MACHINE LEARNING-BASED TRANSIENT BROKERS FOR REAL-TIME CLASSIFICATION OF THE LSST ALERT STREAM

GAUTHAM NARAYAN
STSCI
LASKER DATA SCIENCE FELLOW

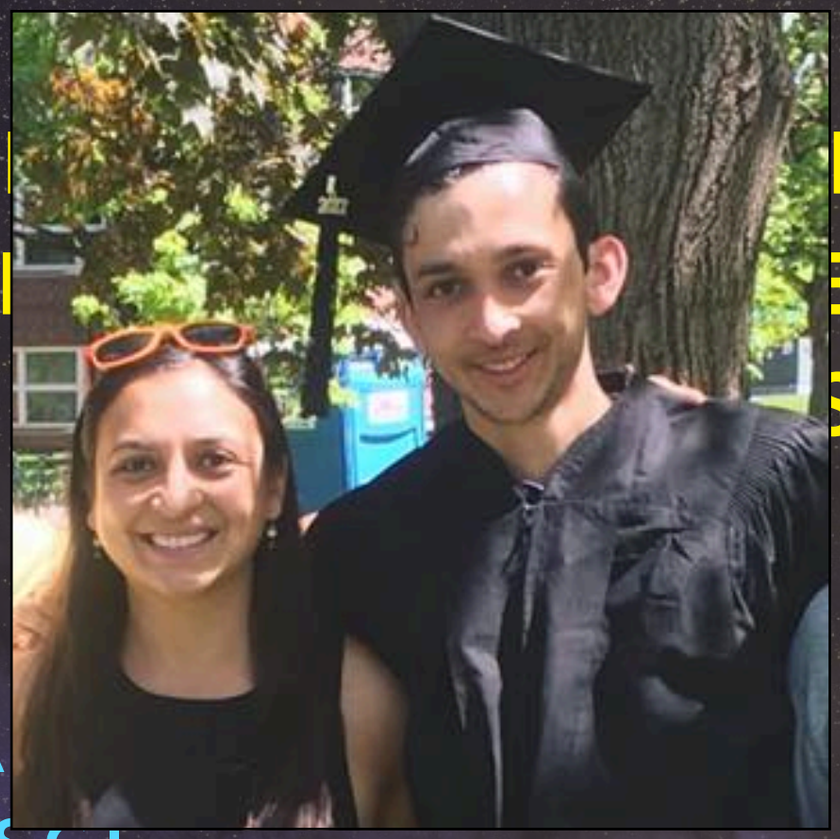
@gsnarayan

NOAO: Monika Soraisam, Abhijit Saha, Tom Matheson
UA CS: Rick Snodgrass, John Kececioglu, Carlos Scheidegger
UA Math/Statistics: Robert Maier
UA Astro: Tim Axelrod
CSS: Rob Seaman
LSST: Tim Jenness

Grads: Zhe Wang, Clark Taylor, Eric Welch,
Zhenge Zhao, Eric Evans,
Shuo Yang, Navdeep Singh

Former students: Tayeb Zaidi, Jackson Toeniskotter

NING-B
EAL-TIM
SST ALE
NT
ON



GA
STSCI
LASKER DATA SCIENCE FELLOW

@gsnarayan

ANTARES

NOAO: Monika Soraisam, Abhijit Saha, Tom Matheson
UA CS: Rick Snodgrass, John Kececioglu, Carlos Scheidegger
UA Math/Statistics: Robert Maier
UA Astro: Tim Axelrod
CSS: Rob Seaman
LSST: Tim Jenness

Grads: Zhe Wang, Clark Taylor, Eric Welch,
Zhenge Zhao, Eric Evans,
Shuo Yang, Navdeep Singh

Former students: Tayeb Zaidi, Jackson Toeniskotter

Motivation

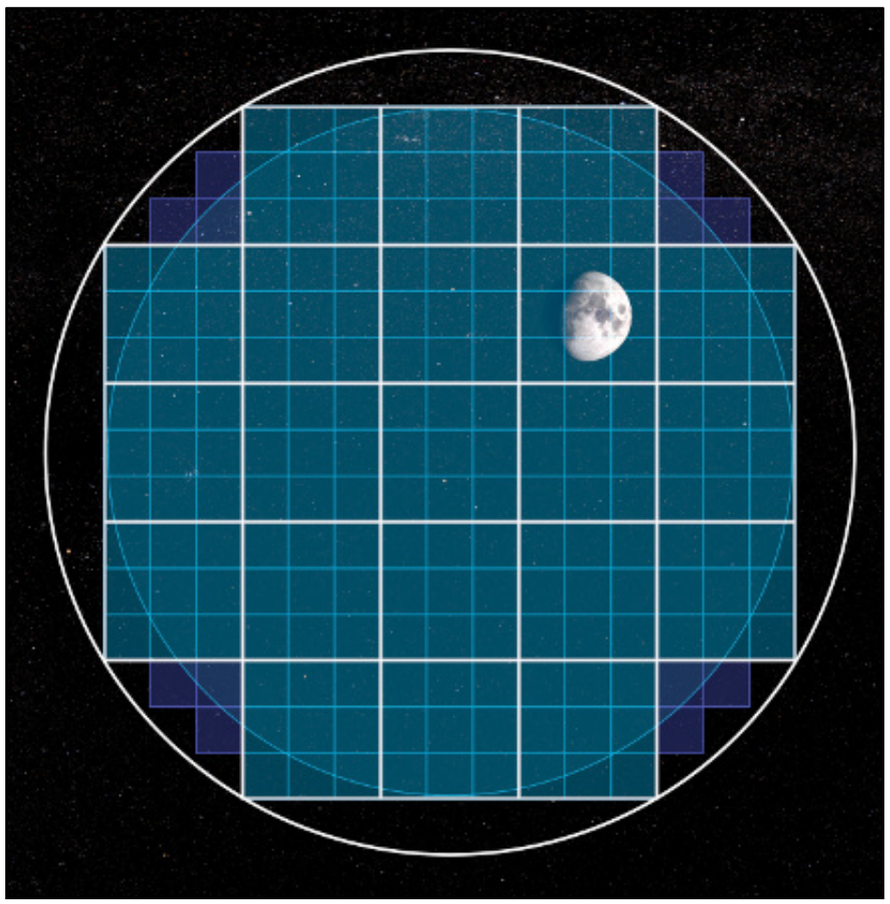
The LSST era is now! From this morning on Cerro Pachon.

We need useful alert brokers **before** first light.

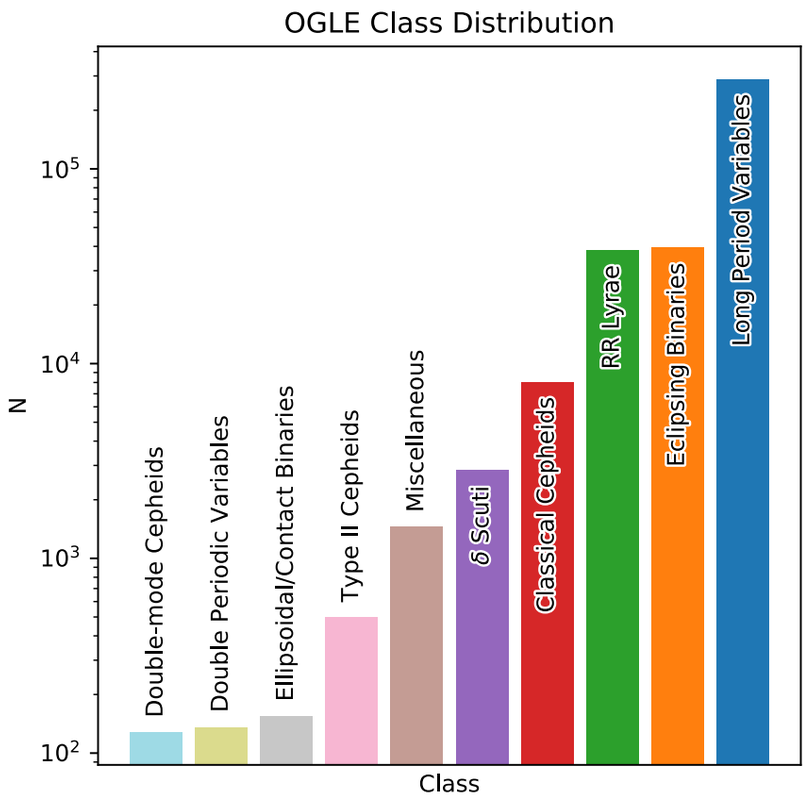
Example broker with real stages and real data in Narayan+ (2018, submitted)

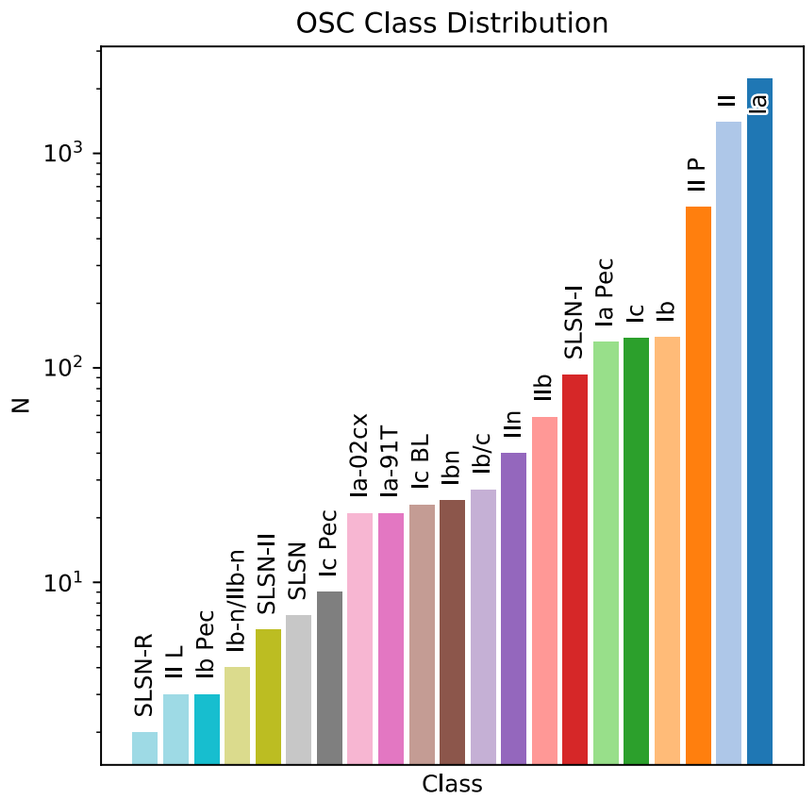
PS1 produced O(10^2) new alerts per night.
We sorted them visually, and could follow ~5/night spectroscopically.

LSST will produce O(10^7) per night.
Theres's no hope to sort this visually.
We can still follow ~5.



Really big, imbalanced, multi class data set of time-series goes in to pipeline.





Datasets | **Filtering** | **Early Categorization** | **Feature Extraction** | **Intermediate Phase Classification**

OGLE — Sec 2.1

OSC — Sec 2.3

VPDF

Timescale Characterization — Sec 3.2

Variable-Transient Split — Sec 4.2.1

Variable Sources

Transients

Feature Derivation — Sec 3.4

Variable-Transient Classifier — Sec 4.2.2

Labelled Output

SNPhotCC Simulated SNe — Sec 2.2

Gaussian Process Regression — Sec 3.4.1

BAGIDIS

SWT

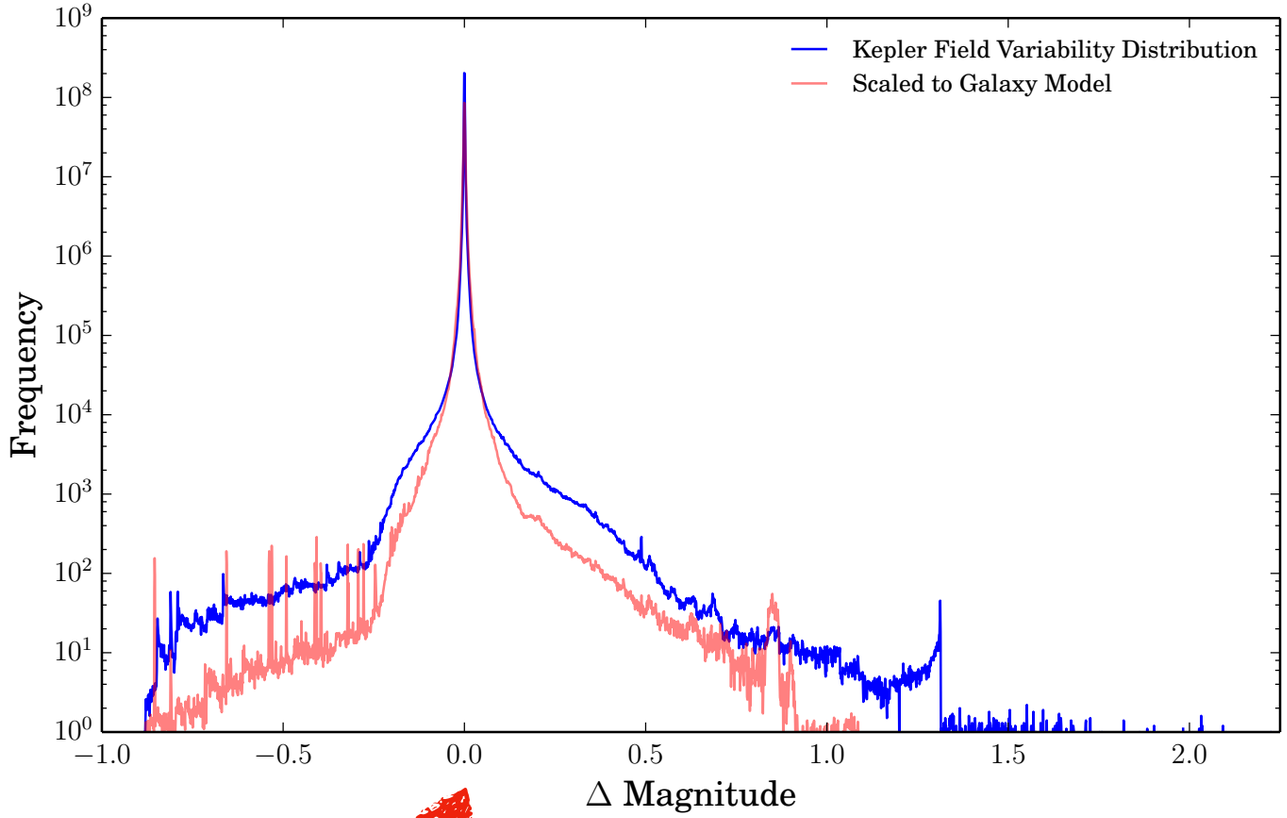
PCA

Sec 3.4.2

Supernova Classifer — Sec 4.3

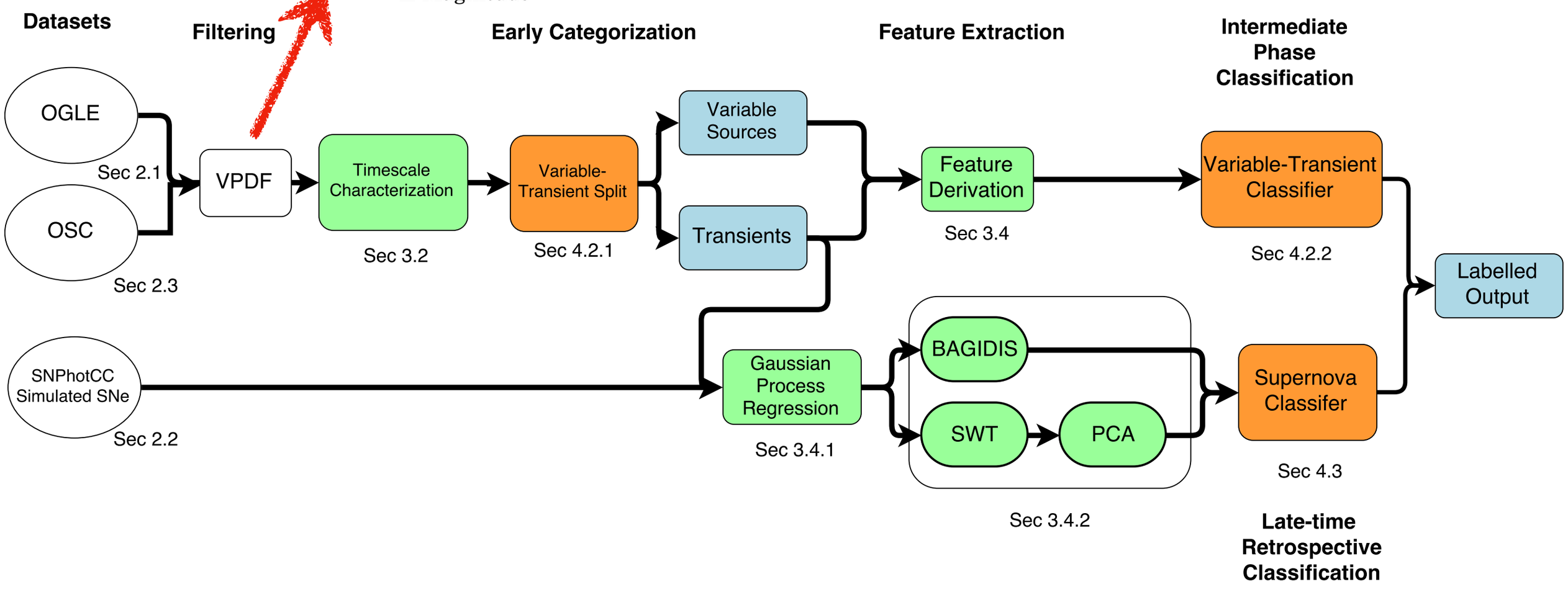
Late-time Retrospective Classification

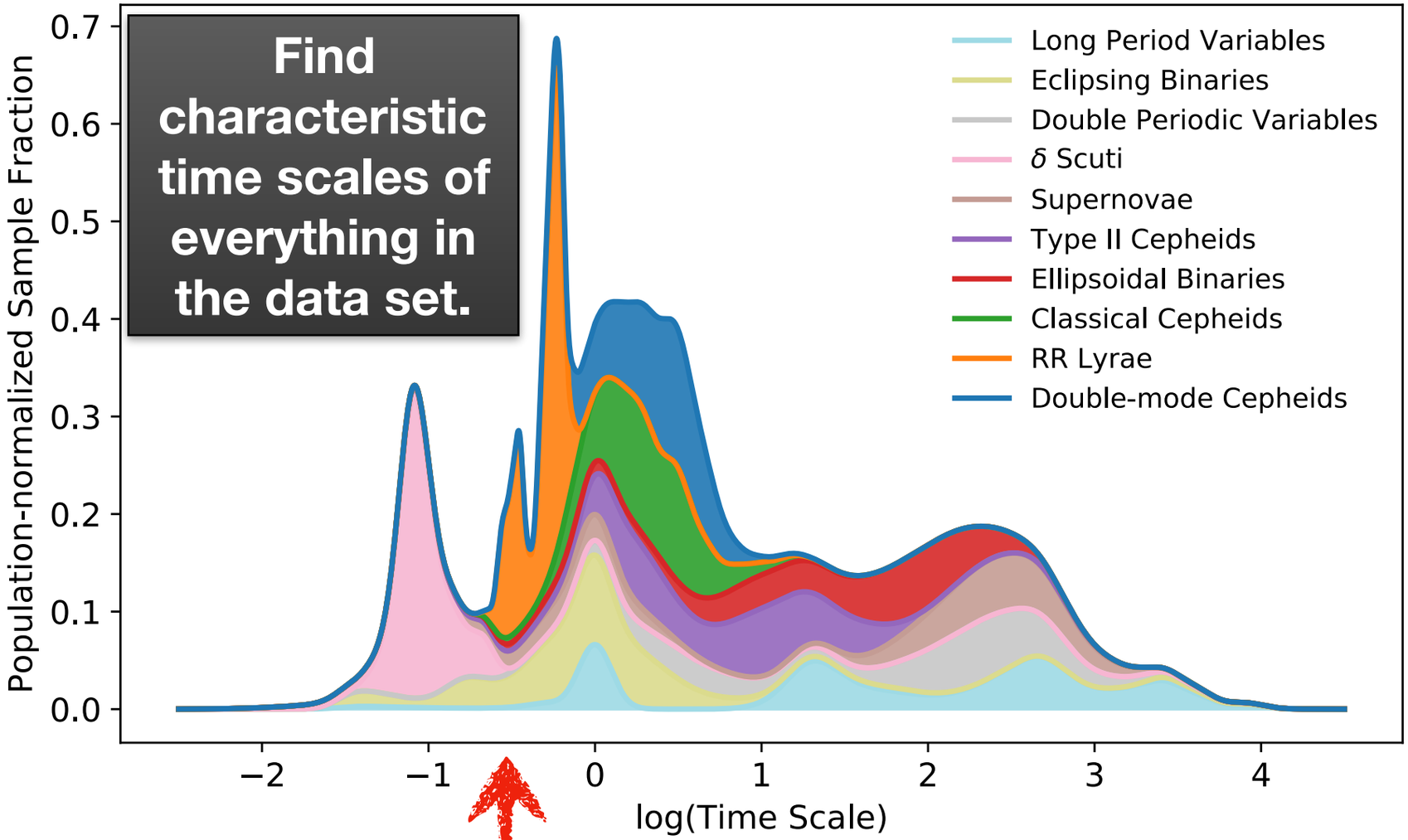
Narayan et al, (2018, submitted) pipeline implements core processing of ANTARES

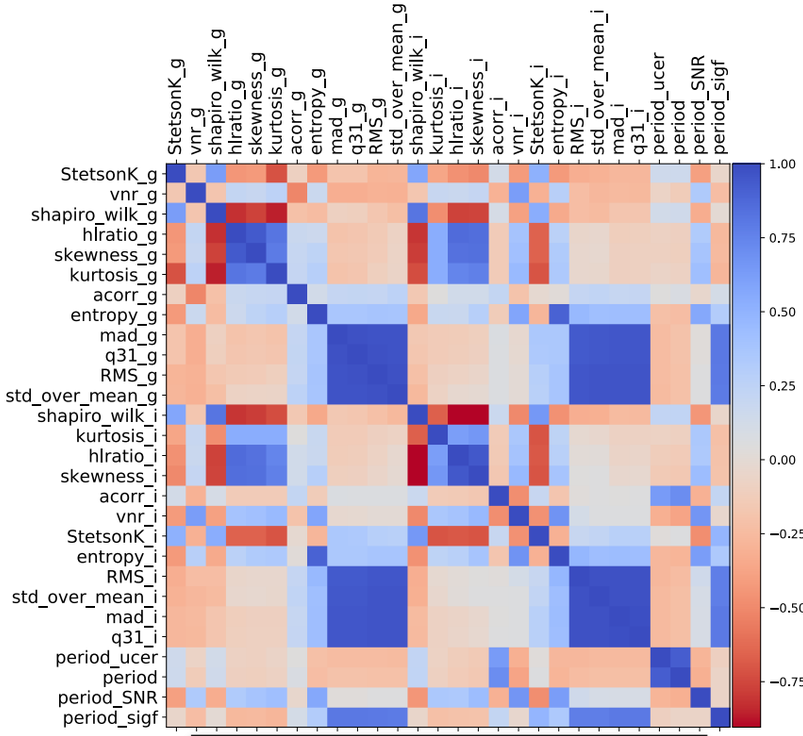


Use known Galactic stellar variability distribution from Kepler to characterize common variables vs explosive transients even with only a few observations (Ridgway et al, 2014)

Can be extended to multi-phase, multi-band (Soraisam et al, in prep)

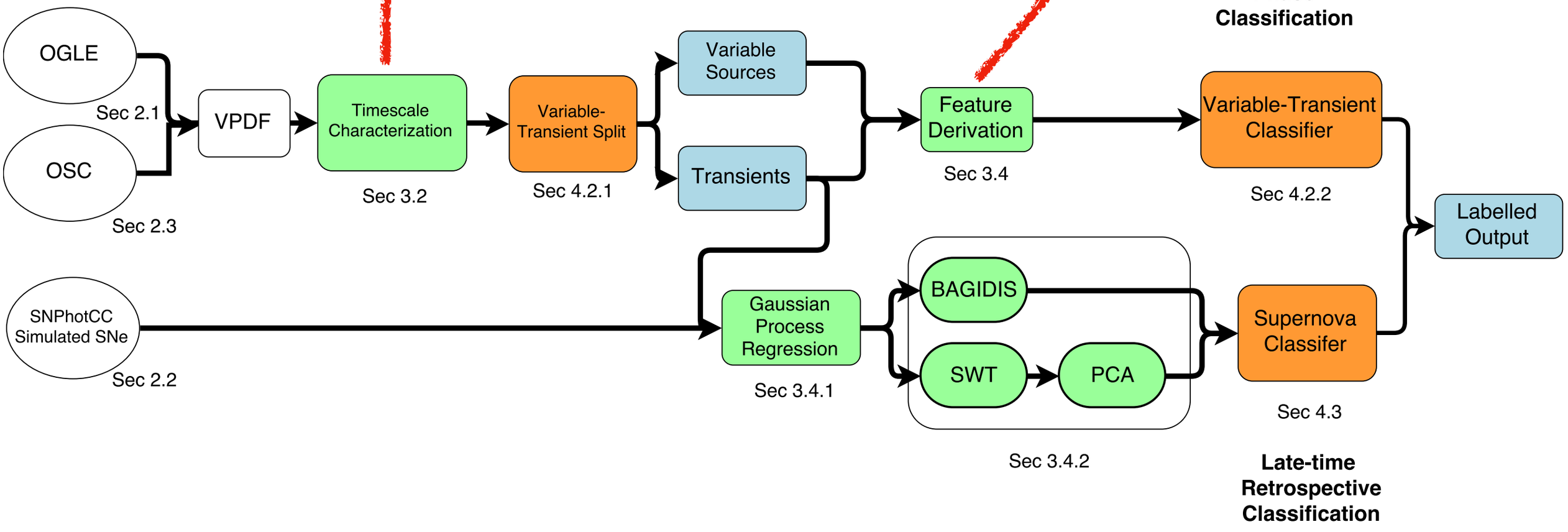


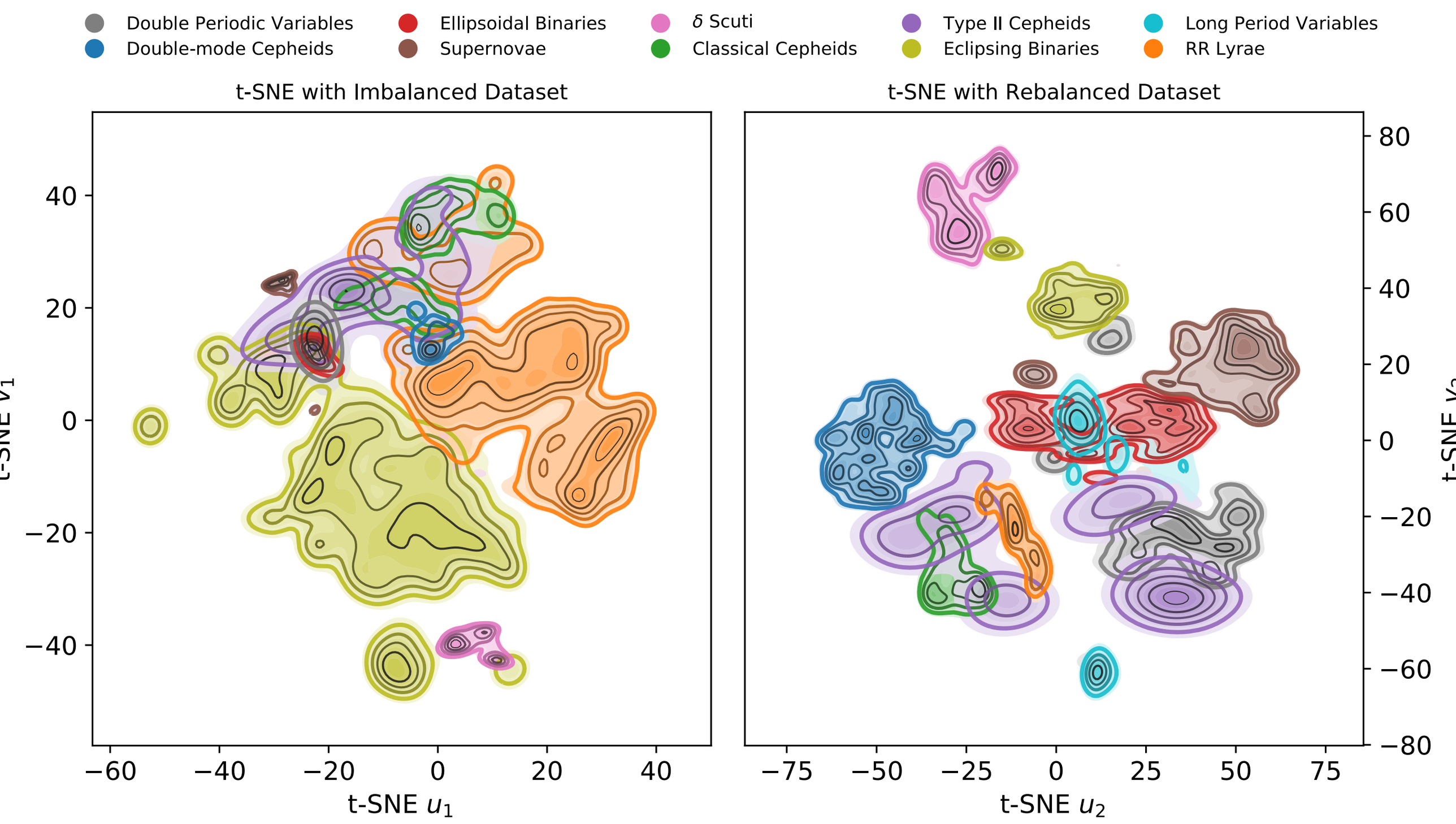




Account for correlations in features.

Intermediate Phase Classification





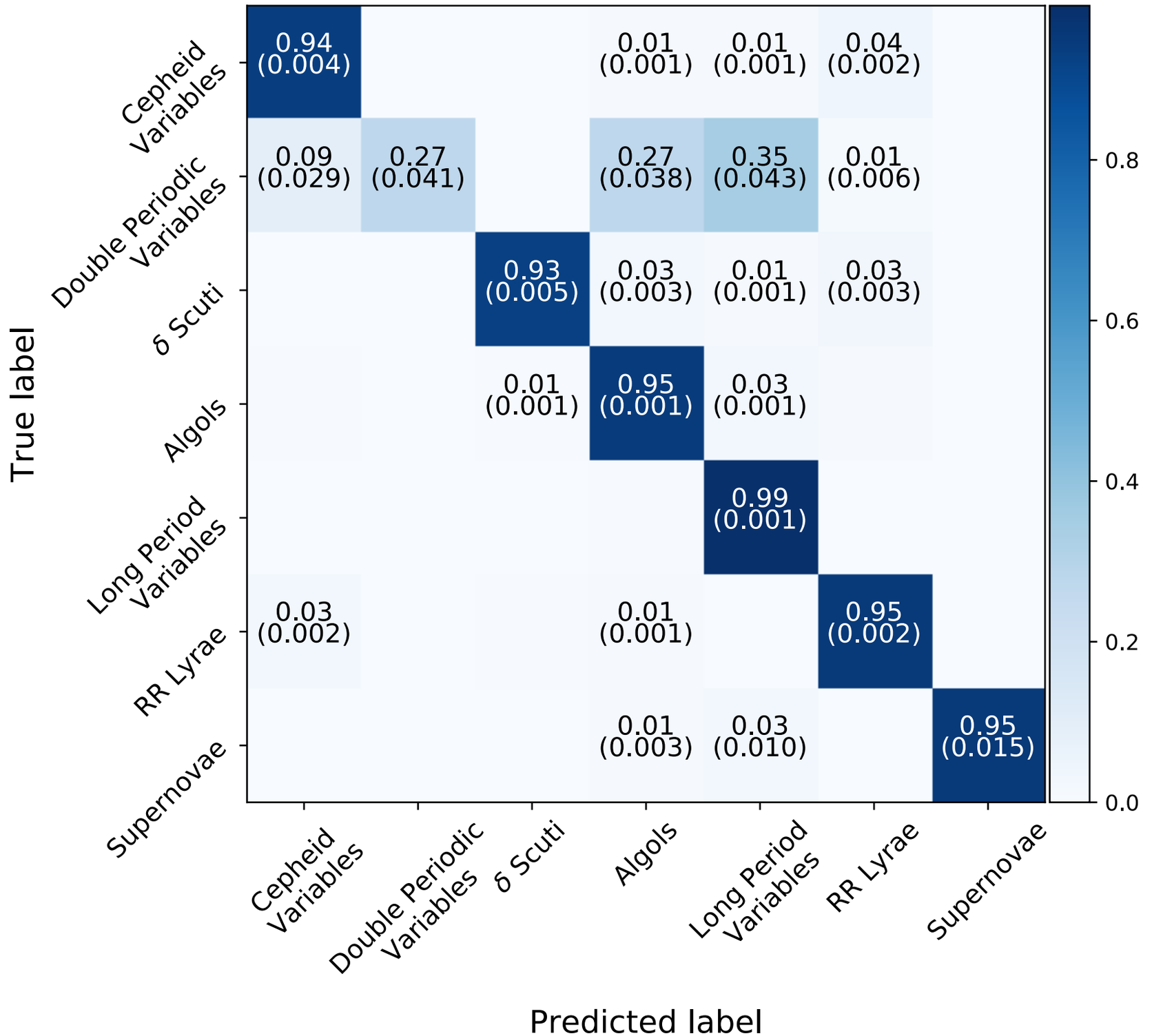
t-SNE to visualize this feature set

SMOTE to deal with class imbalance dramatically helps with distinguishing classes

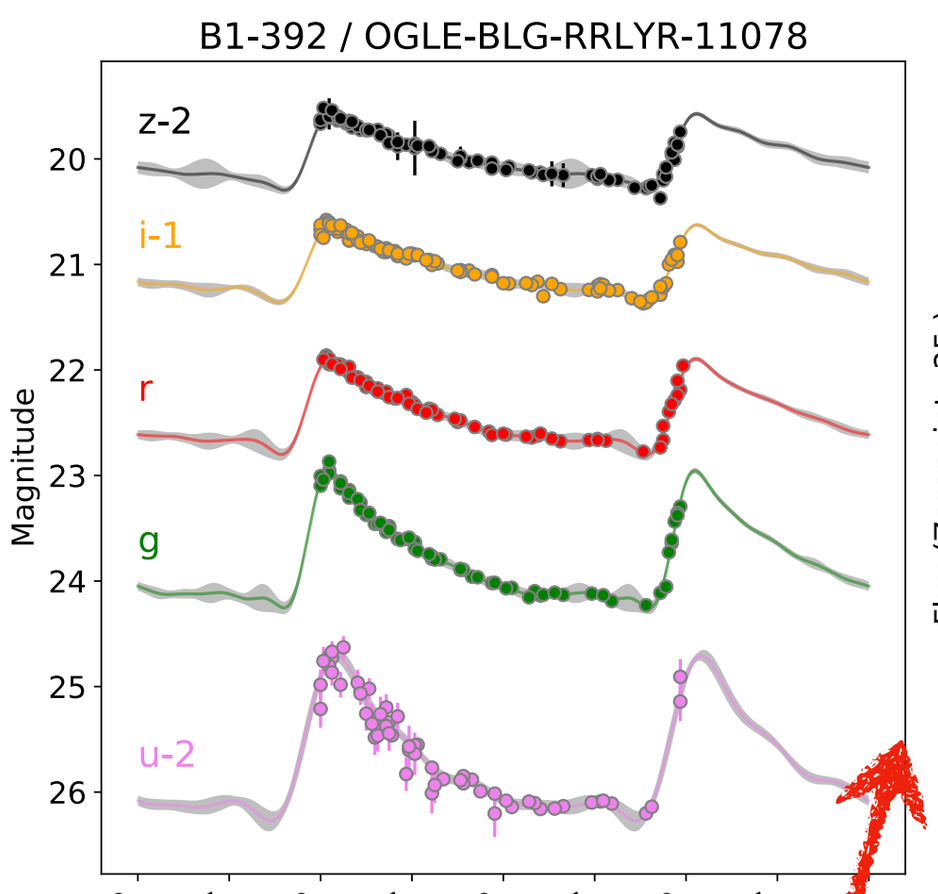
Intermediate-phase classification

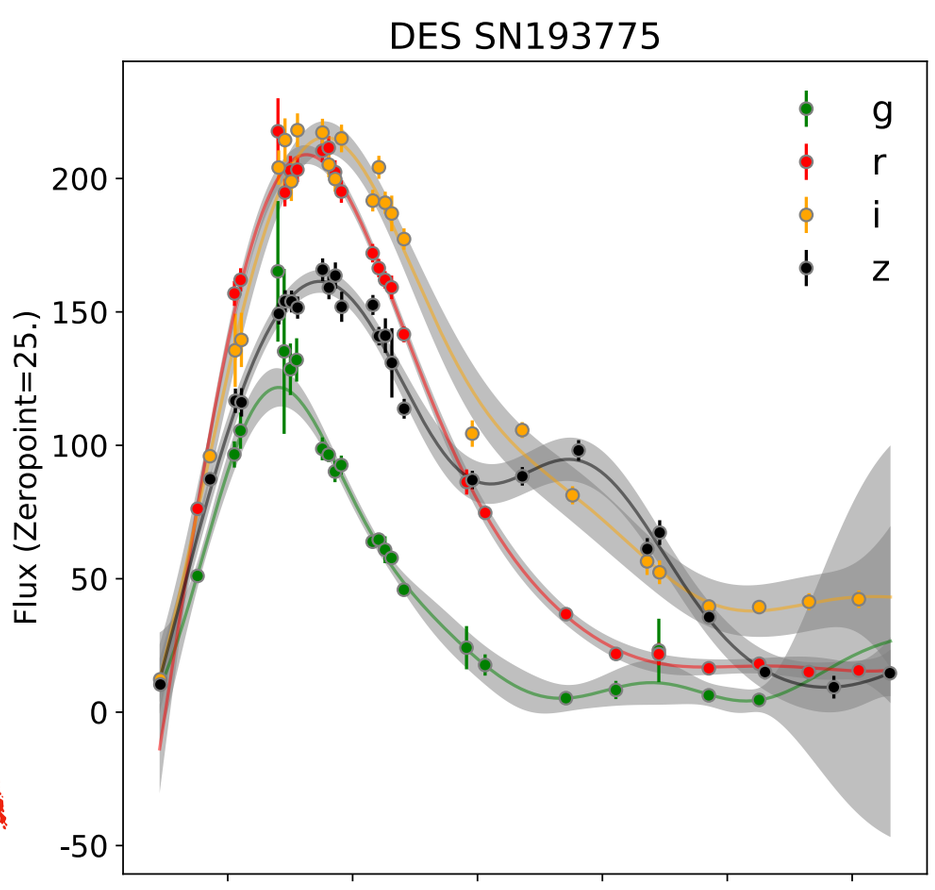
Point estimates + PCA + imbalanced learning are pretty good at getting most of the large classes right, even with sparse light curves (~10 observations).

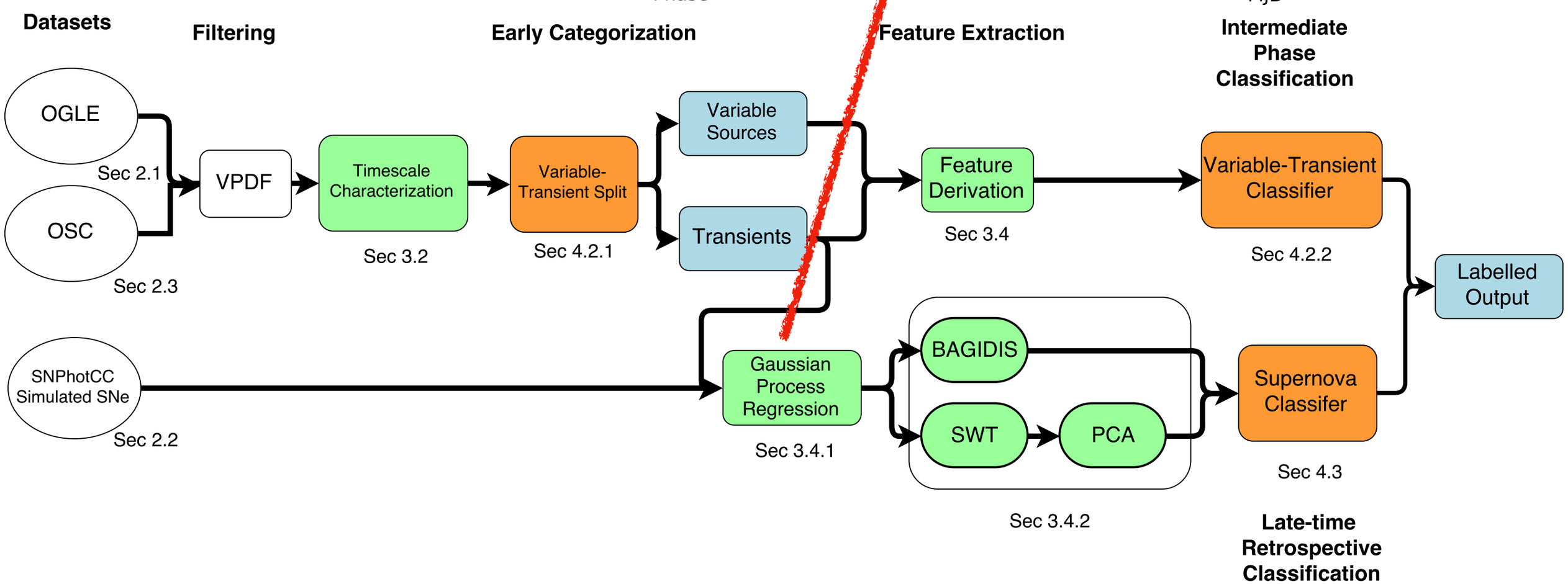
Probably useful for most science cases during LSST operations



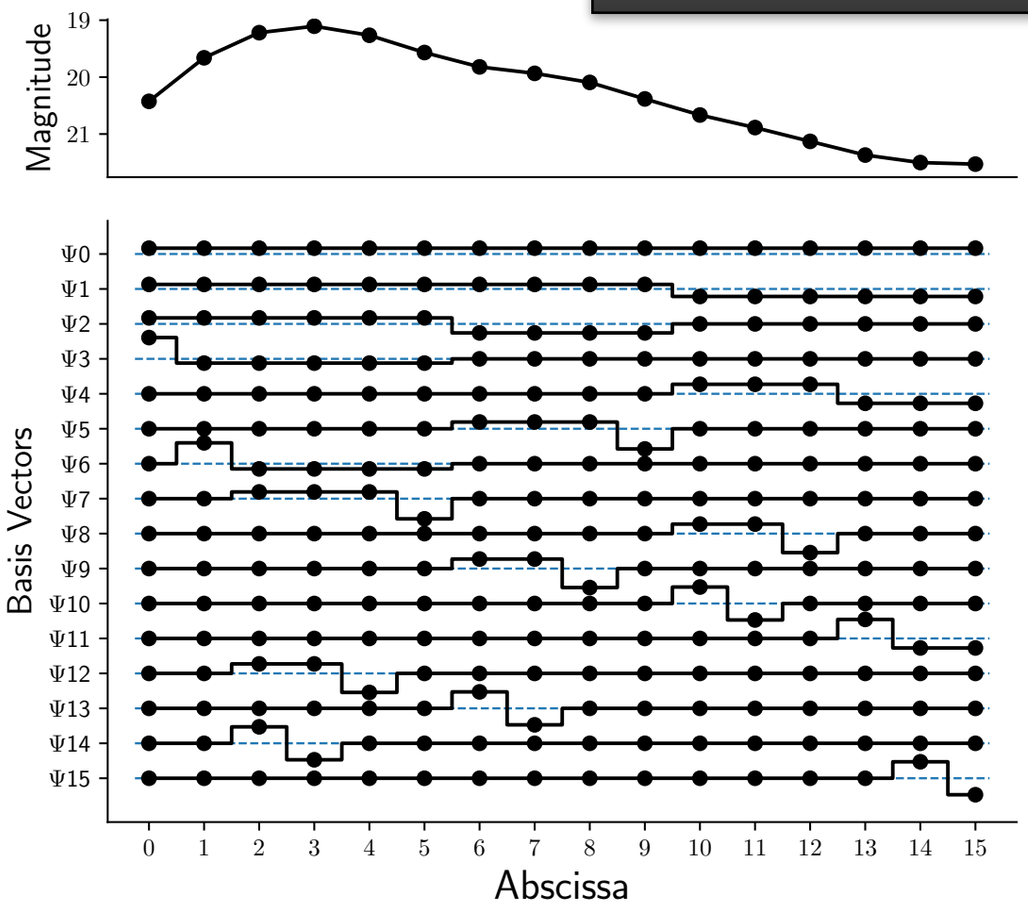
Use Gaussian Processes to interpolate and model time-series observations

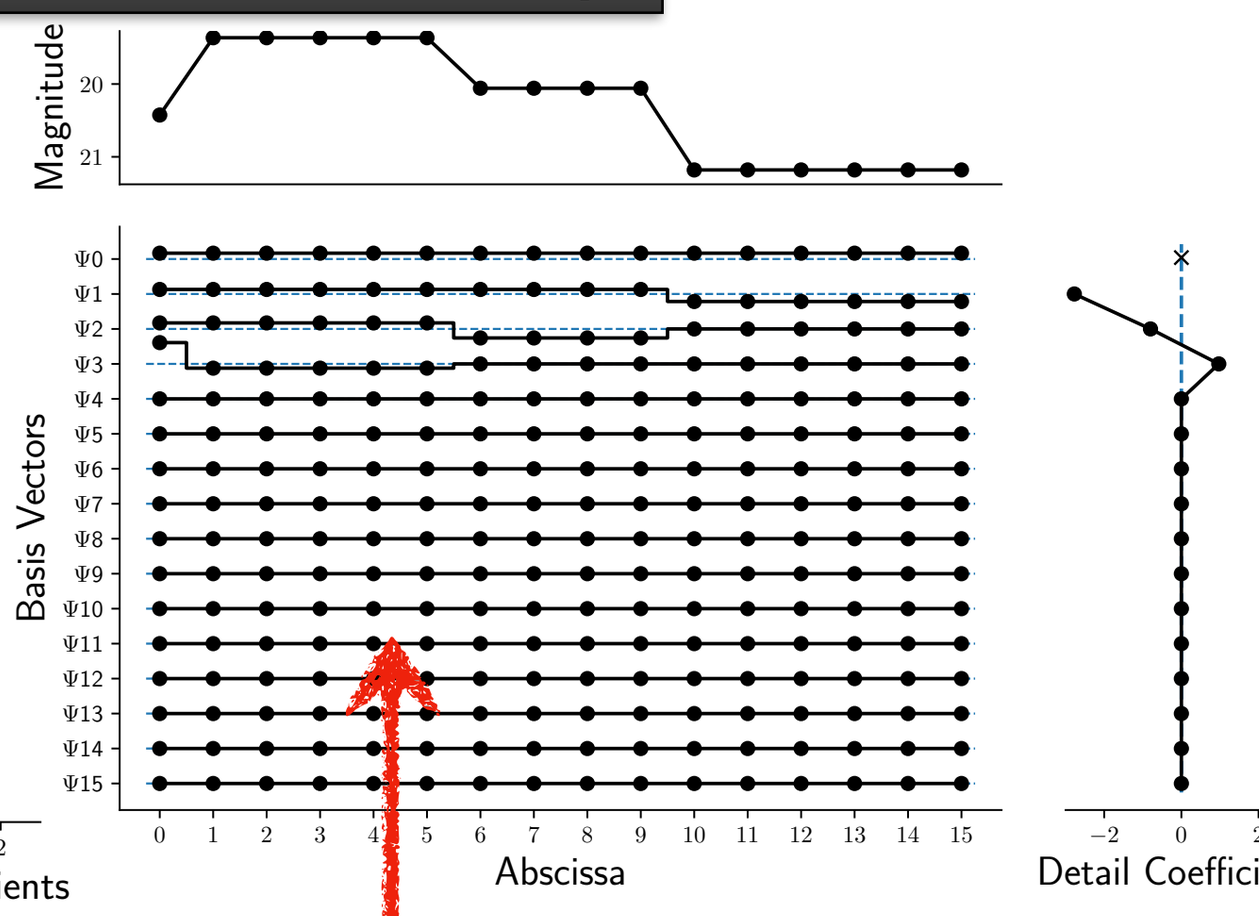


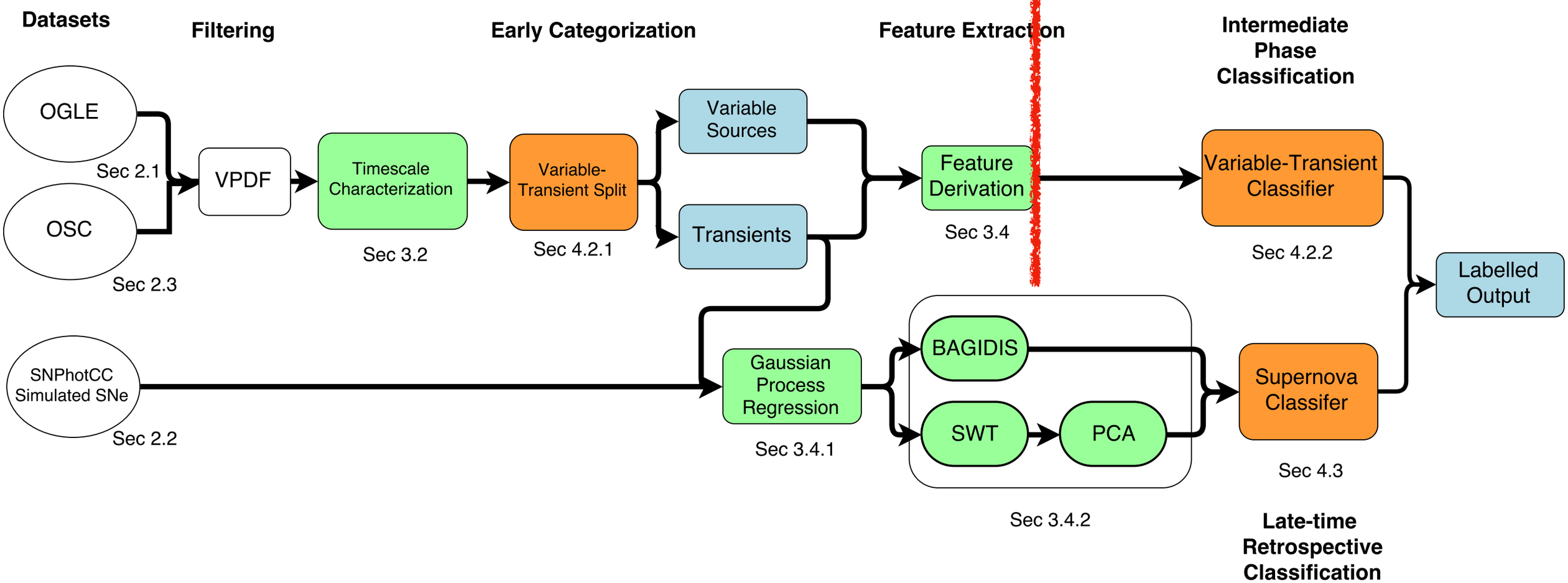


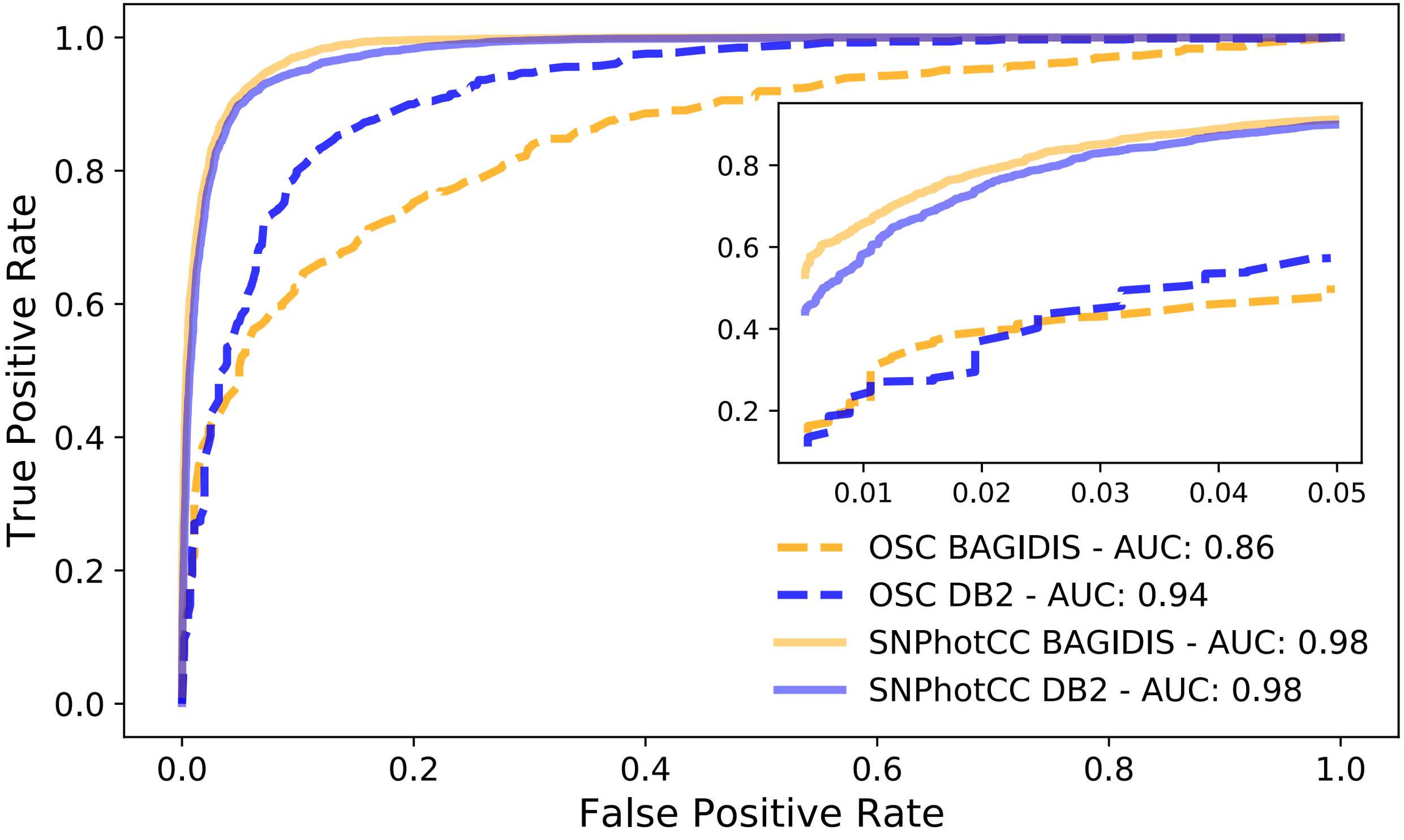


Use wavelet transformation to describe shape









ROC Curves

Very comparable to Lochner ’16 for SNPhotCC. Real heterogenous data is not like homogenous sims.

Take aways

- If you are visually inspecting survey data to find transients and variables, we can help.

- Narayan, Zaidi, Soraisam et al (2018, ApJS, submitted)

- Next: Apply to PS1 MDS light curves + contextual information, and ongoing Kepler K2 C16.

- Apply to PLAsTiCC simulated data, ZTF public alerts (~April)

- Postage stamps as features with CNNs, unsupervised learning techniques

- WE ARE HIRING!
  Associate or Assistant Scientist at NOAO
  (https://goo.gl/bceKu7)