SEMI-SUPERVISED NOVELTY DETECTION WITH ADAPTIVE EIGENBASES, AND APPLICATION TO RADIO TRANSIENTS

DAVID R. THOMPSON[1,2], WALID A. MAJID[2], COLORADO J. REED[2], AND KIRI L. WAGSTAFF[2]

Abstract. We present a semi-supervised online method for novelty detection and evaluate its performance for radio astronomy time series data. Our approach uses adaptive eigenbases to combine 1) prior knowledge about uninteresting signals with 2) online estimation of the current data properties to enable highly sensitive and precise detection of novel signals. We apply the method to the problem of detecting *fast transient* radio anomalies and compare it to current alternative algorithms. Tests based on observations from the Parkes Multibeam Survey show both effective detection of interesting rare events and robustness to known false alarm anomalies.

1. Introduction

Recent discoveries in high time resolution radio astronomy data have drawn attention to a new class of sources. *Fast transients* are rare pulses of radio-frequency energy lasting from microseconds to seconds that might be produced by a variety of exotic astrophysical phenomena [6, 5, 11, 12]. For example, X-ray bursts, neutron stars, active galactic nuclei, and extraterrestrial intelligence (ETI) are all possible sources of short-duration transient radio signals. Such events are often discovered serendipitously in high time resolution data collected for other purposes. These transients are generally faint and difficult to detect, so improved detection algorithms can directly benefit the science yield of all such commensal monitoring. Existing detection approaches rely on a specific *dispersed pulse* model of the signal shape. This paper presents a new method for analyzing real-time high-resolution radio astronomy data that operates without this model assumption. Therefore, it can potentially detect a far broader class of anomalous events in real time, as well as unexpected events that do not match a known profile.

We have formulated fast transient monitoring as statistical anomaly detection in a time series [16, 17]. The main challenges of our domain are:

- **High dimensionality:** Signals of interest span multiple antenna power measurements that could include hundreds of time steps and frequency channels.
- **Real time processing:** With the exception of a few dedicated surveys, most high time resolution data is too voluminous to archive. Therefore, events must be detected in real time to select only the most interesting candidates for storage and later exhaustive analysis.
- **Nonstationarity:** Background noise characteristics change over time on medium to long scales, manifesting as narrow-band noise or large-scale gain fluctuations that can appear and disappear in response to hardware and observing conditions. Detection of anomalous "fast" signals should be robust to these changes.
- **False alarms:** Certain known classes of events, such as momentary Radio Frequency Interference (RFI), are not astronomically interesting but are easily mistaken for fast transients. It is important to avoid flagging these events as novel to avoid filling the detection buffer with these false alarm events. Further, false alarms waste valuable astronomer time in reviewing the results.

[1] david.r.thompson@jpl.nasa.gov
[2] Jet Propulsion Laboratory, California Institute of Technology.

This work proposes a new solution that learns a low-dimensional linear manifold for describing the "normal" data. The novelty of our approach lies in combining basis vectors learned in an unsupervised, online fashion from the data stream with supervised basis vectors learned in advance from known false alarms. We thereby achieve adaptive, data-driven anomaly detection that also exploits prior domain knowledge about signals that may be statistically anomalous but are not interesting and should therefore be ignored. We identify truly interesting anomalies by compressing and reconstructing the data [9] using the combined basis. High reconstruction error indicates a signal that does not match the learned profile of the normal data. The unsupervised component uses the incremental method of Lim et al. [13], an efficient online algorithm that satisfies real-time constraints.

We evaluated the new method using data from the Parkes Multibeam Survey. This data set was originally collected to search for pulsars, which are astronomical sources that emit radio pulses at regular periods. However, several non-pulsar anomalies have recently been discovered in this dataset [3], making it a compelling test case. We found that by explicitly filtering known false alarm patterns, our approach yields significantly better performance than current transient detection methods. This method shows promise for use in current and future astronomical surveys, including data to be collected by the Square Kilometre Array, a radio telescope currently under development that will be 50 times more sensitive than any existing instrument.

2. Related Work

Generic approaches to anomaly detection are data-driven: they typically learn a representation of the "normal" or uninteresting data, then identify any observations that do not match this model. One such method is one-class support vector machine (SVM) classification [18], in which an SVM is trained only on examples from the normal class and then detects any new data belonging to a different, previously unobserved class. More recent efforts seek to include user-labeled examples. Blanchard et al. [2] propose a semi-supervised technique that trains a classifier using two kinds of data: labeled data known to be normal and an additional unlabeled sample that may contain anomalous data. Both approaches aim to train a binary classifier that labels new items as either "normal" or "anomalous." The Blanchard technique further accommodates an upper limit on the false anomaly detection rate. Our approach differs from these methods in that it specifically incorporates known examples of false alarms to further improve the system's precision. In addition, our system is designed for online operation rather than batch processing of previously collected data.

In contrast with statistical novelty detection, radio astronomers generally use physical models of the anticipated events. If the precise shape of the event is known in advance, matched filtering provides maximum sensitivity to detect faint transient pulses. These models reflect the fact that signals from remote astronomical sources are *dispersed*. As the signal travels through the interstellar medium that lies between the source and the observer, it encounters free electrons that absorb some of the signal's energy and delay its propagation. This affects lower frequency components more than higher frequency components. The slight difference accumulates over long distances and ultimately causes a broadband signal to appear dispersed in time, so that the lower frequency components arrive later.

Real-time transient detection uses *incoherent* radio array data organized as a matrix of discretely channelized measurements. In other words, observations occur across a range of radio frequency channels at each time step. This data is typically portrayed as a two-dimensional image in which the axes correspond to time and frequency. The pixel intensity shows observed power, the accumulated squared voltage received by the antenna. Figure 1 (left) shows a pulse from pulsar J0742-2822 that displays a typical dispersed "sweep." Dispersion manifests as a time delay Δt_{delay} that is inversely proportional to the signal's frequency. Following [14]:

$$\Delta t_{\text{delay}} = 4.1\text{ms}\frac{DM\ k}{\Delta\nu^2_{\text{GHz}}} \tag{1}$$

Here $\Delta\nu$ is the frequency difference. The amount of dispersion, or the Dispersion Measure (DM), correlates with the number of interfering electrons present between the source and the observer [1]. It is commonly reported in parsecs per cm^3. For regions of constant electron density, the amount of dispersion suggests the physical distance to the source.

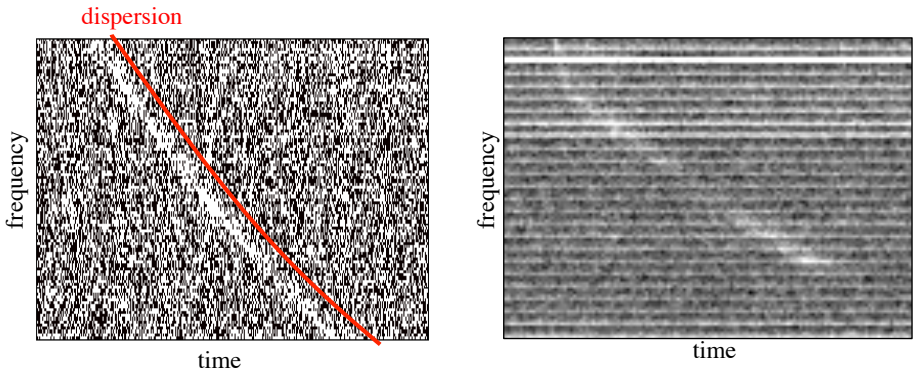

FIGURE 1. Examples of typical and atypical transient signals. The image at left shows a single pulse from pulsar J0742-2822, with a classic dispersed pulse profile. Such signals can be found by inverting the dispersion effect prior to matched filtering. More exotic and poorly understood phenomena, like the *peryton* signal pictured at right, do not match typical dispersion and could benefit from model-free detection strategies with fewer assumptions. This example shows a distinctive "kink" in the curved signal. The narrow horizontal lines are narrow-band interference; such behavior is time-variable but not astronomically relevant and would ideally not affect the detection decision.

The most common approach to detecting remote transient signals is tailored to the known properties of dispersion. Data is exhaustively *dedispersed* using a variety of different candidate DMs [1, 4]. Dedispersion re-aligns the observations in time to undo the effects of a given assumed DM, and then sums the resulting signal across all frequency channels to yield a detection statistic. This tailored summation is equivalent to a matched filter, and increases detection sensitivity over a naive sliding window detection using all frequency channels. By seeking the maximum dedispersed sum across all DMs, one can characterize the signal (and roughly the distance to the source). A dedispersion search can also help separate genuine astronomical signals from Radio Frequency Interference (RFI). Broadband RFI manifests as a vertical signal with no dedispersion (DM = 0); the pulse originates locally and all frequencies arrive simultaneously.

This approach has proven effective for the detection of pulsars and other astronomical phenomena [7, 11, 6, 12]. It can be implemented efficiently to keep up with streaming data using FPGAs, GPUs, or other parallel architectures; dedispersion over multiple DMs is inherently highly parallelizable. The weakness of this approach, however, is that it is sensitive to only one kind of signal. While dispersion is a known phenomenon of all remote signals, some recently-discovered sources (Figure 1, right) exhibit deviations from the expected shape which renders them difficult to detect. Further, it is not known how many other exotic source types may currently be overlooked due to the detection method's dependence on one kind of signal model. The next section presents a more flexible strategy that could operate in parallel with dedispersion searches, providing the capability to detect both dispersed pulses and unanticipated novel events.

3. Approach

We propose a new approach that combines 1) prior knowledge about uninteresting signals with 2) online estimation of the current data properties to enable flexible detection of novel signals. We treat the data as a sequence of observations that arrive sequentially from the antenna. We combine n such observed data points $\mathbf{x}_i \in \mathbb{R}^d$ as columns of a $d \times n$ data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_n]$. Here, d is the number of frequency channels observed at each time step. The goal is to compute a discriminant function that maps each observation to a novelty score, $f(\mathbf{x}_i) : \mathbb{R}^d \mapsto \mathbb{R}$. The discriminant value should be small for typical data but large for interesting or novel data.

3.1. **Constructing an eigenbasis.** We exploit a popular strategy, detailed in [19] and [9], of measuring the distance from the signal to a low-dimensional manifold learned from the data stream. We will start by describing the simpler case of novelty detection in a static (non-adaptive) subspace. We hypothesize that the "regular" data lies on a linear subspace in $\mathbb{R}^{d'}$ with $d' \ll d$. Subtracting the data mean $\bar{\mathbf{x}}$ yields a translated matrix $\tilde{\mathbf{X}} = [(\mathbf{x}_1 - \bar{\mathbf{x}}), (\mathbf{x}_2 - \bar{\mathbf{x}}), \ldots, (\mathbf{x}_n - \bar{\mathbf{x}})]$. Singular Value Decomposition (SVD) provides $\tilde{\mathbf{X}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. The columns of $\mathbf{U}$ are the *principal components*: an orthonormal basis with axes in the order of decreasing data variance. We form a low-dimensional basis $\mathbf{A}$ using the first d' columns of $\mathbf{U}$. When $n > d$ the SVD decomposition is undefined, but one can still compute the matrix $\mathbf{A}$ via classical Principal Component Analysis (PCA), e.g., using the eigenvectors corresponding to the largest eigenvalues of the covariance matrix $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$.

We quantify the novelty of observation $\mathbf{x}_i$ using the Euclidean distance to the subspace, equivalent to the L2-norm *reconstruction error* after first transforming $\mathbf{x}_i$ into the low-dimensional basis and then reconstructing an approximation $\hat{\mathbf{x}}_i$. This suggests the following discriminant function which is large for novel data and zero for points on the linear manifold.

$$f(\mathbf{x}_i) = \|\mathbf{x}_i - \hat{\mathbf{x}}_i\| = \|(\mathbf{x}_i - \bar{\mathbf{x}}) - \mathbf{A}\mathbf{A}^T(\mathbf{x}_i - \bar{\mathbf{x}}))\|_2 \tag{2}$$

The eigenvalue decomposition makes computing $\mathbf{A}$ difficult for large n. However, it is important that our basis accommodate large data sets and long-timescale changes in the background. One solution is to periodically recompute the entire matrix $\mathbf{A}$ in batch mode using a recent subset of the data. In this work we favor the approach of Lim et al. [13] for efficient online updates to the mean $\bar{\mathbf{x}}$ and eigenbasis $\mathbf{A}$. This approach updates an SVD decomposition defined by some previous data $\tilde{\mathbf{X}}_p = \mathbf{U}_p\mathbf{\Sigma}_p\mathbf{V}_p^T$. Each block update has a data matrix $\mathbf{X}_q$ with mean $\bar{\mathbf{x}}_q$ and decomposition $\tilde{\mathbf{X}}_q = \mathbf{U}_q\mathbf{\Sigma}_q\mathbf{V}_q^T$. This gives a combined dataset $\mathbf{X}_r = [\mathbf{X}_p|\mathbf{X}_q]$. Fortunately one can compute an updated mean $\bar{\mathbf{x}}_r$ and eigenbasis $\tilde{\mathbf{X}}_r = \mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^T$ without having to store the old data explicitly. We refer the reader to the Lim et al. text for details, but summarize their approach in Algorithm 1 below. It relies on the widely-studied R-SVD procedure [8] which exploits the fact that a low-rank update to the eigenbasis is decomposable into efficient block operations. The method extends R-SVD to the case where the data are not assumed to have zero mean.

Algorithm 1: Lim et al. Algorithm for Sequential Eigenbasis Updates.

Input: Previous mean $\bar{\mathbf{x}}_p$
Previous decomposition $\mathbf{U}_p\mathbf{\Sigma}_p\mathbf{V}_p^T$
Additional data $\mathbf{X}_q$
Output: Revised mean $\bar{\mathbf{x}}_r$
Revised decomposition $\mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^T$
Compute $\bar{\mathbf{x}}_r = \frac{n}{n+m}\bar{\mathbf{x}}_p + \frac{m}{n+m}\bar{\mathbf{x}}_q$, where $n = |\mathbf{X}_p|$ and $m = |\mathbf{X}_q|$
Compute $\mathbf{E} = \left(\mathbf{X}_q - \bar{\mathbf{x}}_r 1_{(1\times m)} | \sqrt{\frac{nm}{n+m}}(\bar{\mathbf{x}}_p - \bar{\mathbf{x}}_q)\right)$
Use $\mathbf{U}_p\mathbf{\Sigma}_p\mathbf{V}_p^T$ with $\mathbf{E}$ as input to R-SVD to compute $\mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^T$

An advantage of the Lim et al. method is that one can downweight the old basis to introduce a forgetting factor that allows the influence of old data to decay gradually as new points are added.

This lets the basis shift to track a nonstationary distribution, and it accommodates observations of arbitrary length.

3.2. **Semi-supervised eigenbases.** Automated novelty detection may need to exclude certain rare events that are known in advance to be uninteresting. For example, there may be known *false alarms* due to rare but recognizable noise. Another case where false alarms can be anticipated in advance is through feedback from a human user based on follow-up processing of previous results. This feedback could determine which previous detections had been erroneous. We incorporate information about known false alarm patterns with a second basis trained to model these known examples. Our semi-supervised novelty detection method uses a combined subspace with both supervised and unsupervised components, shifting with long-term trends while still excluding the false alarms. Algorithm 2 summarizes the procedure. We train an initial low-dimensional basis $\mathbf{U}_s$ using data known to be uninteresting. At runtime, we compute an adaptive mean $\bar{\mathbf{x}_r}$ and basis $\mathbf{U}_r$ using the Lim et al. method, and also define a combined basis $\mathbf{U}_c = [\mathbf{U}_r|\mathbf{U}_s]$ to span both the supervised basis and the unsupervised data. We orthogonalize the new basis using QR decomposition with the Gram-Schmidt method. The reconstruction error with respect to the combined basis yields a more reliable novelty score.

Algorithm 2: Semi-Supervised Eigenbasis Novelty Detection.

Input: Supervised training data $\mathbf{X}_s$ of size l
Size m block updates of streaming, unsupervised data $\mathbf{X}_u$
Output: Novelty scores $f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots$
Compute $\tilde{\mathbf{X}}_s = [\mathbf{x}_{s1} - \bar{\mathbf{x}}_s,\ \mathbf{x}_{s2} - \bar{\mathbf{x}}_s,\ \ldots, \mathbf{x}_l - \bar{\mathbf{x}}_s]$
Use PCA with $\tilde{\mathbf{X}}_s\tilde{\mathbf{X}}_s^T$ or SVD with $\tilde{\mathbf{X}}_s = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^T$ to compute an orthonormal basis $\mathbf{U}_s$
Using the first block $\mathbf{X}_{u1}$, compute an initial mean $\bar{\mathbf{x}}_p$ and eigenbasis $\mathbf{U}_p\mathbf{\Sigma}_p\mathbf{V}_p^T$
For each subsequent $\mathbf{X}_u$:
  Compute a revised mean $\bar{\mathbf{x}}_r$ and a revised decomposition $\mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^T$ using the Lim et al. method
  Define a combined basis $\mathbf{U}_c = [\mathbf{U}_r|\mathbf{U}_s]$
  Use QR decomposition to find a basis $\mathbf{U}_c'$ that makes $\mathbf{U}_c$ orthonormal.
  For each $\mathbf{x}_i \in \mathbf{X}_u$:
    Compute $f(\mathbf{x}_i) = \|(\mathbf{x}_i - \bar{\mathbf{x}}_r) - \mathbf{U}_c'\mathbf{U}_c'^T(\mathbf{x}_i - \bar{\mathbf{x}}_r)\|_2$

Note that the proposed approach does not preserve the mean of the initial false alarm distribution, which is assumed to drift in a similar fashion as the mean of the dynamic distribution. User feedback would permit a more sophisticated system that also updates the false alarm mean and basis online, but we focus here on the simpler case where all training occurs in advance.

4. Evaluation

This semi-supervised anomaly detection method was motivated by applications in radio astronomy. We performed a comparative evaluation on a test set of radio array data using five linear and nonlinear novelty detection algorithms: the traditional dedispersion approach, kernel PCA novelty detection [9], one-class SVM novelty detection [18], unsupervised adaptive novelty detection using PCA, and the proposed semi-supervised approach.

4.1. **Data set.** We use a selected portion of data from the Parkes Multibeam Survey, an extensive search for Pulsars using the Parkes radio telescope of CSIRO [7, 15, 10]. This instrument views the sky simultaneously through 13 receivers, effectively providing 13 independent antennas covering adjacent, and slightly overlapping, areas in the sky. Receiver measurements are recorded at high time resolution and transformed into channelized power measurements corresponding to the squared voltage response at various discrete frequency channels. This specific data sequence contains examples of events known as *perytons*, first discovered by Burke-Spoloar and Bailes in their analysis

of Parkes pulsar surveys [3]. Perytons are still poorly understood, and they are scientifically interesting because they vary in frequency and approximate a dispersion curve. However, they do not exactly match a dispersion profile, and their spatial distribution in the sky suggests that they are of terrestrial (possibly atmospheric) origin.

In addition to these features, structured interference is often visible in the form of channel-specific noise and gain fluctuations appearing as horizontal stripes. Such noise is pervasive and typical for highly sensitive, cryogenically cooled receiver feeds. Our tests focus on approximately five minutes of observation time in each of the 13 receivers. This span includes several tens of thousands of timesteps recorded at a cadence of 0.125 milliseconds in each of 96 frequency channels near 1450 MHz. Figure 2 shows three examples of perytons. The red rectangle shows the size of an example data window used to construct $\mathbf{x}_i$.

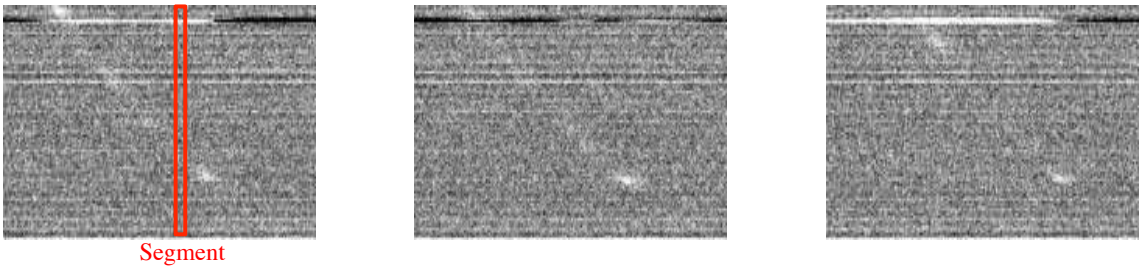

FIGURE 2. True anomalies: Peryton events from the Parkes Multibeam Survey.

Figure 3 shows some examples of false alarms that are statistically uncommon but not scientifically interesting. These specific examples are broadband pulses of Radio Frequency Interference (RFI), probably emitted by some local artificial source. Such features are rare enough that they are not well-represented in an unsupervised eigenbasis, but typical enough that they would dominate novelty detection results if not handled explicitly.

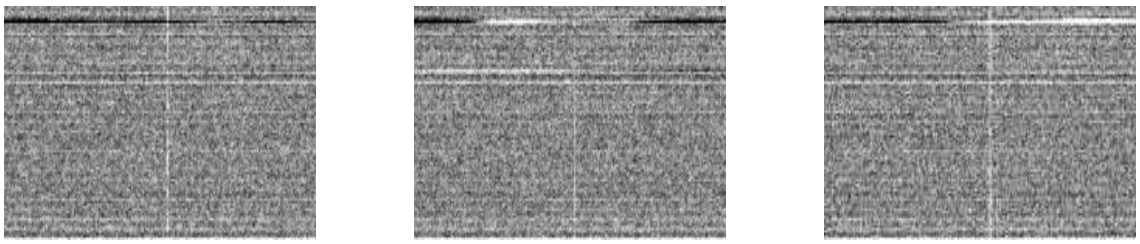

FIGURE 3. False anomalies: Vertical stripes due to broadband RFI that is statistically anomalous but uninteresting.

4.2. **Methodology.** We subsample the data by a temporal factor of 20 so that it has a resolution of 2.5 ms, and then analyze the data as a sequence of short non-overlapping *segments* that cross all 96 vertical frequency channels and 6 horizontal timesteps. This segment width corresponds to a $15ms$ time interval, found empirically to be the best-performing value for all methods. We reorder each segment into a single column vector $\mathbf{x} \in \mathbb{R}^{384}$. Finally, we unify data from all beams into one large dataset, witholding five beams (38%) for training purposes.

We compare five different detection methods that are broadly representative of different linear and nonlinear anomaly detection approaches. First, we consider the proposed semi-supervised method that combines supervised and unsupervised components and reports reconstruction error $f_{ss}(\mathbf{x}_i)$.

Here we trained the subspace $\mathbf{U}_s$ using 30 overlapping segments ($\mathbf{X}_s$) drawn from three manually-selected broadband RFI pulses. The second method is a purely unsupervised eigenbasis approach based on reconstruction error from a low-dimensional basis $f_u(\mathbf{x}_i)$. It does not explicitly account for RFI. The third method is the one-class SVM novelty detection of Scholkopf et al. [18]. Here we use a radial basis kernel function selected with a grid search, and treat each test point's distance to the decision boundary as a real-valued novelty score.

The fourth method is kernel PCA: a non-linear extension of PCA. Kernel PCA novelty detection first maps the data to a higher (generally infinite) dimensional features space, computes the principal components in this space, projects the transformed data to a lower-dimension manifold, and defines a novelty measure as the reconstruction error in the feature-space. Kernel functions allow the reconstruction error to be calculated without explicitly [9]. However, this method never explicitly calculates the principal components so it cannot be used as an adaptive technique in the manner discussed in Algorithm 2. Instead we use the implementation of Hoffmann et al [9]. We use a radial basis kernel function with parameters selected by a grid search.

Finally, we consider a state-of-the-art incoherent dedispersion and summation method which searches DM values from 0 and 500. We correct each time step separately for each DM, use and the maximum response from all DMs as the novelty score $f_d(\mathbf{x}_i)$. Time averaging did not improve performance, so we report the single-timestep result.

In addition to labeling RFI, we obtained the precise locations of all peryton events noted in the study by Burke-Spolaor et al [3]. These appeared to some degree in all antennas, though the signal strength and character varied somewhat even for simultaneous observations. The concatenated dataset provided 88 real novel events for our evaluation. We assigned each peryton an enclosing time interval; any detection in this range counted as having successfully detected the peryton. Note that we use the same time interval for all beams regardless of the actual signal strength. Perytons that are weak in one or more beams penalize all detection methods equally, so we assume events are always present for our (relative) performance assessment.

We evaluated each method by first computing novelty scores for the entire data set, sorting these scores across all beams, and then counting the result of each trigger in order of decreasing novelty. Each peryton can only be captured once, though multiple triggers within the same event do not count as false positives. However, any detection falling outside a peryton interval counts as a false positive.

4.3. **Results.** A visualization of the unsupervised and supervised bases created by our method appears in Figure 4. Here we use 4 principal components as an unsupervised basis with online updates from the data stream. These *eigensignals* (Figure 4, left) show high magnitude in the most variable channels; at the time this eigensignal snapshot was captured, such channels comprised the major axis of variance for the data set. A supervised basis of 10 dimensions models the known broadband RFI. The top five such eigensignals appear in Figure 4, Right; they reflect the vertical profile of momentary RFI pulses. Together, the two can accurately reconstruct a slow shift in channelized RFI conditions along with any additive broadband pulses. This image shows the orthonormal segments after QR factorization.

Figure 5 compares novelty detection scores for the entire observation sequence of the first test beam, computed with a purely unsupervised basis (standard PCA), f_u, and the semi-supervised approach, f_{ss}. Interesting peryton events are noted by black triangles; the other signal spikes correspond to various kinds of RFI. Five peryton signals were barely visible in the reconstruction error of either method, due possibly to the alignment of non-overlapping segments or the inherently weak visibility of those events in this beam. We exclude these five from the diagram for clarity. In general the semi-supervised approach responds to the novelty of peryton events while filtering most of the RFI. In contrast, broadband RFI contaminates the purely unsupervised approach; it accounts for the three strongest responses by f_u for this sequence.

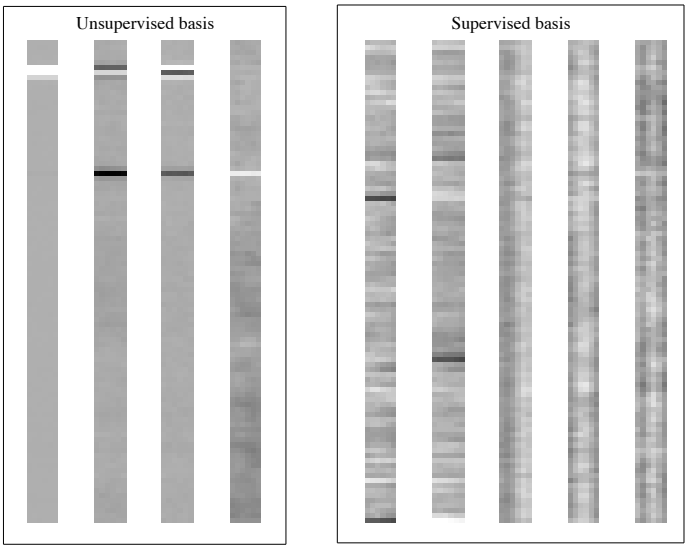

FIGURE 4. Orthonormal principal components used to construct $\mathbf{U}_c$ from $\mathbf{U}_r$ (left) and $\mathbf{U}_s$ (right). The unsupervised portion (left) models channelized interference, while the vertical structures in the supervised portion (right) represent momentary broadband RFI.

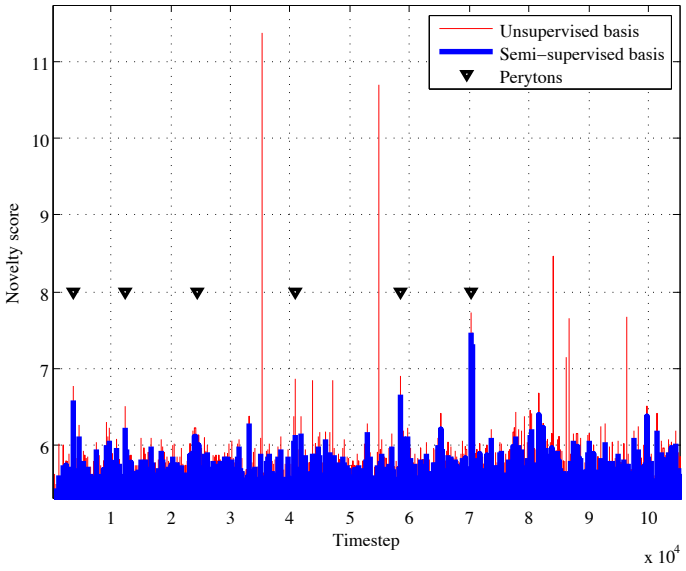

FIGURE 5. Semi-supervised learning filters out RFI events that would otherwise dominate the detection results. This time series plot shows per-timestep novelty evaluated for the first beam in the test set. Not all perytons are clearly distinguishable in this beam.

Figure 6 shows a Receiver Operating Characteristic (ROC) curve describing the tradeoff in precision and recall rates. We report the number of perytons captured for a variety of false positive budgets. For real-time observations, false positive budgets beyond 10 are excessive. Generating more than 10 false positives would represent greater than one detection event for every 5 seconds of

observations, imposing an unrealistic burden on manual post-analysis. Future commensal campaigns with constant observations and higher data volumes will demand even stricter limits. For this low error budget, the semi-supervised approach considerably outperforms the competing methods: the top 12 signals detected via f_{ss} are due to perytons, while the kernel PCA technique detects 5 false-positives before the first peryton, and the unsupervised method reports more than 40 false positives before finding the first real peryton. These runner-up methods require 250 and 200 false positives respectively before they match the error-free retrieval rate of the semi-supervised approach. Note that one might improve performance of any method further with additional hand-crafted RFI excision rules, such as a ban on zero-DM detections that are likely to be terrestrial. Naturally, such rules are less general than a learning-based approach and might filter other interesting but unanticipated phenomena.

The preceding results used a data segment size of 15 ms (6 time steps) to compose $\mathbf{x}_i$. We evaluated sensitivity to segment length (see Figure 7). Segments of duration $10 - 15$ ms performed best for this data set. It is possible that smaller segments are susceptible to noise while larger sizes dilute the perytons. It might improve performance for large segments to use a higher-dimensional basis for the unsupervised component. Such models might do a better job of modeling temporal structure (such as switching interference) that begins to appear at these scales.

We also assessed the runtime of each method to determine whether they could be used in a realistic real-time setting. Using a single core of a modern desktop processor, the runtime of the dedispersion search method averaged 0.16 seconds per DM for the entire subsampled sequence, or ≈ 80 seconds for a typical search over 500 DM values. This could be divided easily among multiple processors to provide faster processing for multiple beams. The eigenbasis approaches' runtimes depend strongly on the size of the block updates to the eigenbasis. For a single desktop processor core performing block updates of size $m = 100$, the entire observation from a single beam was processed at $5\times$ real time (≈ 10 seconds/beam for the entire dataset). The time required was slightly larger (up to ≈ 20 seconds/beam) for smaller block updates where constant-time overhead costs had a larger impact. The accuracy of these techniques was nearly indistinguishable for all block update sizes we tried. Varying the segment sizes also affected run time by up to a factor of two. Kernel PCA and one-class SVM performed considerably slower than the dedispersion and eigenbasis approaches as all computations were performed with an RBF kernel representation of the data: a representation of size $|\mathbf{x}|^2 = 384^2$ for this work. In our experiments we found these techniques required $\approx 200 - 400$ seconds/beam with block updates of $m = 400$. Furthermore, unlike the dedispersion and eigenbasis techniques, the Kernel PCA and one-class SVM computation times scale quadratically with the size of m. This reduces the generality of these methods, and in combination with their large computational run-times, makes them unfeasible as real-time techniques. On the other hand, we found the dedispersion and eigenbases approaches to be readily employable for real-time use on general purpose computing hardware.

5. Discussion

Semi-supervised eigenbases have general applicability for anomaly detection in domains with real-time requirements, high-dimensional input, and prior knowledge about false alarm events. Of course, it is not necessary to incorporate this false alarm information directly into the novelty detection model as we have done here. One could perform pre-classification to filter these events prior to a purely unsupervised novelty detection stage. Nevertheless, there may be other advantages to a combined approach. It is simple and easy to implement. The projection shifts to reflect any underlying drift in the mean signal levels, so that a basis trained on previous false alarms remains relevant. Further work will investigate ways to combine multi-scale models when the temporal extent of the interesting events is not known in advance. Finally, application to the broader Parkes survey catalogue will increase practical experience with the technique, and may even reveal additional classes of RFI and astronomical transient events.

6. Acknowledgements

This work was made possible through the assistance of Sarah Burke-Spolaor, formerly of CSIRO and now affiliated with the Jet Propulsion Laboratory. More generally we would also thank Swinburne University and CSIRO for a generous data access policy. We also thank J-P Macquart (CSIRO), and Dayton Jones, Robert Preston and Joseph Lazio (Jet Propulsion Laboratory) whose advice and ideas have influenced our investigation. This research was performed at the Jet Propulsion Laboratory, California Institute of Technology, under the Research and Technology Development Strategic Initiative Program.

References

[1] D. Bhattacharya. Detection of radio emission from pulsars. *NATO ASIC Proc. 515: The Many Faces of Neutron Stars*, 1998.

[2] G. Blanchard, G. Lee, and C. Scott. Semi-supervised novelty detection. *Journal of Machine Learning Research*, 11:2973–3009, 2010.

[3] S. Burke-Spolaor, M. Bailes, R. Ekers, J. Macquart, and F. Crawford III. Radio Bursts with Extragalactic Spectral Characteristics Show Terrestrial Origins. *The Astrophysical Journal*, 727:18, 2011.

[4] J. Chennamangalam et al. Software data-processing pipeline for transient detection. *The Low-Frequency Radio Universe, ASP. Conf. Series*, LFRU, 2009.

[5] J. M. Cordes et al. The dynamic radio sky. *New Astronomy Reviews*, 48:1459–1472, 2004.

[6] J. M. Cordes and M. A. McLaughlin. Searches for fast radio transients. *The Astrophysical Journal*, 596:1142–1154, October 2003.

[7] R. T. Edwards, M. Bailes, W. van Straten, and M. Britton. The Swinburne Intermediate Latitude Pulsar Survey. *MNRAS*, 326:358–374, 2001.

[8] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins Univ Pr, 1996.

[9] H. Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition*, 40(3):863–874, 2007.

[10] B. A. Jacoby, M. Bailes, S. M. Ord, R. T. Edwards, and S. R. Kulkarni. A large-area survey for radio pulsars at high galactic latitudes. *The Astrophysical Journal*, 699:401–411, 2009.

[11] E. Keane. The search for nearby RRATs and other transient radio bursts. In *Third Estrela Workshop*, 2008.

[12] J. Lazio, J. S. Bloom, G. C. Bower, J. Cordes, S. Croft, S. Hyman, C. Law, and M. McLaughlin. The dynamic radio sky: An opportunity for discovery. *Astro2010: The Astronomy and Astrophysics Decadal Survey White Paper no. 176*, 2009.

[13] J. Lim, D. Ross, R. Lin, and M. Yang. Incremental learning for visual tracking. *Advances in Neural Information Processing Systems*, 1:793–800, 2004.

[14] A. G. Lyne and F. G. Graham-Smith. *Pulsar Astronomy*. Cambridge University Press, 1998.

[15] R. N. Manchester et al. Parkes multibeam pulsar survey: I. observing and data analysis systems, discovery and timing of 100 pulsars. *MNRAS*, 328:17–35, 2001.

[16] M. Markou and S. Singh. Novelty detection: a review. Part 1: statistical approaches. *Signal Processing*, 83(12):2481–2497, 2003.

[17] M. Markou and S. Singh. Novelty detection: a review. Part 2: neural network based approaches. *Signal Processing*, 83(12):2499–2521, 2003.

[18] B. Schöolkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1472, 2001.

[19] S. O. Song, D. Shin, and E. S. Yoon. Analysis of novelty detection properties of auto-associators. In *Proceedings of COMADEM*, pages 577–584, 2001.

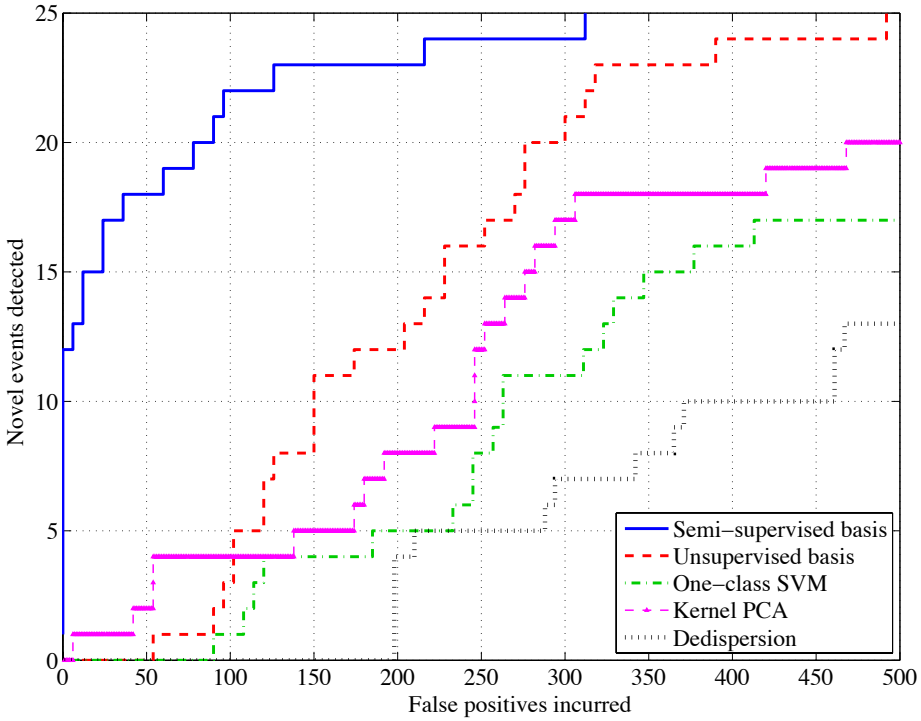

FIGURE 6. ROC curves comparing eigenbasis novelty detection approaches with the traditional dedispersion search.

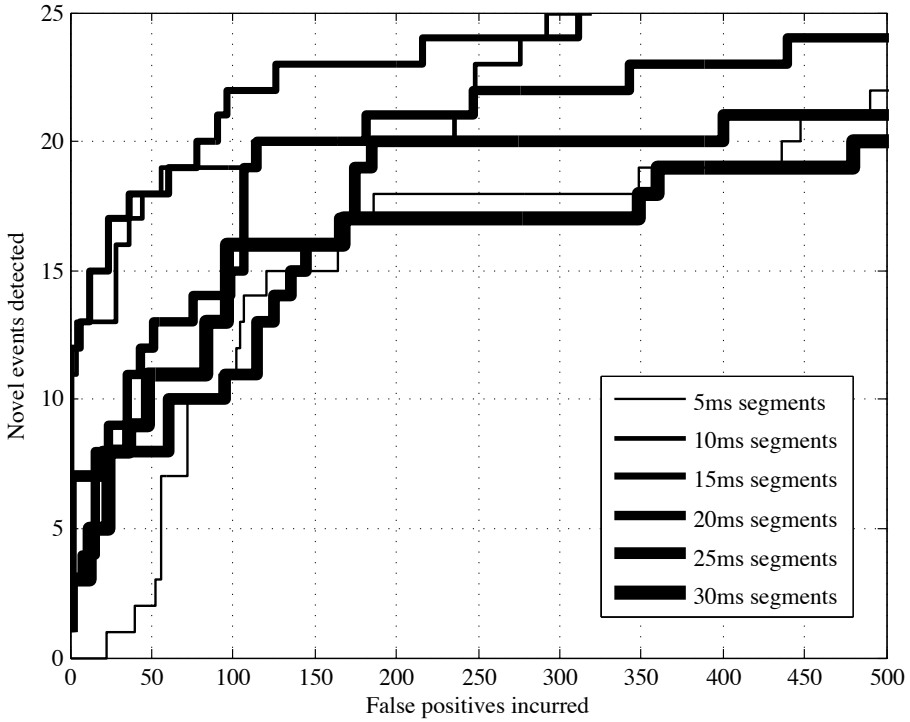

FIGURE 7. ROC curves to assess the sensitivity of semi-supervised detection to data segment sizes.