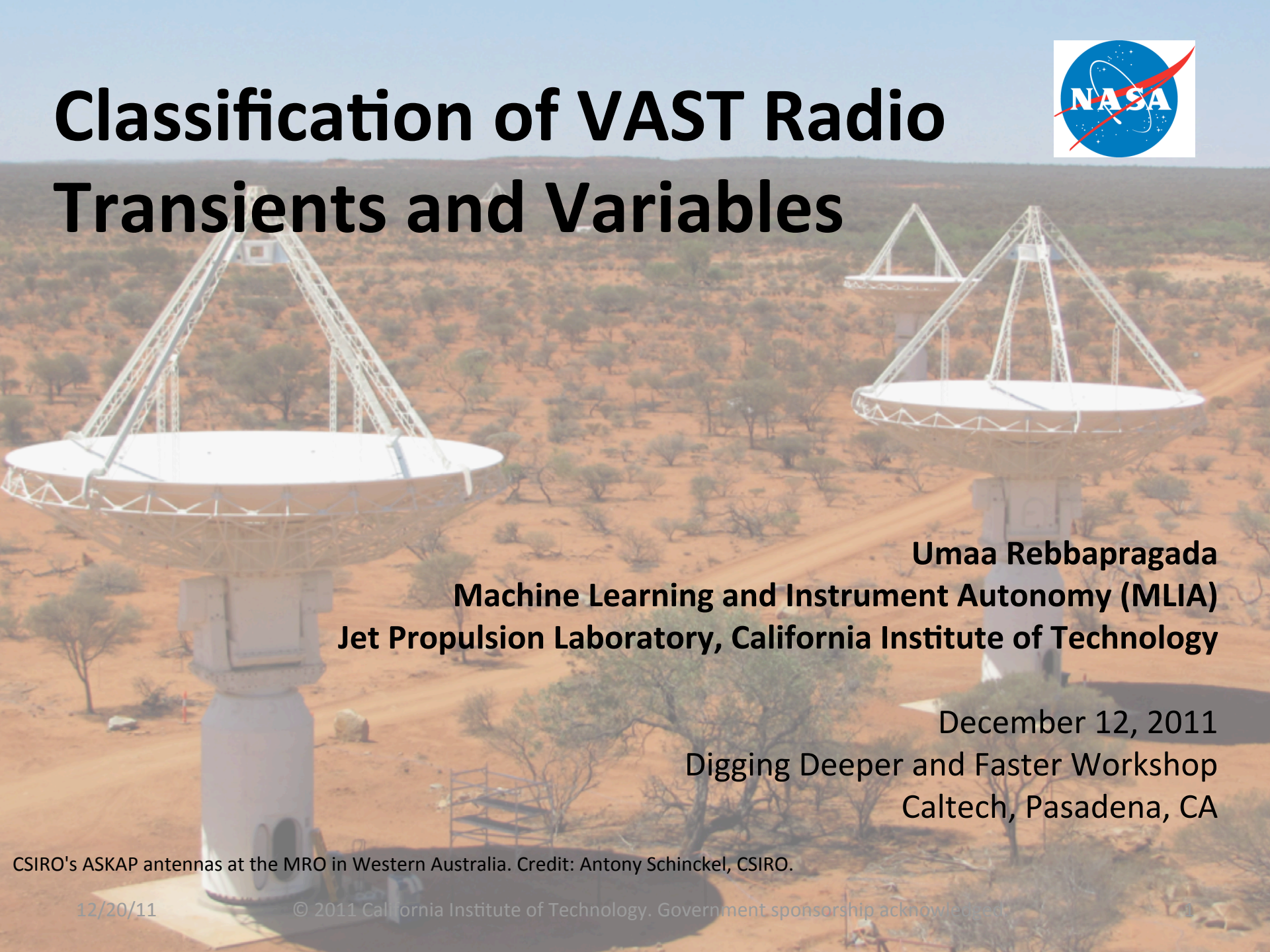
Classification of VAST Radio Transients and Variables



Umaa Rebbapragada
Machine Learning and Instrument Autonomy (MLIA)
Jet Propulsion Laboratory, California Institute of Technology

December 12, 2011
Digging Deeper and Faster Workshop
Caltech, Pasadena, CA

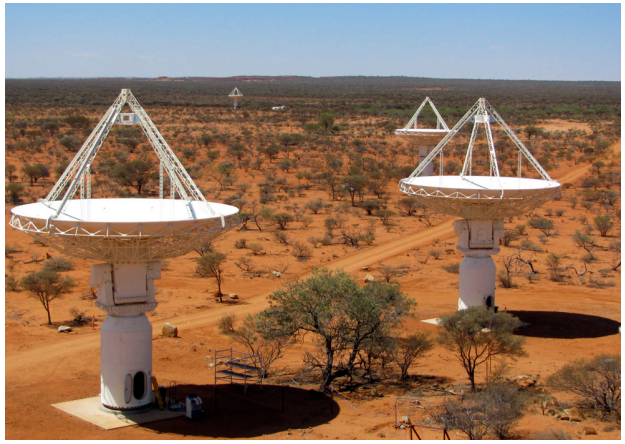
CSIRO's ASKAP antennas at the MRO in Western Australia. Credit: Antony Schinckel, CSIRO.

Acknowledgements

- Sydney Institute for Astrophysics, U. Sydney
  - Kitty Lo
  - Dr. Tara Murphy
- MLIA, JPL
  - Dr. Kiri Wagstaff
  - Dr. David Thompson
  - Colorado Reed (University of Iowa)

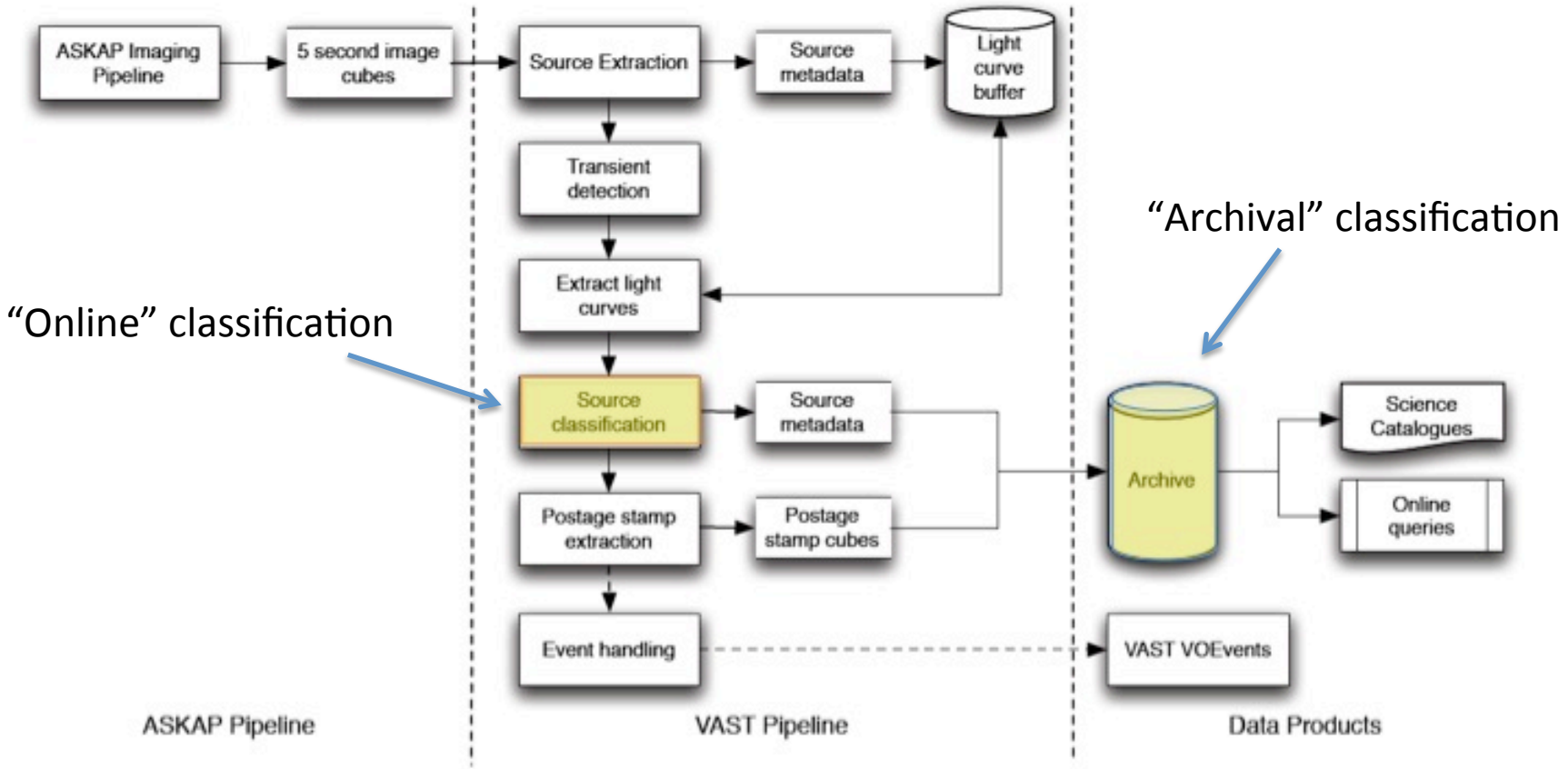
ASKAP and VAST

- **A**ustralian **SKA** **P**athfinder
- Observes radio sky in single day
- sub-mJy sensitivity
- 5 second cadence



- **V**ariables **a**nd **S**low **T**ransients
- Real-time data processing pipeline
- Objects of interest: SNe, Novae, IDVs, ESEs, etc.
- Potential to discover new objects and object classes
- ASKAP BETA online in 2012

VAST Pipeline



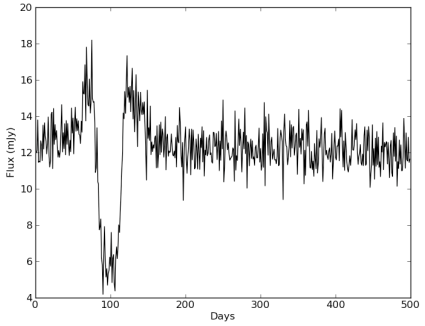
Study Goals

- Identify
  - **feature representations**
  - **learning algorithms**

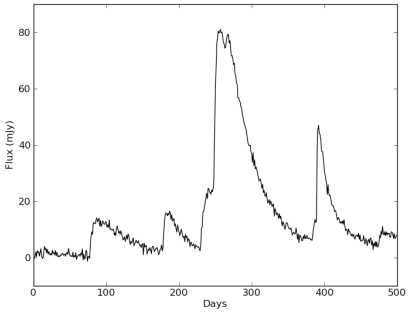
- Estimate classification performance per
  - **source type**
  - **observing strategy**

Simulated Radio Source Types

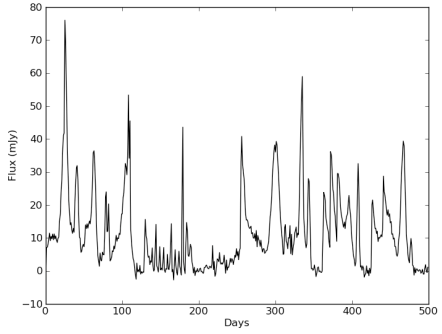
- 200 per source type, sampled 1x a day, SNR at 3, 5, 7, 10 σ, 400 days



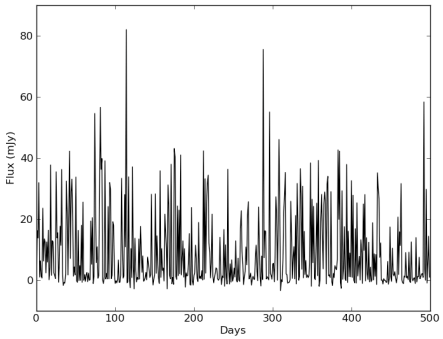
Extreme Scattering Event (ESE)



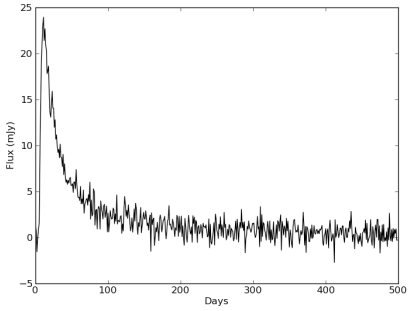
Xray Binary (XRB)



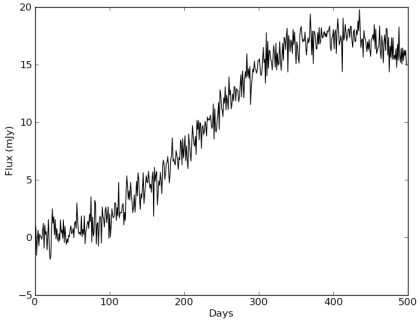
Flare Stare RSCVn



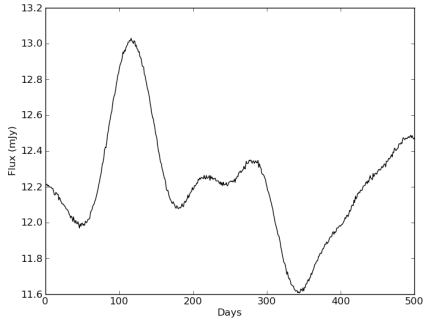
Flare Star dMe



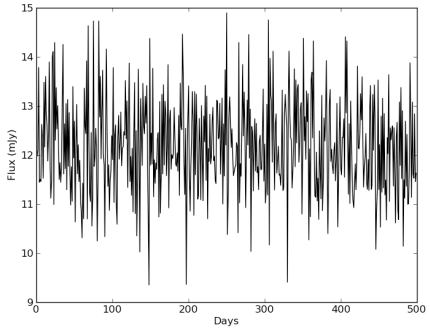
Supernova (SNe)



Nova



Intraday Variable (IDV)



Background Source (BG)

Observational Strategies

	VAST Wide	VAST Galactic Plain	VAST Deep	patches	monthly	log
Sampling	Daily	Weekly, plus a pattern of 4, 2, 1, 2, 4 days every 3 months	Days 1, 2, 3, 4, 17, 21	Random 3 conse-cutive days per month	Every 30 days	Days 1, 2, 4, 8, etc.
RMS	0.5mJy	**0.1mJy**	**0.05mJy**	0.5mJy	0.5mJy	0.5mJy

Classifiers and Features

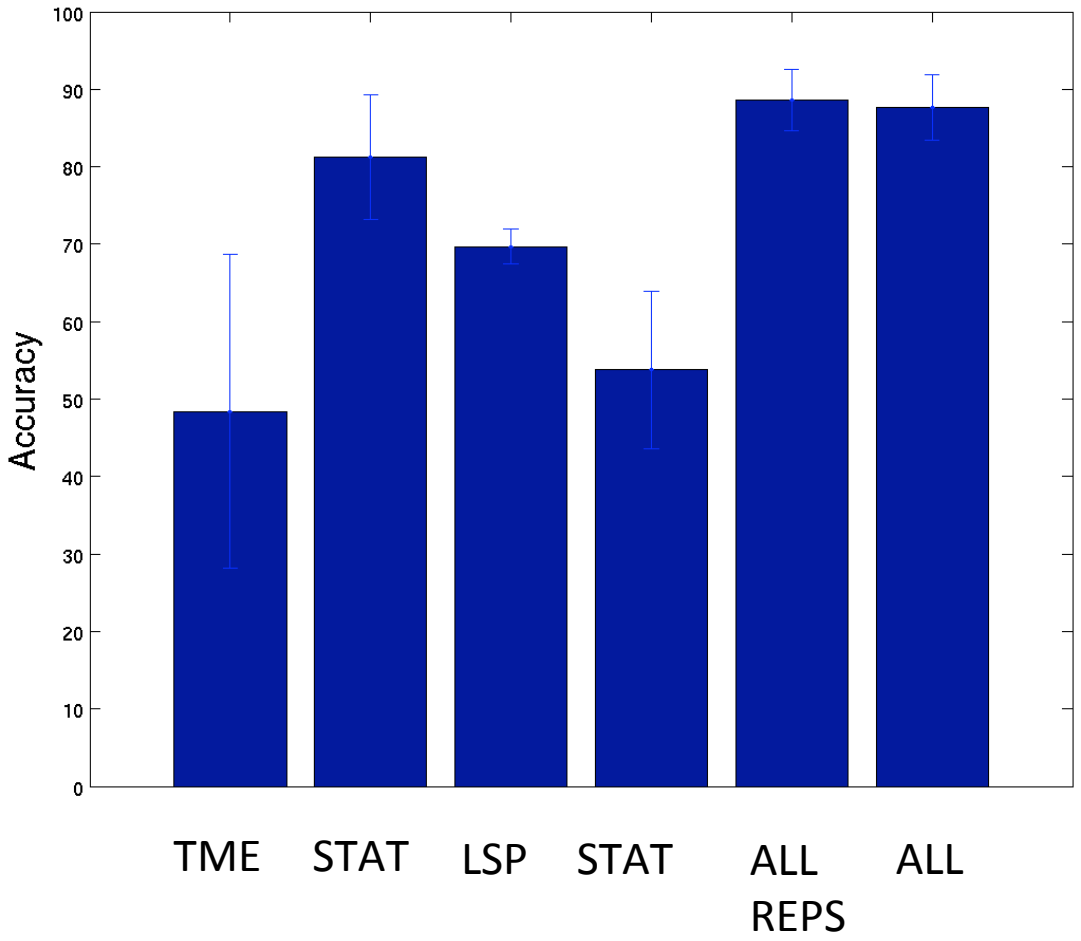
- **Support Vector Machine**
- **Decision Tree**
- **Random Forest**
- Naïve Bayes
- Logistic Regression

- Frequency Domain
  - Lomb-Scargle Periodogram (**LSP**)
  - Haar Wavelets (**WLET**)
- Statistical Representations (**STAT**)
  - Moment statistics
  - Non-periodic features from [1]
- Time observations (**TME**)
- LSP+WLET+STAT (**all-reps**)
- TME + LSP+WLET+STAT (**all**)

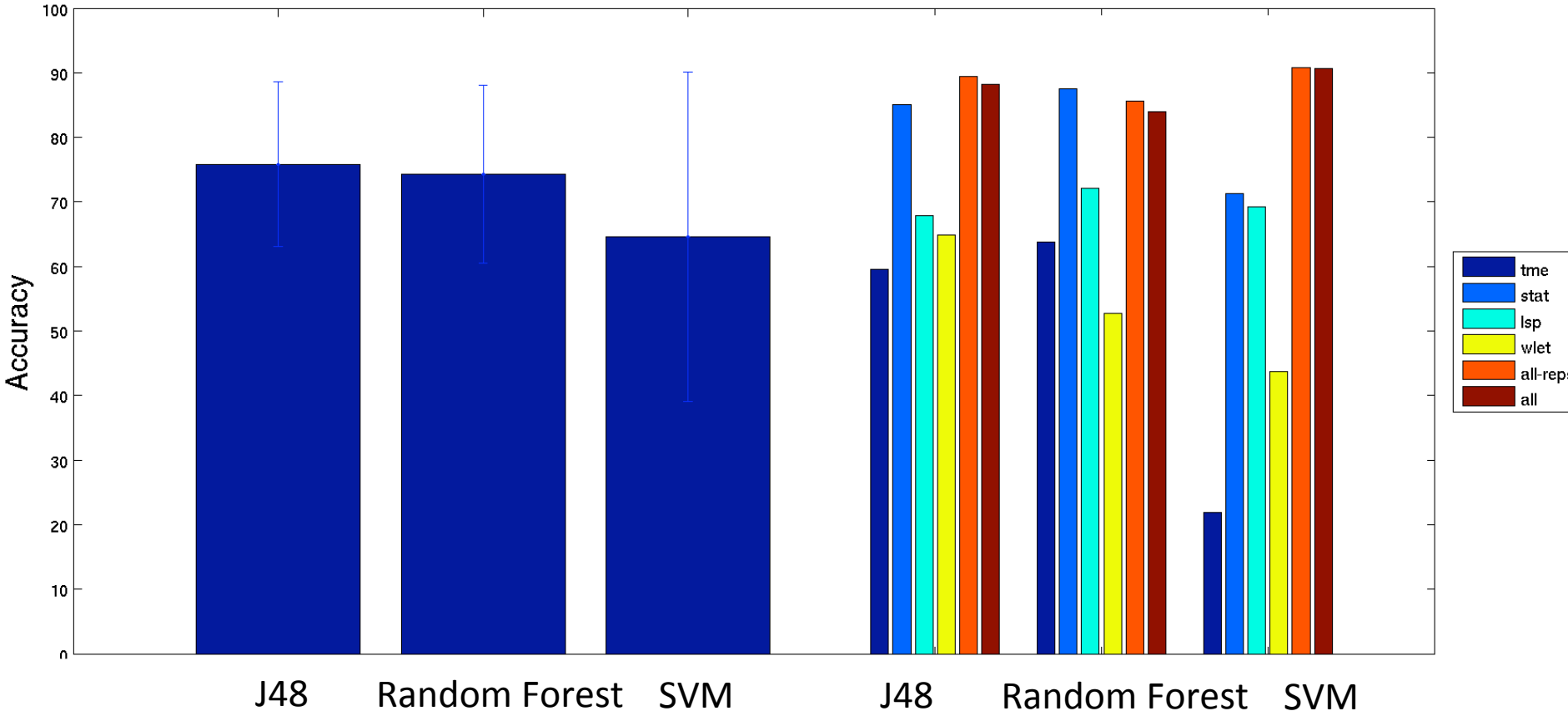
[1] Richards et al. (2011) On machine- learned classification of variable stars with sparse and noisy time-series data. arXiv 1101.1959

Archival: Accuracy by Feature

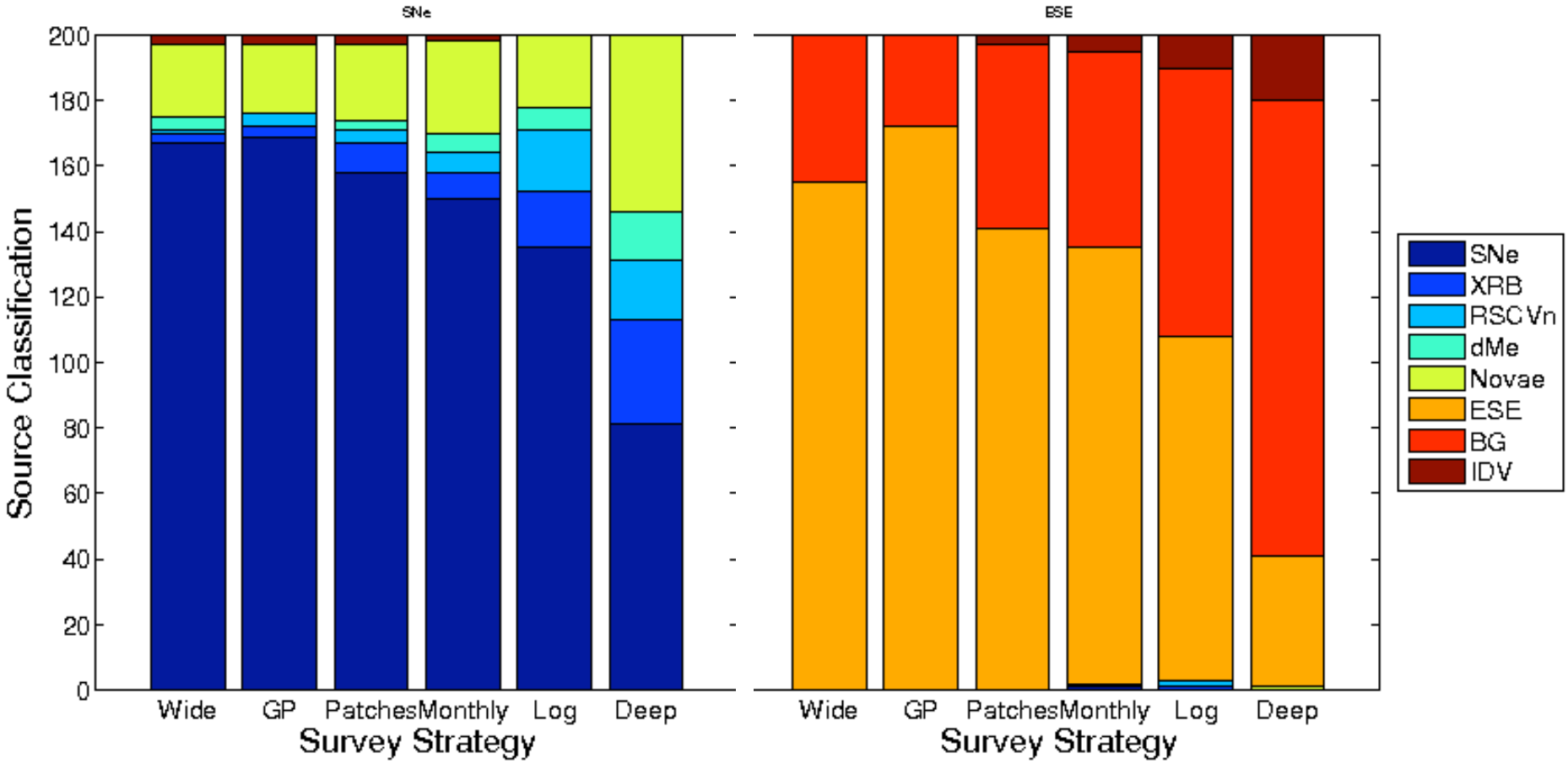
- Y-axis measures accuracy
- VAST Wide
- Averaged over all SNR, classifiers
- 10-fold CV
- **Concatenating feature representations performs best**



Archival: Accuracy by Classifier



Archival: Class Confusions



SNe **ESE**

Archival: Conclusions

- VAST GP achieved same classification performance with ¼ observations
- Best performers: SVM using STAT+LSP+WLET
- Two major confusion groups:
  - BG, IDVs, ESE
  - Sne, Novae, Flare Stars (RSCVn and dME), XRBs

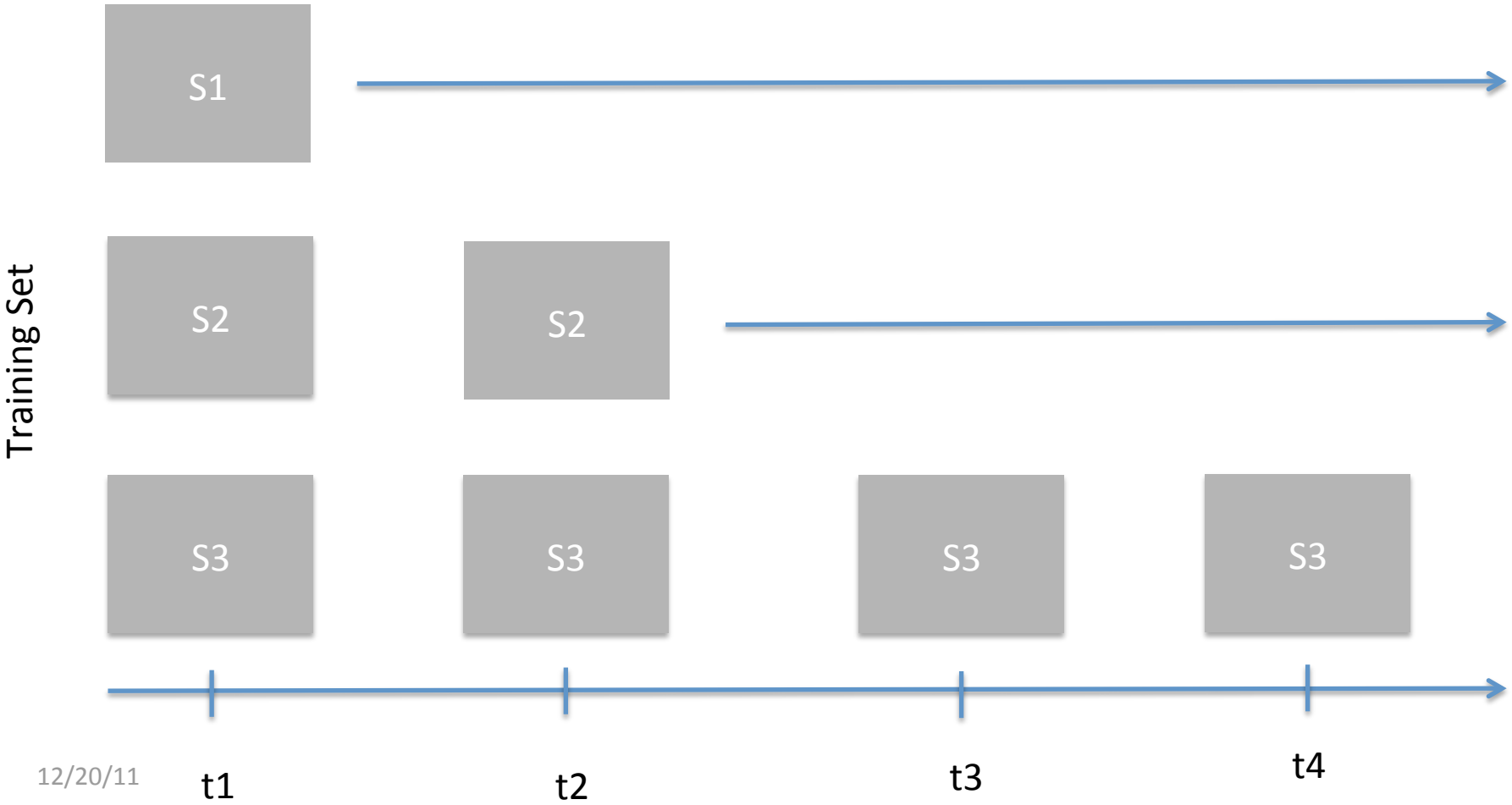
Online: Methodology

1. Train on full archival knowledge, test on partially-observed

2. Train and test on partial observations

3. Delay-Sensitive Ensemble Classification (DSEC)

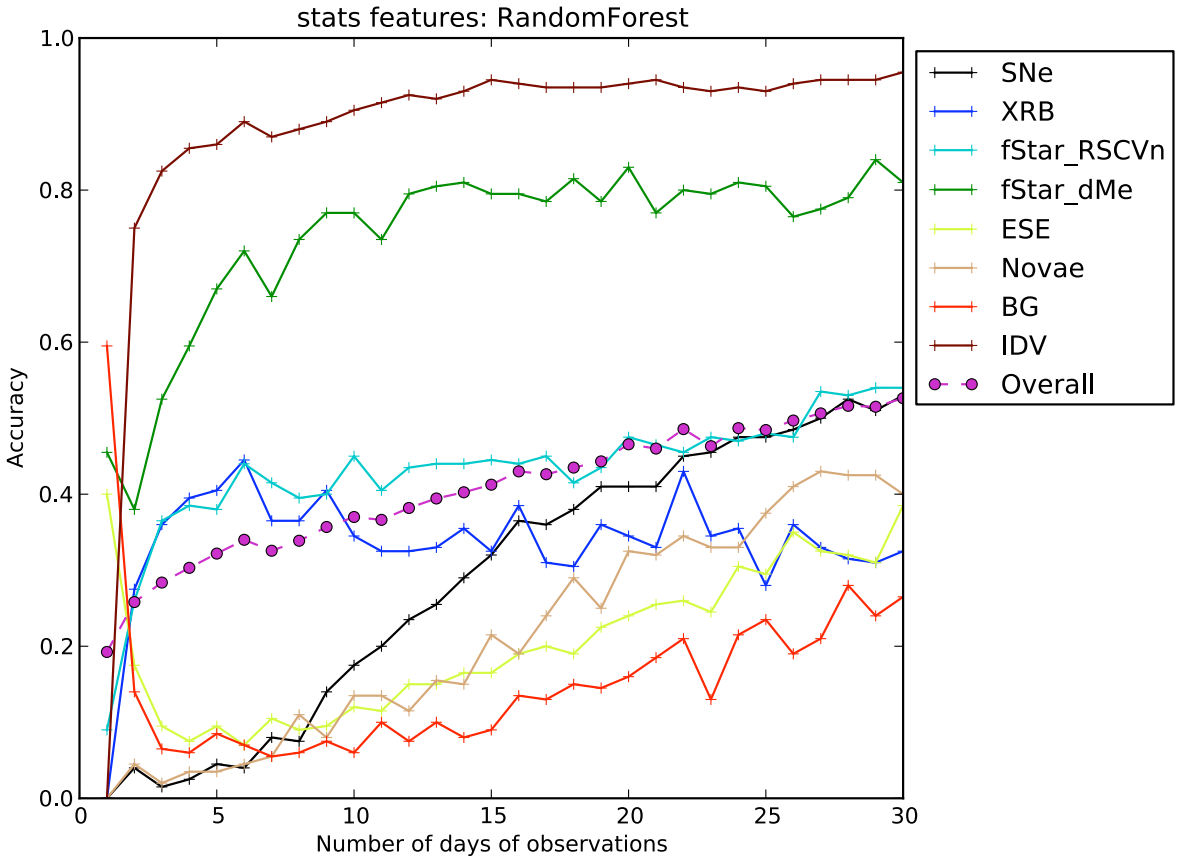
DSEC



Online: Methodology

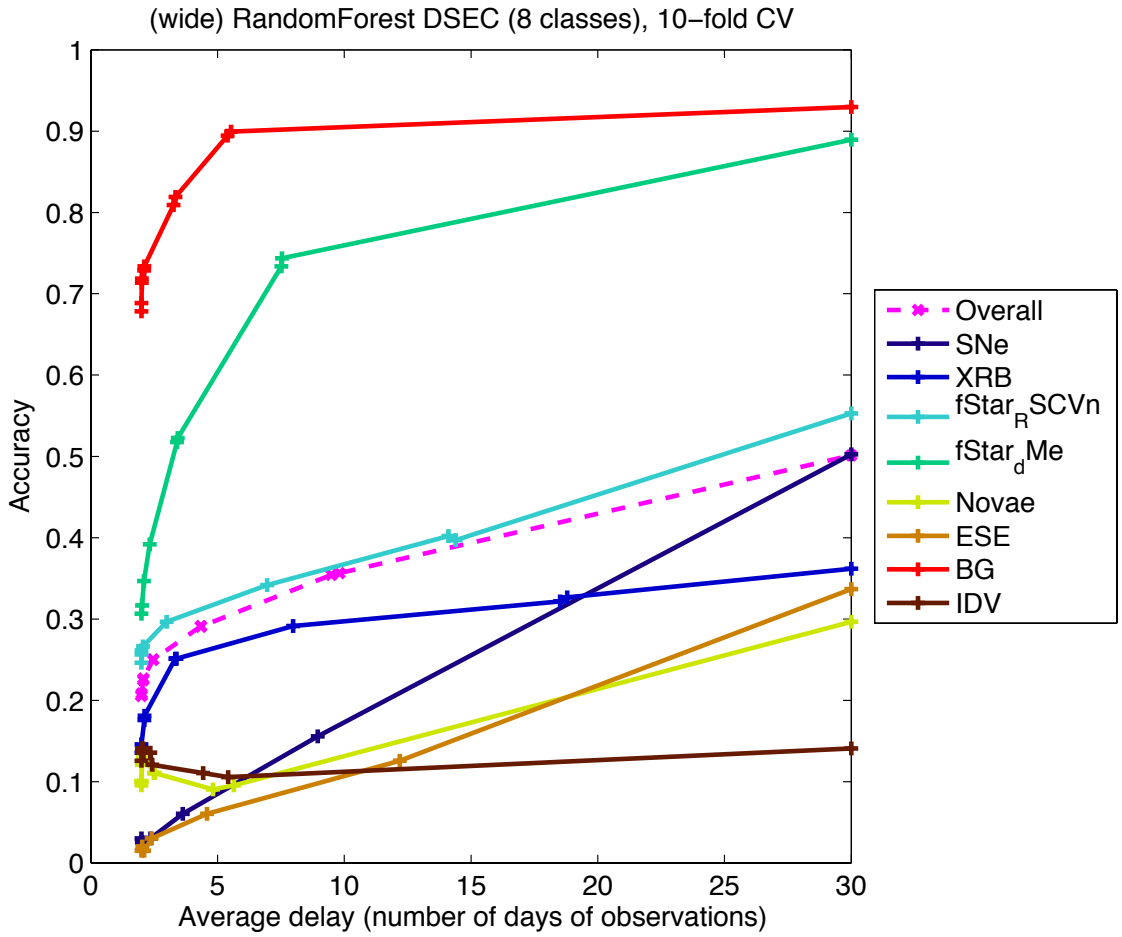
- Up to 30 days VAST Wide observations
- Features: time, stat
- Classifiers: J48, Random Forest

Online: Results (Method 2)



- Average 30% after 5 observations
- BGs largely confused
- IDVs classified well
- Decision tree, RF results similar

Online: Results (DSEC)



- Similar average performance, ordering
- Ensemble complexity justified?

Future Work

- VAST Memo
- Improve results, focus on additional feature representations
- Benchmark methods on optical data sets
- Integration into VAST data processing pipeline