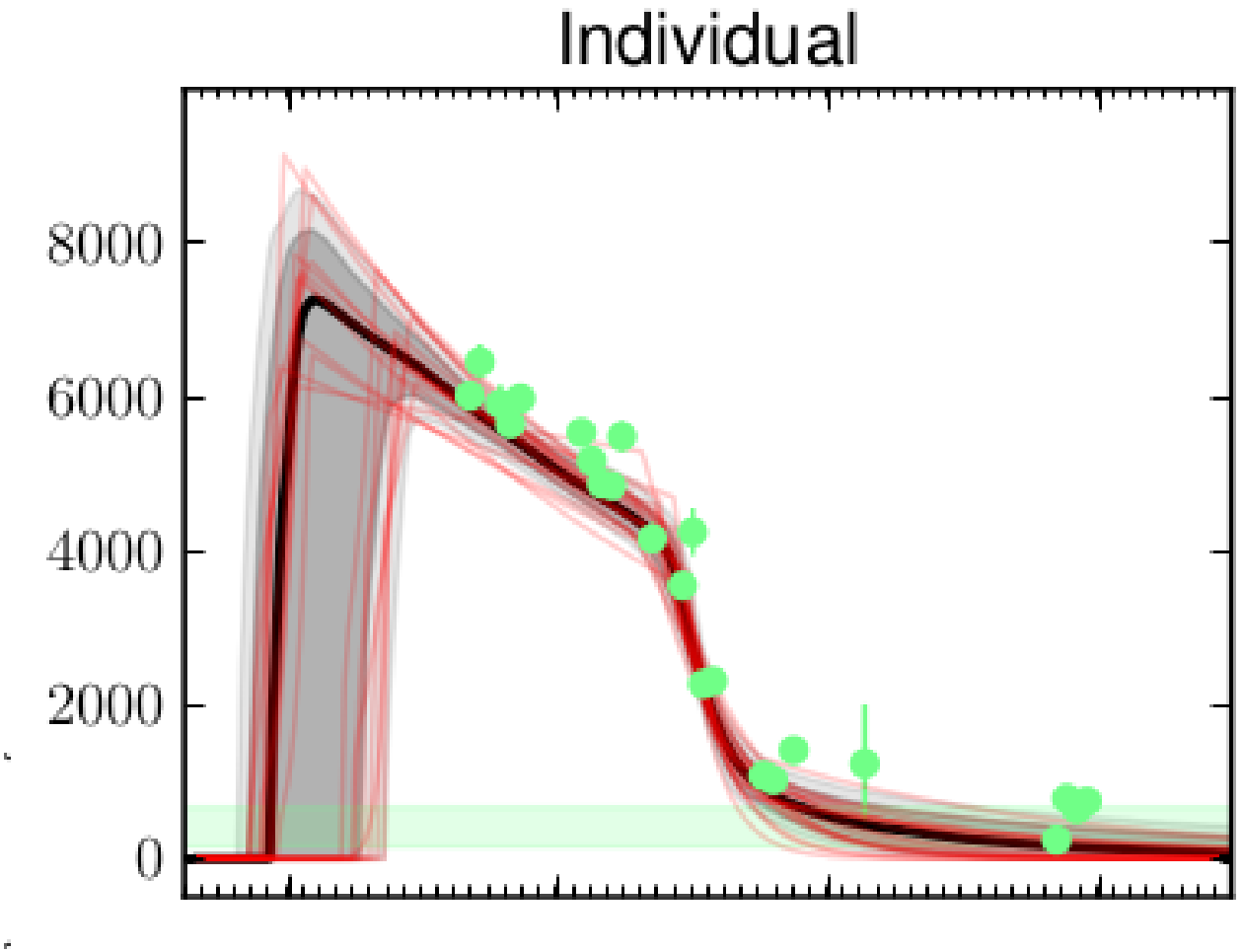
Outline

Piecewise Linear Supernovae Model

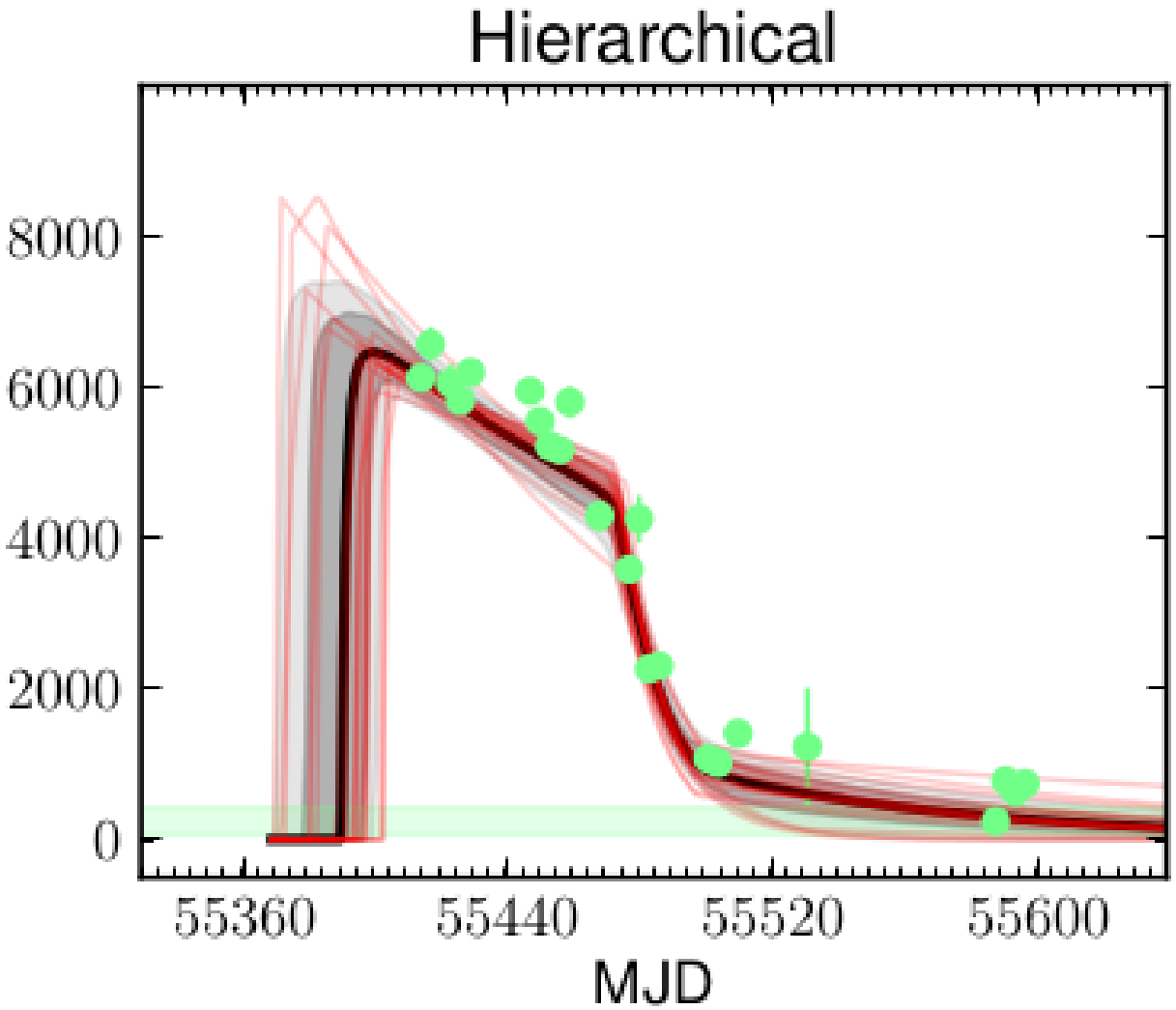


Piecewise linear Type IIP supernovae lightcurve model from [3].

Fit for PanStarrs Supernovae



Incorporate Information from Other Supernovae



Major Themes

- Bayesian hierarchical models
- MCMC
- Hierarchical models for other types of data
  - Mira light curves
  - Galaxy SED fitting
  - Type Ia Supernovae
    - Cosmology, estimation of Ω_m and Ω_Δ

Classification of Variable Stars

Period Estimation for Variable Stars

Photometric Redshift Estimation

Large Scale Structure of the Universe

Bayesian Hierarchical Modeling of Supernovae

Bibliography I

[1] Francisco S Kitaura, Jens Jasche, Cheng Li, Torsten A Enßlin, R Benton Metcalf, Benjamin D Wandelt, Gerard Lemson, and Simon DM White.
Cosmic cartography of the large-scale structure with sloan digital sky survey data release 6.
Monthly Notices of the Royal Astronomical Society, 400(1):183–203, 2009.

[2] Joseph W Richards, Dan L Starr, Nathaniel R Butler, Joshua S Bloom, John M Brewer, Arien Crellin-Quick, Justin Higgins, Rachel Kennedy, and Maxime Rischard.
On machine-learned classification of variable stars with sparse and noisy time-series data.
The Astrophysical Journal, 733(1):10, 2011.

[3] NE Sanders, AM Soderberg, S Gezari, M Betancourt, R Chornock, E Berger, RJ Foley, P Challis, M Drout, RP Kirshner, et al.
Toward characterization of the type iip supernova progenitor population: A statistical sample of light curves from pan-starrs1.
The Astrophysical Journal, 799(2):208, 2015.

[4] Chad M Schafer.
A framework for statistical inference in astrophysics.
Annual Review of Statistics and Its Application, 2:141–162, 2015.