Interference-resistant real time adaptive detection

David R. Thompson
David.r.thompson@jpl.nasa.gov

With Kiri L. Wagstaff, Walid A. Majid and Colorado J. Reed

Jet Propulsion Laboratory,
California Institute of Technology
December 2011

Copyright 2011 California Institute of Technology. All Rights Reserved. U.S. Government Support Acknowledged.

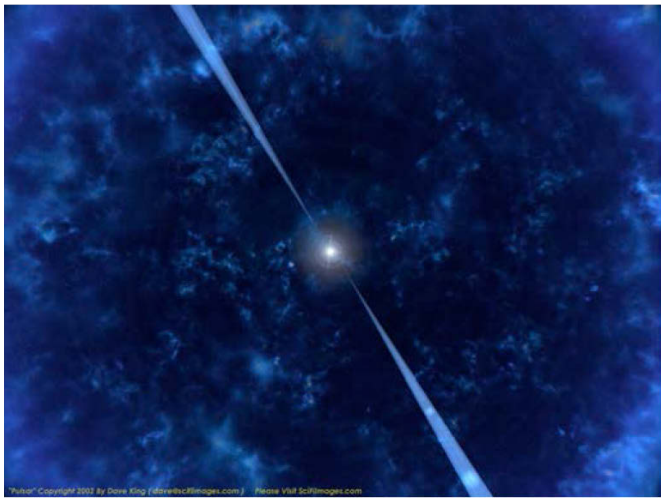


Image courtesy Square Kilometre Array

General formulation

Time series $X = \{x_t: 1 \leq t \leq T,\ x_t \in R^N\}$

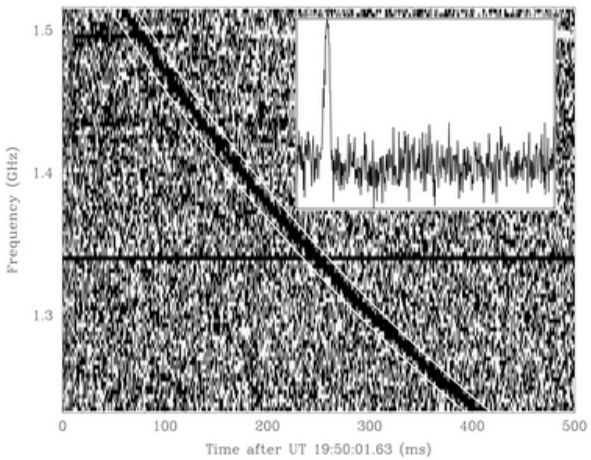
- Radio filterbank time series
- Optical light curves
- Petabyte data streams generally

Seek fast transients

- Brief, energetic pulses
- X-ray bursts, RRATs, and the unknown

Low false positive rate, online performance is vital

- Human effort required to review candidates
- Avoid overflowing data buffer



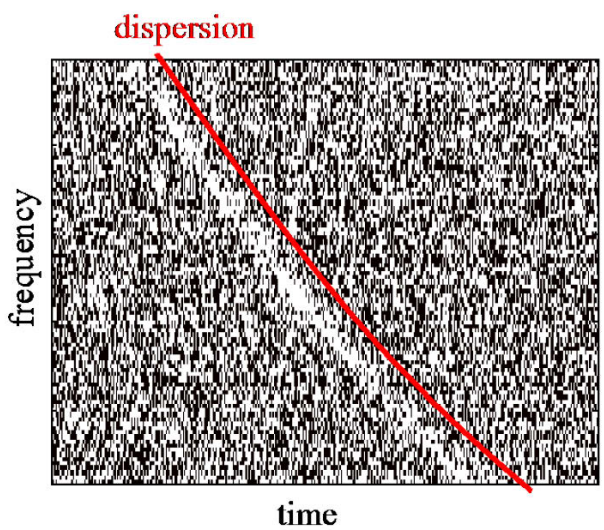
Images courtesy Lorimer 2007, Square Kilometre Array Consortium

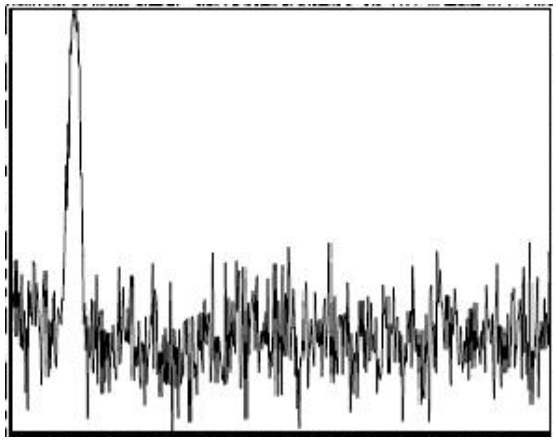
Case 1: we know the target signal

Classical Neyman Pearson detection:

$$f(x) = \frac{L(x_t\,;\,H_1)}{L(x_t\,;\,H_0)} > \tau$$

Implemented with matched filter, with threshold based on false positive tolerance





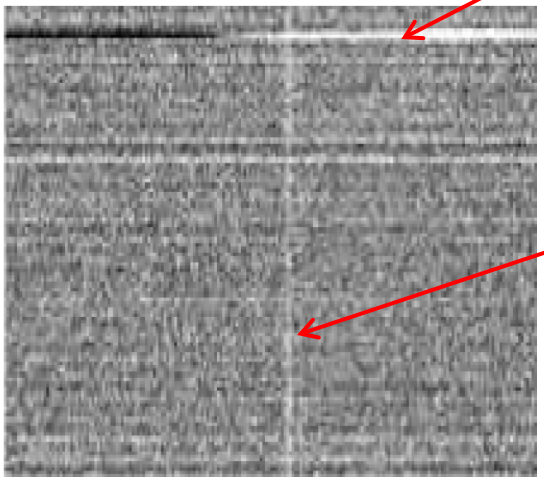
Challenge: noise and interference

$$f(x) = \frac{L(x_t\,;\,H_1)}{L(x_t\,;\,H_0)} > \tau$$

Hard to characterize

Generally handled with *ad hoc* rules, and limits the *practical* capacity of any system

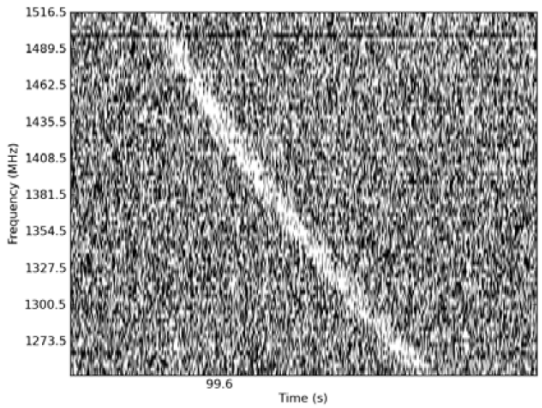
Parkes telescope multibeam receiver (Courtesy CSIRO)



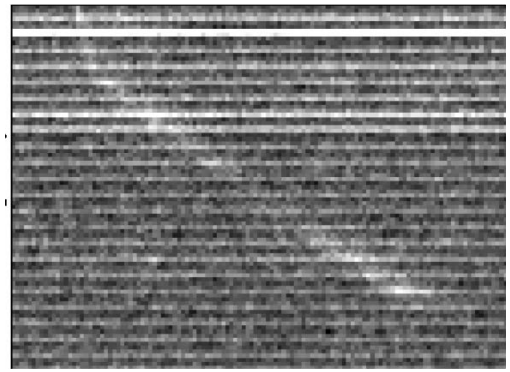
Radio image courtesy CSIRO / Parkes Pulsar Survey / S. B. Spolaor 2011

Case 2: Anomaly detection

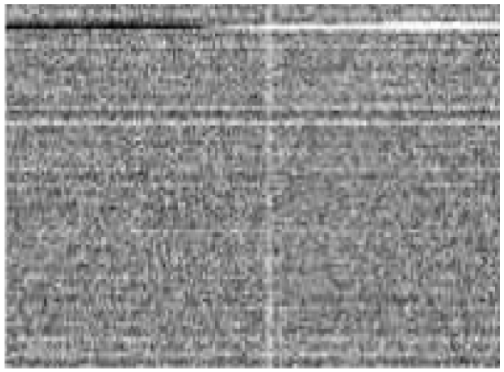
Pulsar pulse



"Peryton"



Radio Freq. Interference



$$f(x) = \frac{1}{L(x_t\,;\,H_0)} > \tau$$

(crude simplification)

Related methods: One-class SVM [Schölkopf et al., 2001], Semi-supervised anomaly detection [Blanchard et al., 2010]

Interference is even more problematic!!

Review: our problem

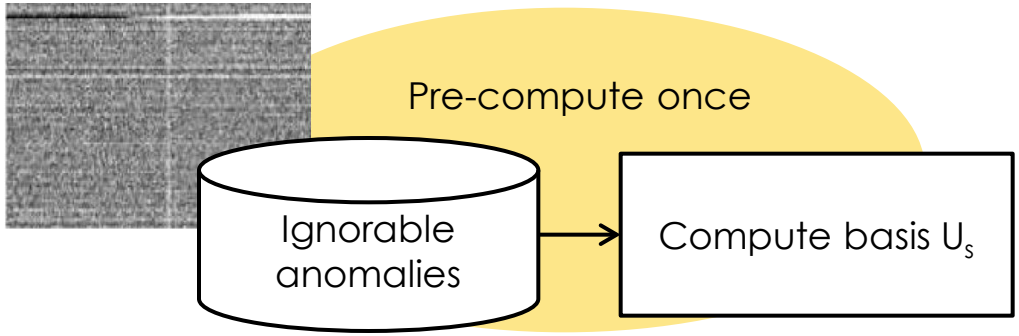
- Weak signals
- High data rates
- Structured background interference (false alarms)
- Changing background interference (difficult to characterize in advance)
- Don't always know what we're looking for
- Must run online in real time

Semi-Supervised Eigenbasis Novelty Detection (SSEND)

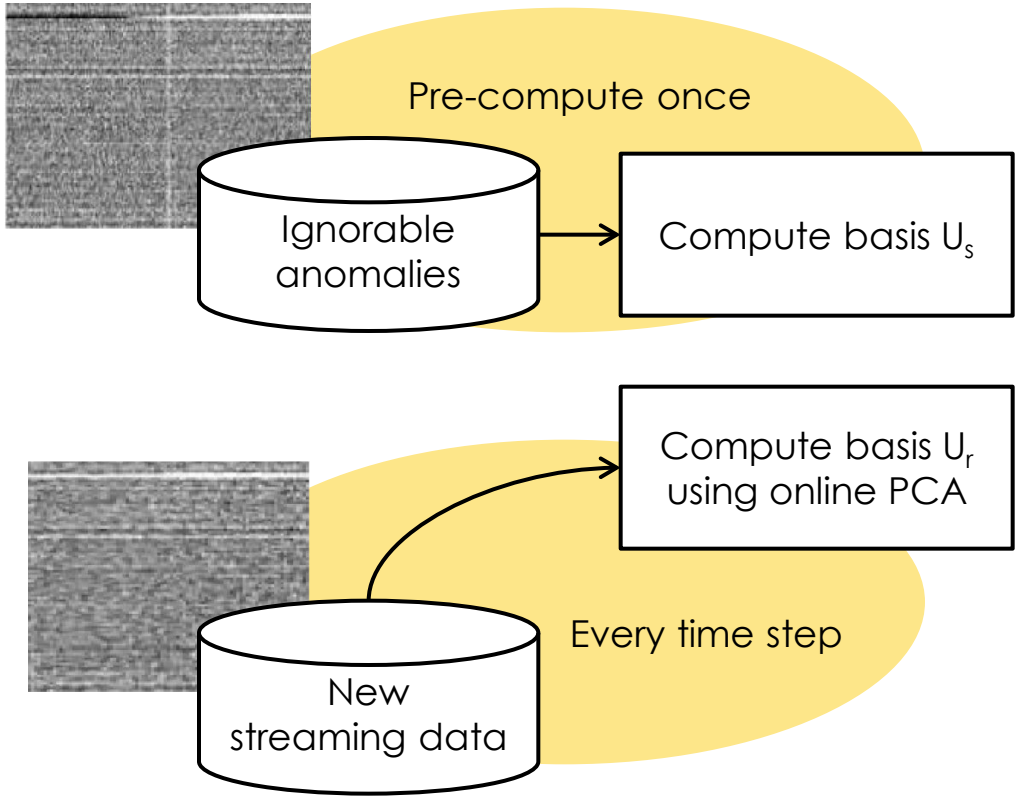
- Online background estimation based on incoming data (assumption – background changes slowly)
- Semi-supervised train with known "ignorable anomalies"
- Use Reconstruction Error as novelty score

$$f(x_i) = ||x_i - \hat{x}_i|| = ||x_i - AA^T x_i||_2$$

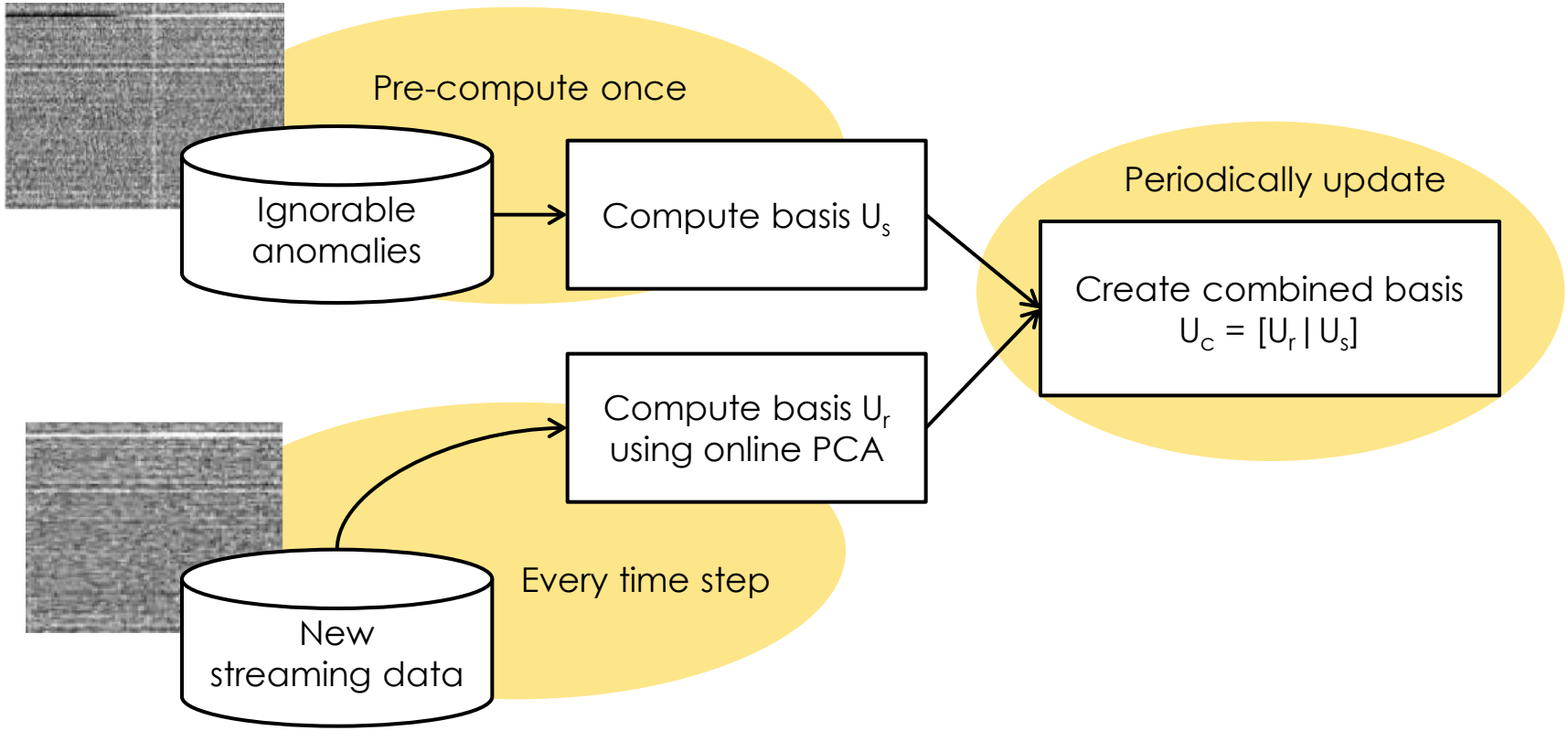
SSEND Concept: online, trainable novelty detection



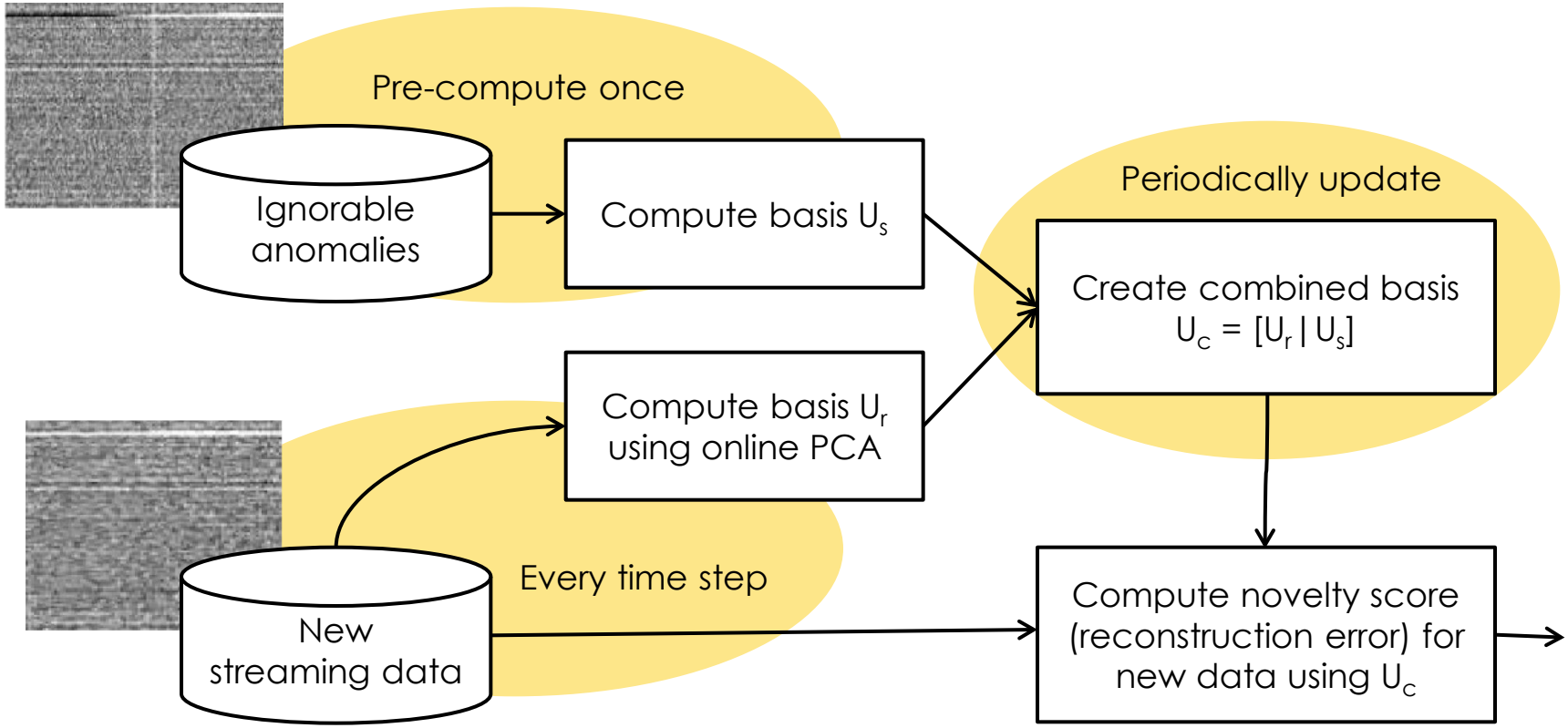
SSEND Concept: online, trainable novelty detection



SSEND Concept: online, trainable novelty detection



SSEND Concept: online, trainable novelty detection

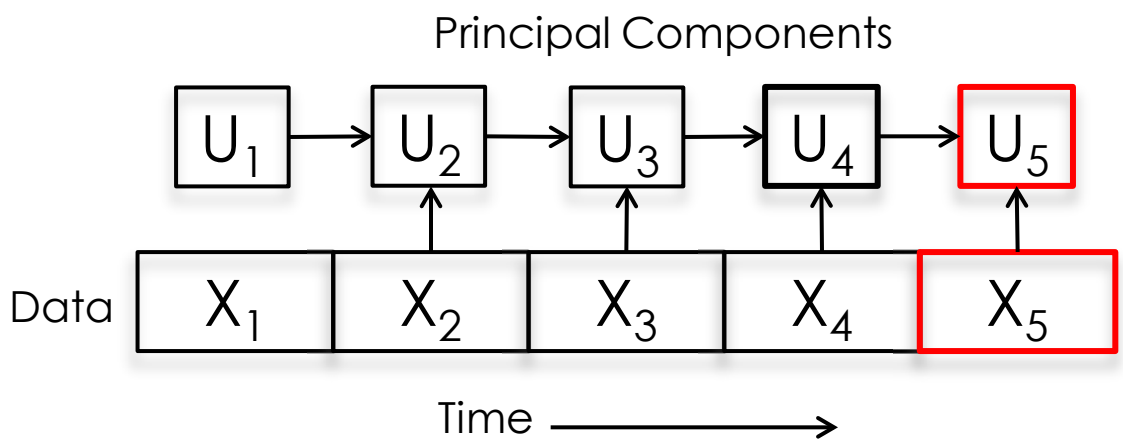


Online Updates

Compute principal components using Online PCA [Ross et al., 2004]

$$X = U\Sigma V^T$$

1. Have old decomposition $U_r \Sigma_r V_r^T$
2. We get a new data point X_q
3. Compute combined $U_p \Sigma_p V_p^T$ without X_p



Semi-supervision

Compute principal components from training data (ignorable anomalies):

$$X_s = U_s \Sigma_s V_s^T$$

Concatenate bases and orthogonalize with QR decomposition:

$$U_c = [U_r | U_s]$$

Retain first few bases in A

Compute reconstruction error:

$$f(x_i) = ||x_i - \hat{x}_i|| = ||x_i - AA^T x_i||_2$$

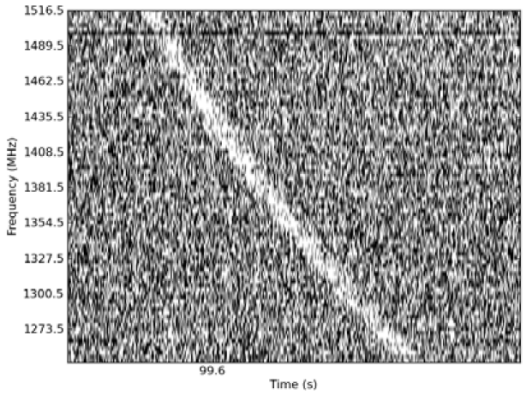
Radio Astronomy Data

Parkes Multibeam Survey [Edwards et al., 2001]

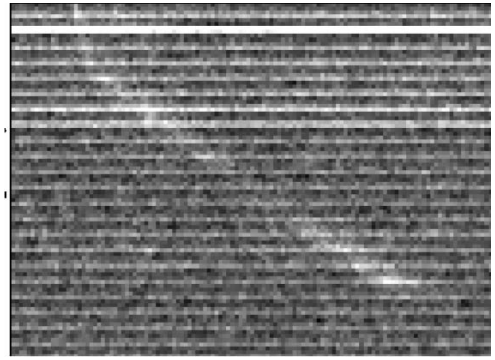
- 1.4 GHz, 125 μs sample time, 96 channels
- Goal: detect pulsars
- … but other anomalies also lurk within

Parkes telescope multibeam receiver

Pulsar pulse

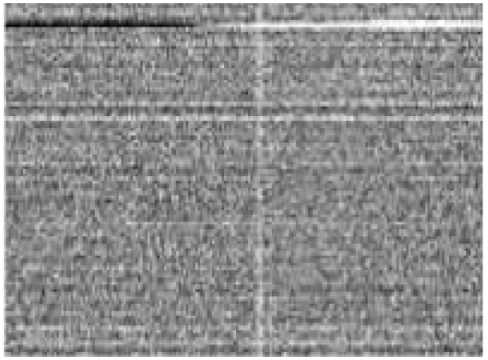


"Peryton"



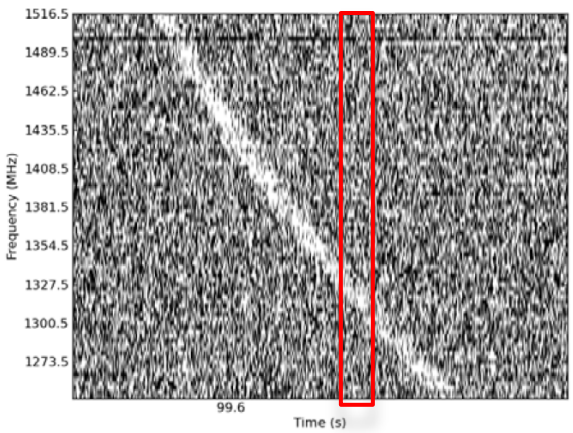
Time

Radio Freq. Interference



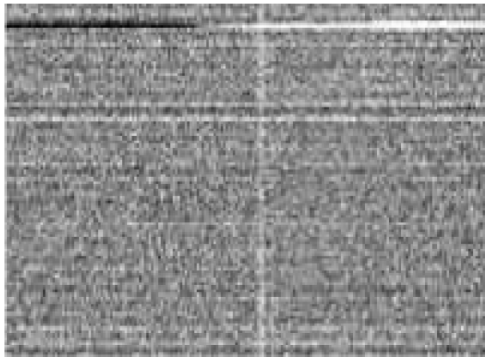
Experiments

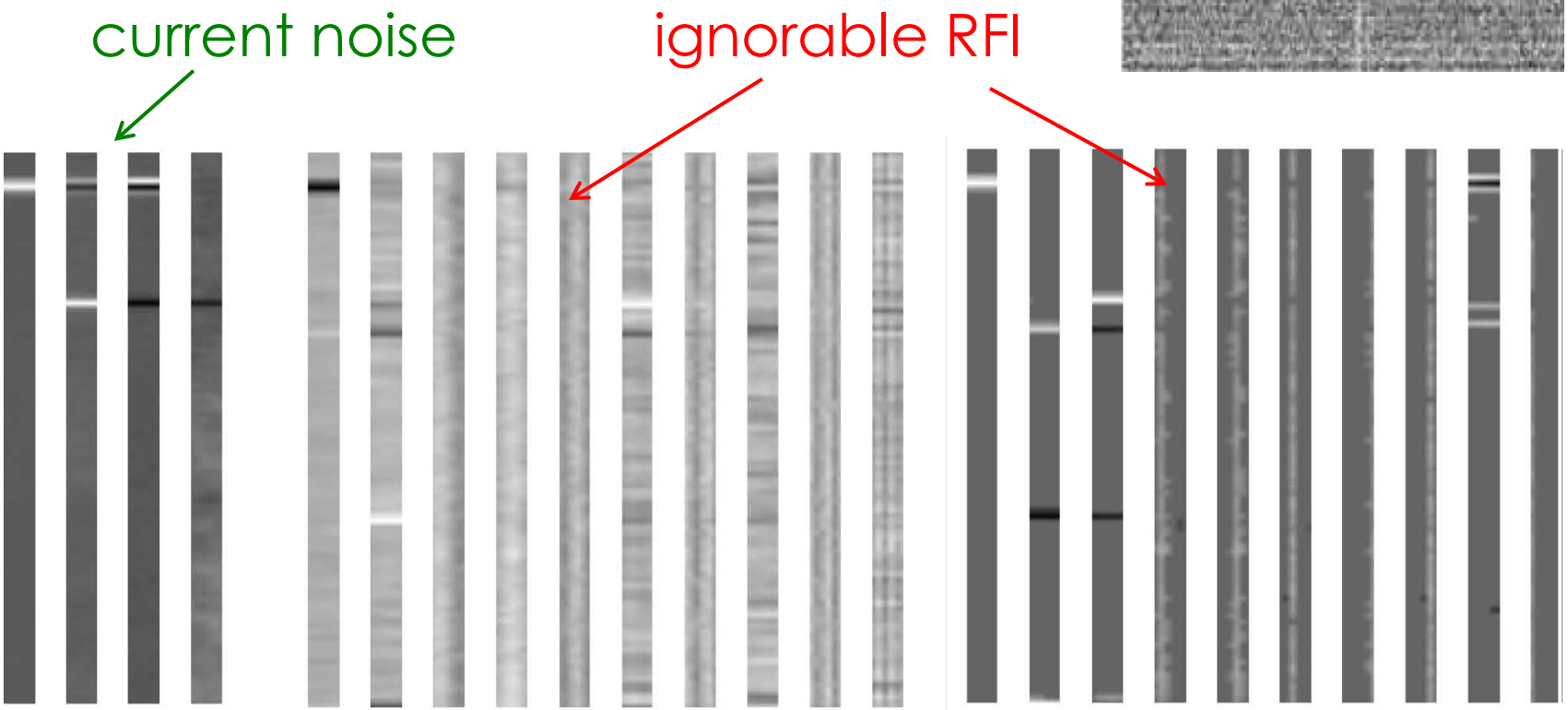
- Subsample and segment data every 15 ms
  - 576 dimensions (6 time steps x 96 channels)



- Construct U_r online; retain 4 bases
- Train U_s using 30 manually selected RFI
  - Collapse to 10 bases

Eigensignals



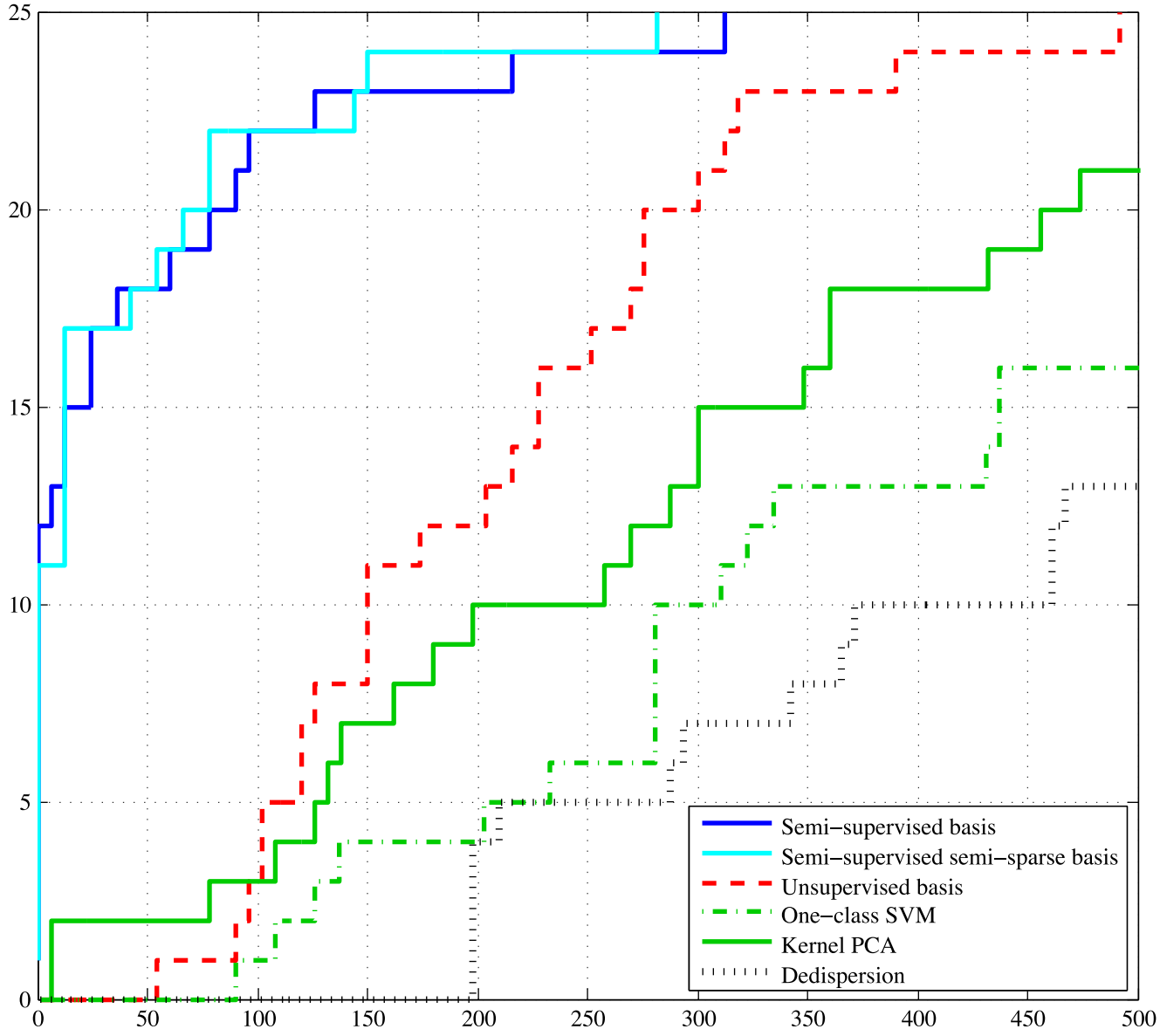


Adaptive
(Online PCA)

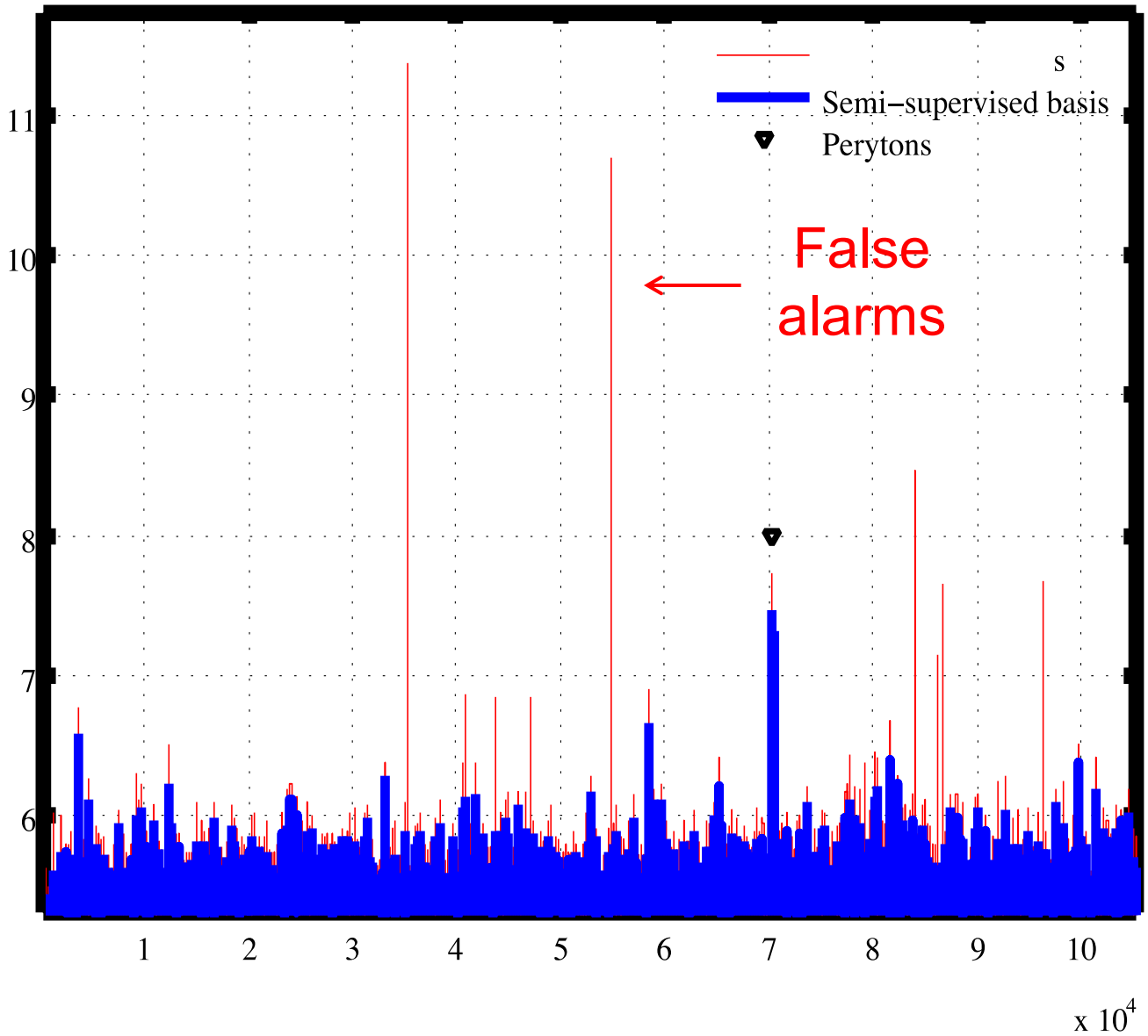
Supervised
(PCA)

Supervised
(Sparse PCA)

Initial Results

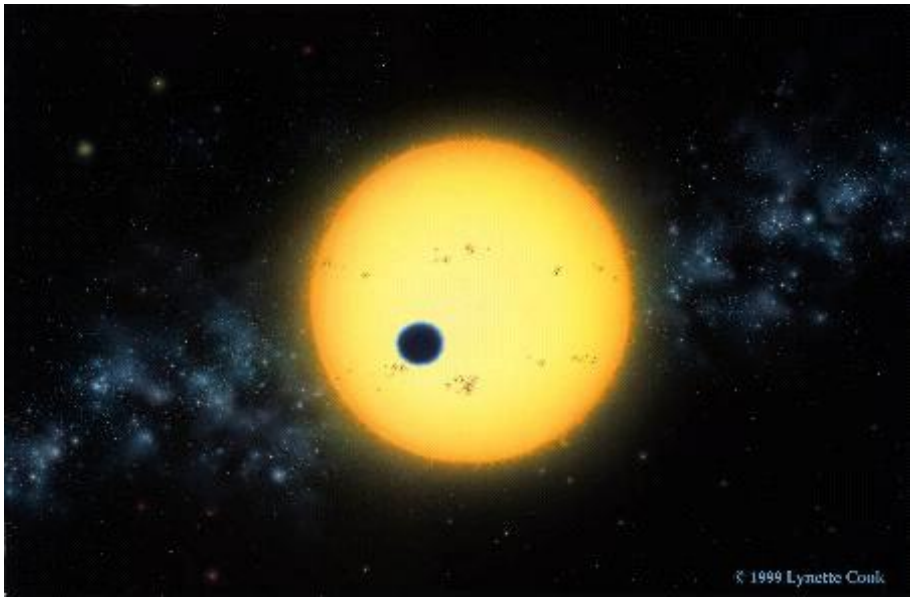


Novelty Scores



SSEND for "Iterative Discovery" with user interaction

- Kepler: the hunt for exoplanets
  - Planet transits are rare, anomalous - but many other anomalies exist



Kepler Mission

Goal: characterize exoplanets in/near habitable zones and their host stars

- 0.95-m telescope, 105 sq deg, >145,000 stars
- To date: 1781 candidates, > 1000 stars

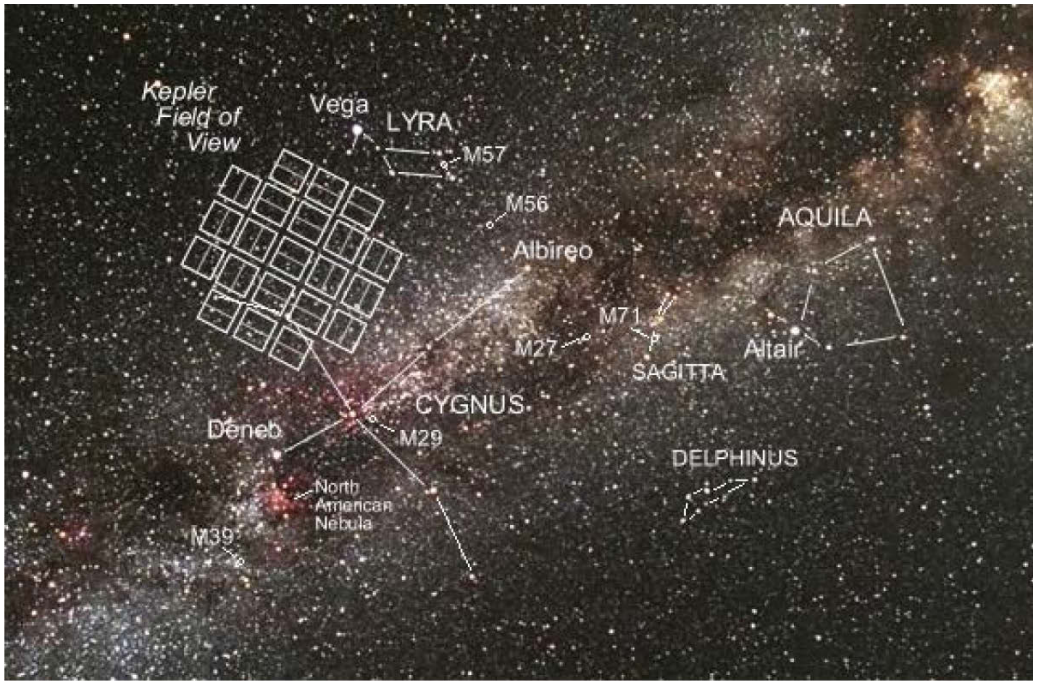


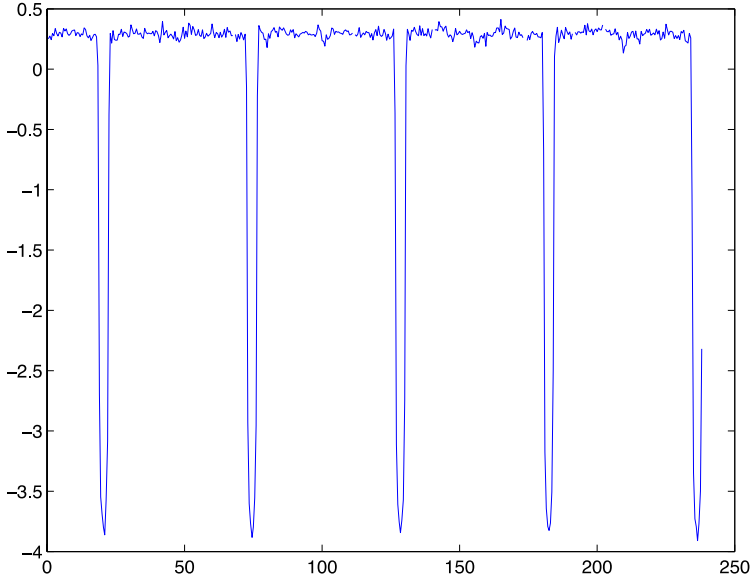
Image courtesy Kepler Mission / NASA

Kepler Time series Data

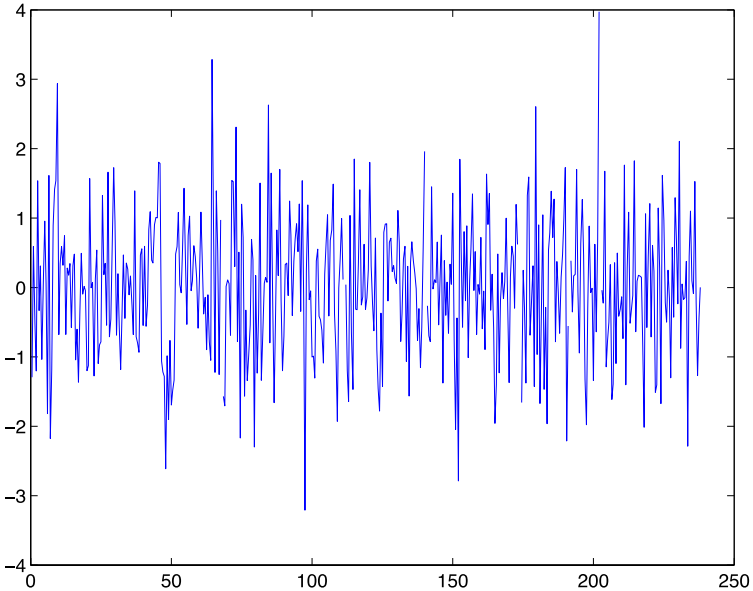
Three quarters released so far

Hunting for anomalous, rare transits

10666592: 1 planet, T=2.2 days

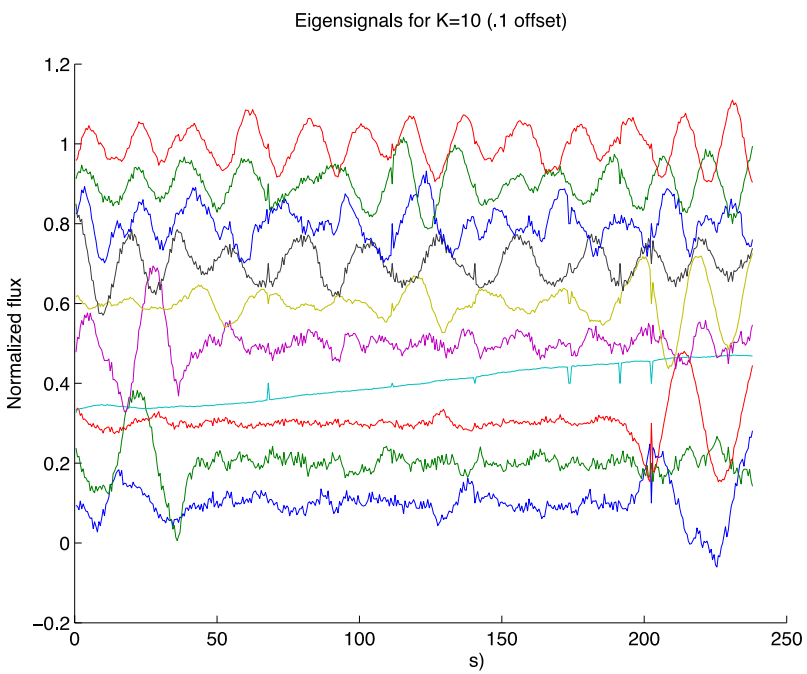


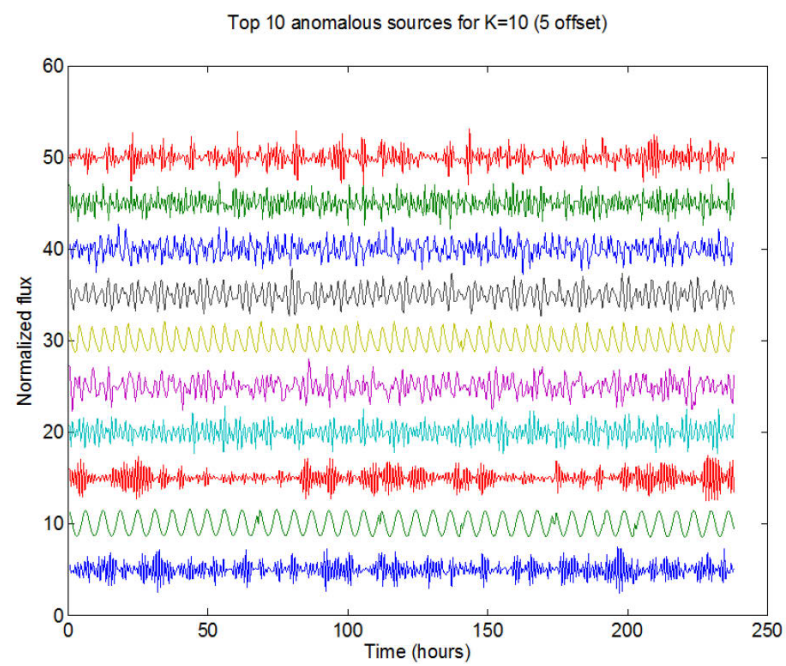
11512246: 3 planets; T=5,7,10 days



Kepler Q0 Data

52,496 light curves, 476 time points



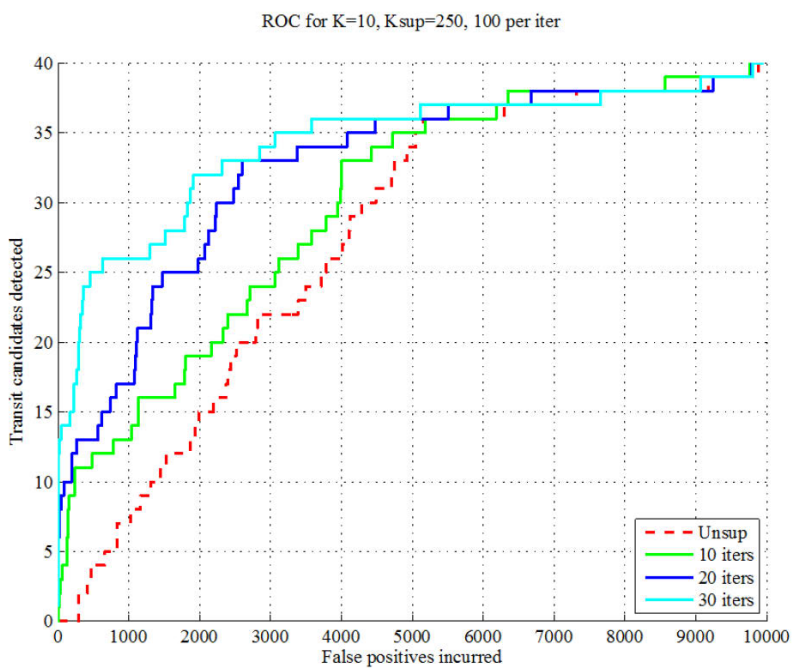


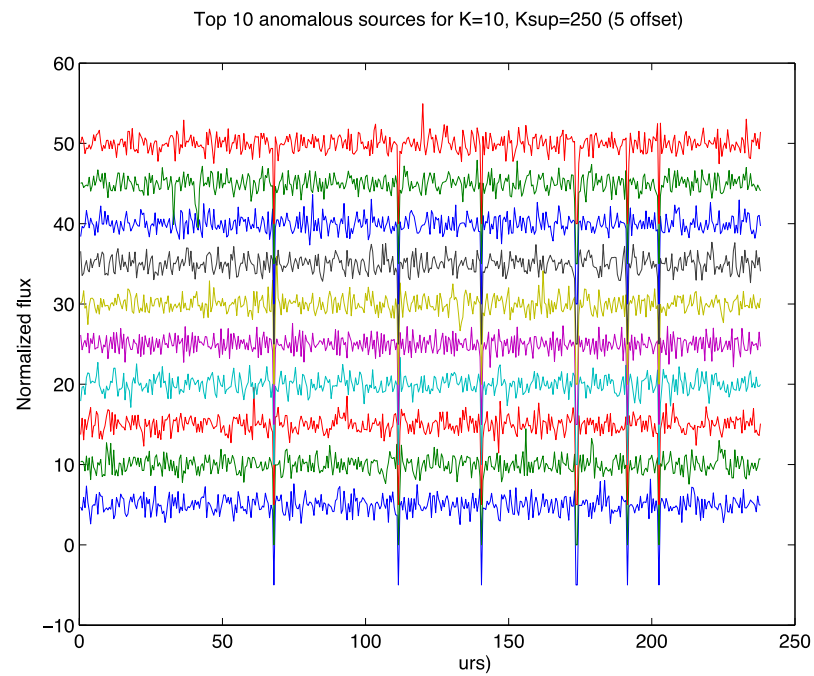
Iterative Discovery with SSEND

Each iteration:

examine top N, add FPs to "ignorable" set

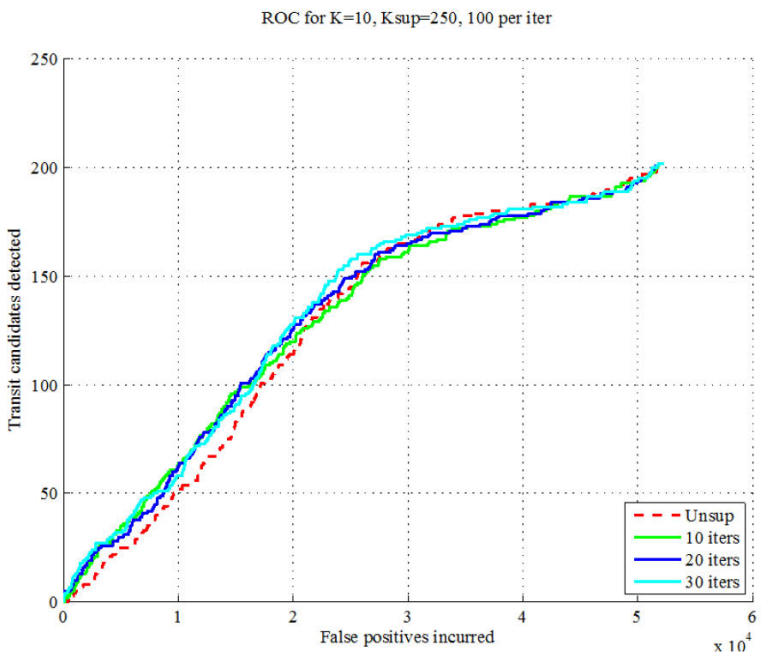
Best results with small K, large Ksup

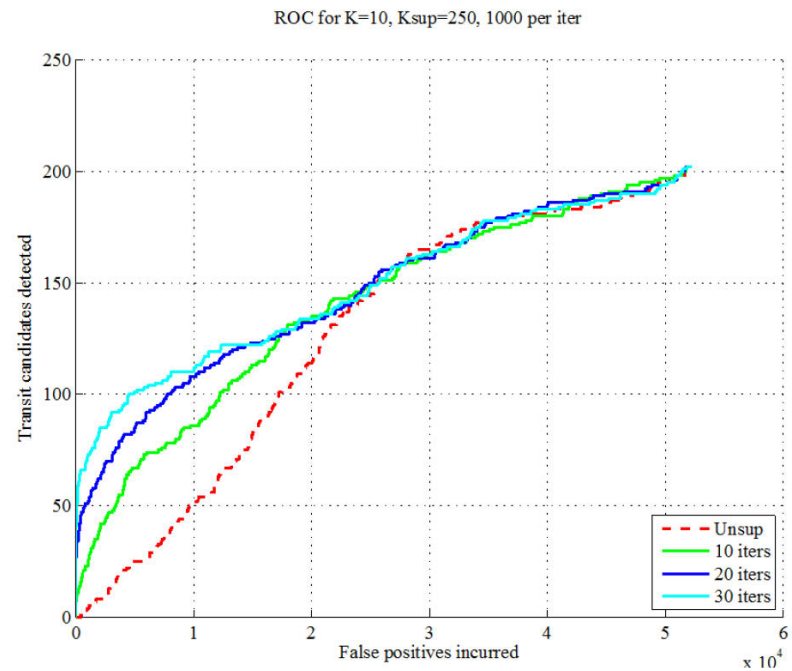




How about the full data set?

- Requires more review, more training
- 100 vs. 1000 per iteration:





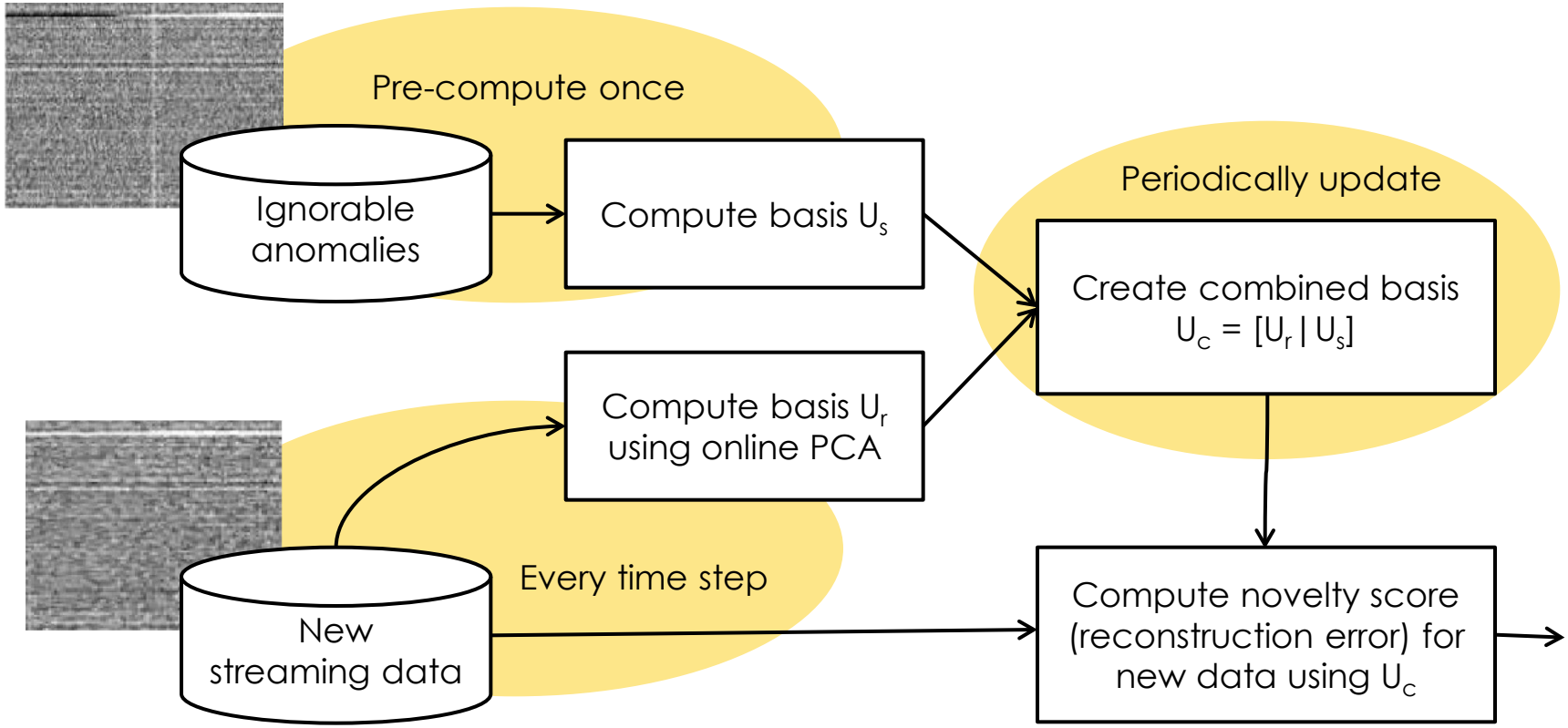
Summary: Digging deeper through interference

Novelty detection that

- Adapts to changing data properties
- Avoids known false alarms

Novelty score based on linear bases: a modular approach that accommodates online and trained components

Promising initial results for radio, optical astronomy

Extra slides…

SSEND Concept: online, trainable novelty detection