ROBUST DETECTION OF RADIATION THREAT USING UNCERTAIN CENSORED ENERGY WINDOWS

Eric Lei[1], Kyle Miller[1], Peter Huggins[1], Karl Nelson[2], Simon Labov[2], Artur Dubrawski[1]

[1] Auton Lab, Carnegie Mellon University

[2] Lawrence Livermore National Laboratory

Acknowledgements

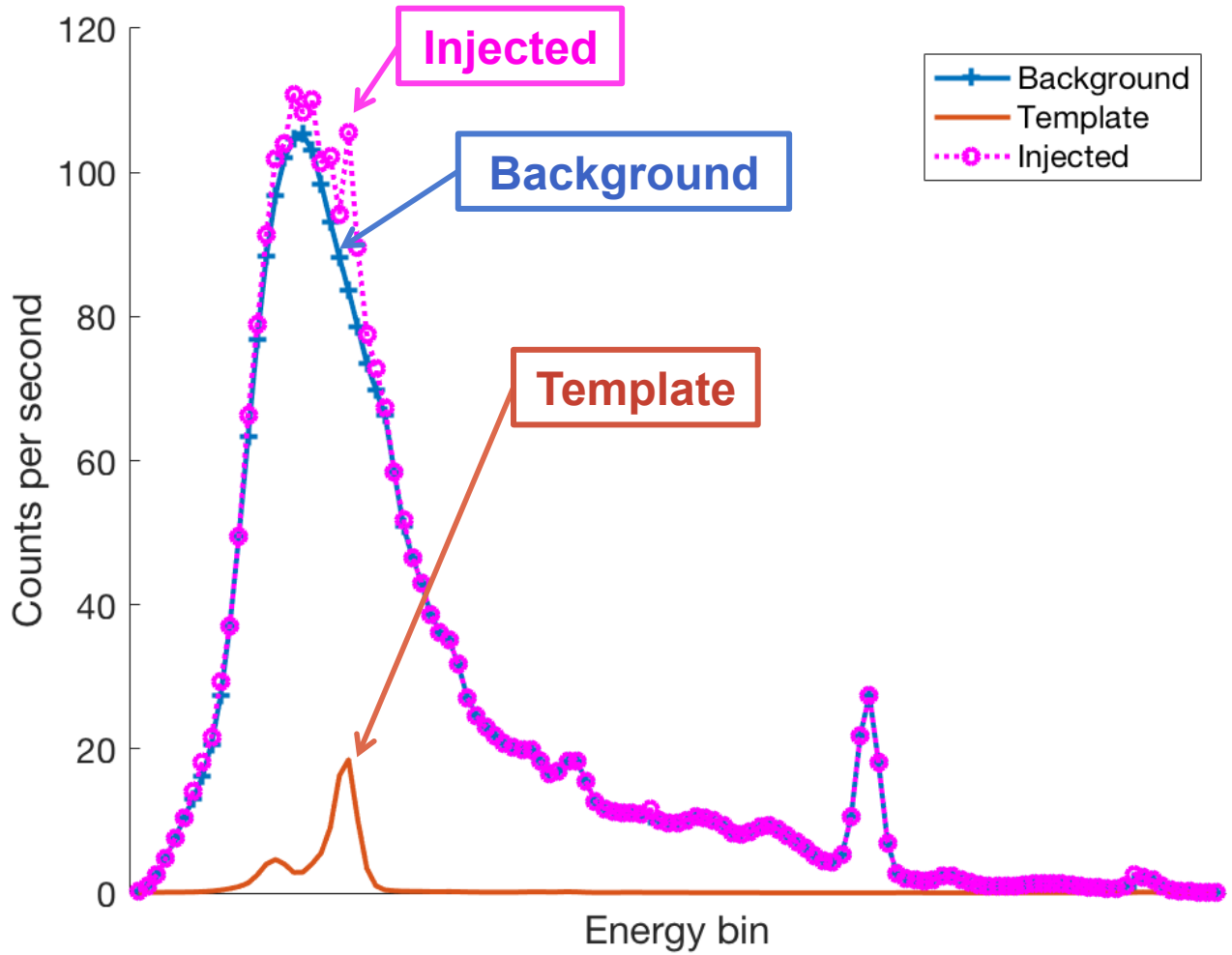
This work has been partially supported by the

- U.S. Department of Homeland Security, Domestic Nuclear Detection Office, under competitively awarded grant 2014-DN-077-ARI087-01.
- U.S. Department of Defense, Defense Threat Reduction Agency under award HDTRA1-13-1-0026.
- National Science Foundation under award 1320347.

This support does not constitute an express or implied endorsement on the part of the U.S. Government.

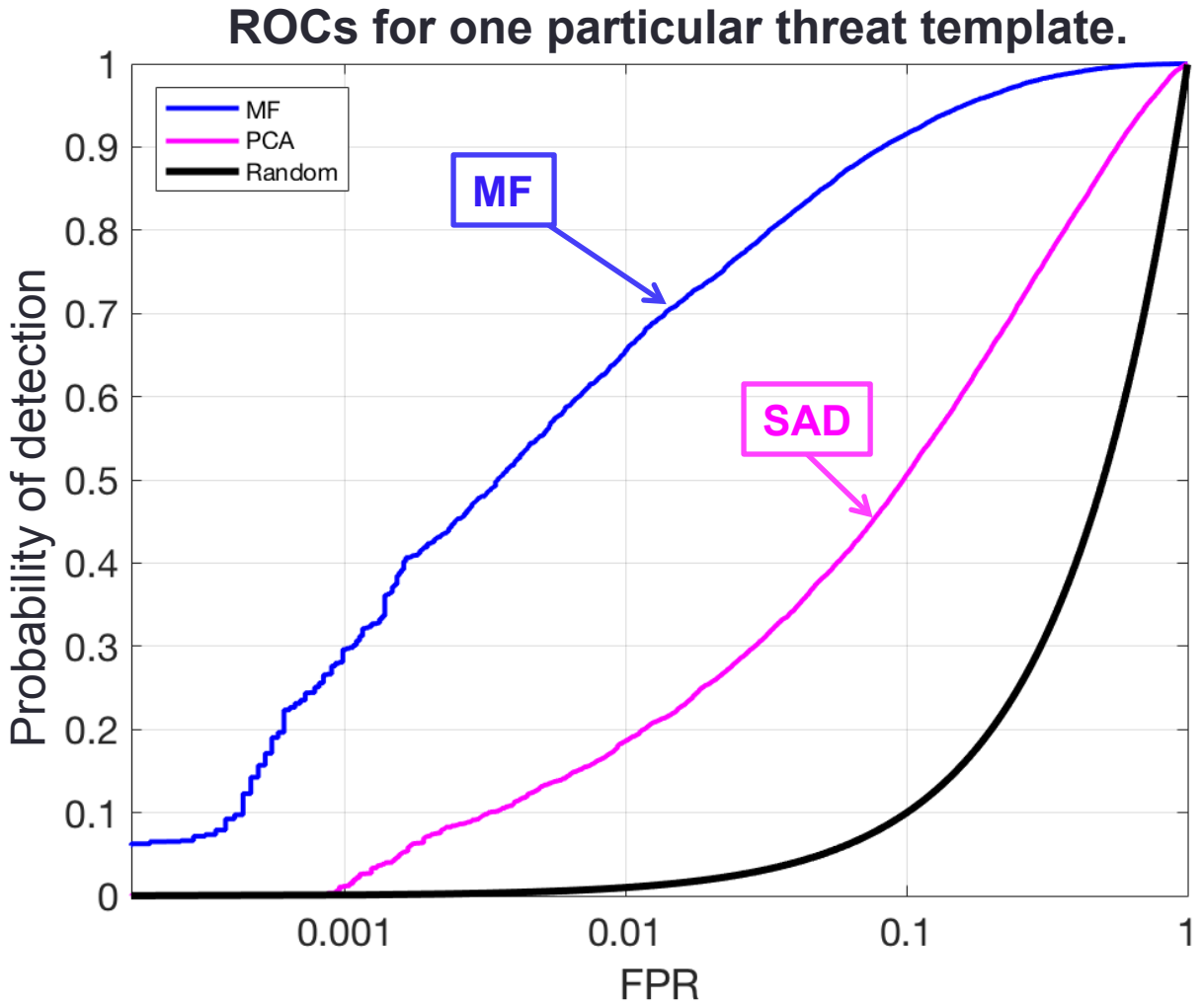
Radiation Threats in Spectral Data

- Our purpose is to detect compact sources of potentially harmful radiation in the presence of background noise.
- We analyze individual gamma-ray spectrometer measurements, some of which may reflect presence of the sources sought.
- We propose an algorithm that tolerates imperfect knowledge of source spectrum templates.



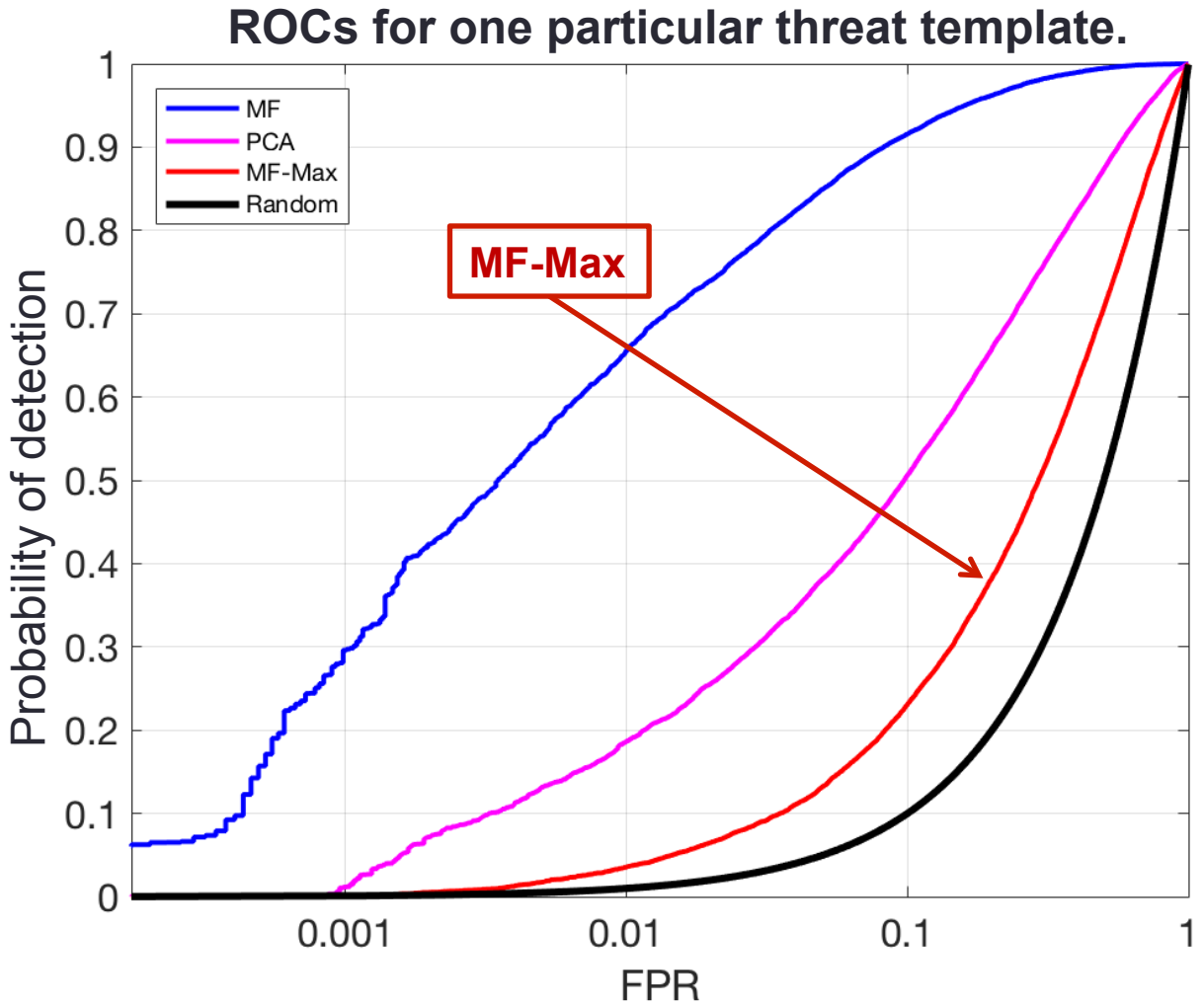
Alternative Circumstances of Threat Detection

- If we do not know what threat design to expect, we can use Spectral Anomaly Detection (SAD or PCA).
- If we have perfect knowledge of the shape of threat spectrum, we can use a Matched Filter (MF).
- In practice, we often have an idea of what threat to expect, but our knowledge of it is usually imperfect.



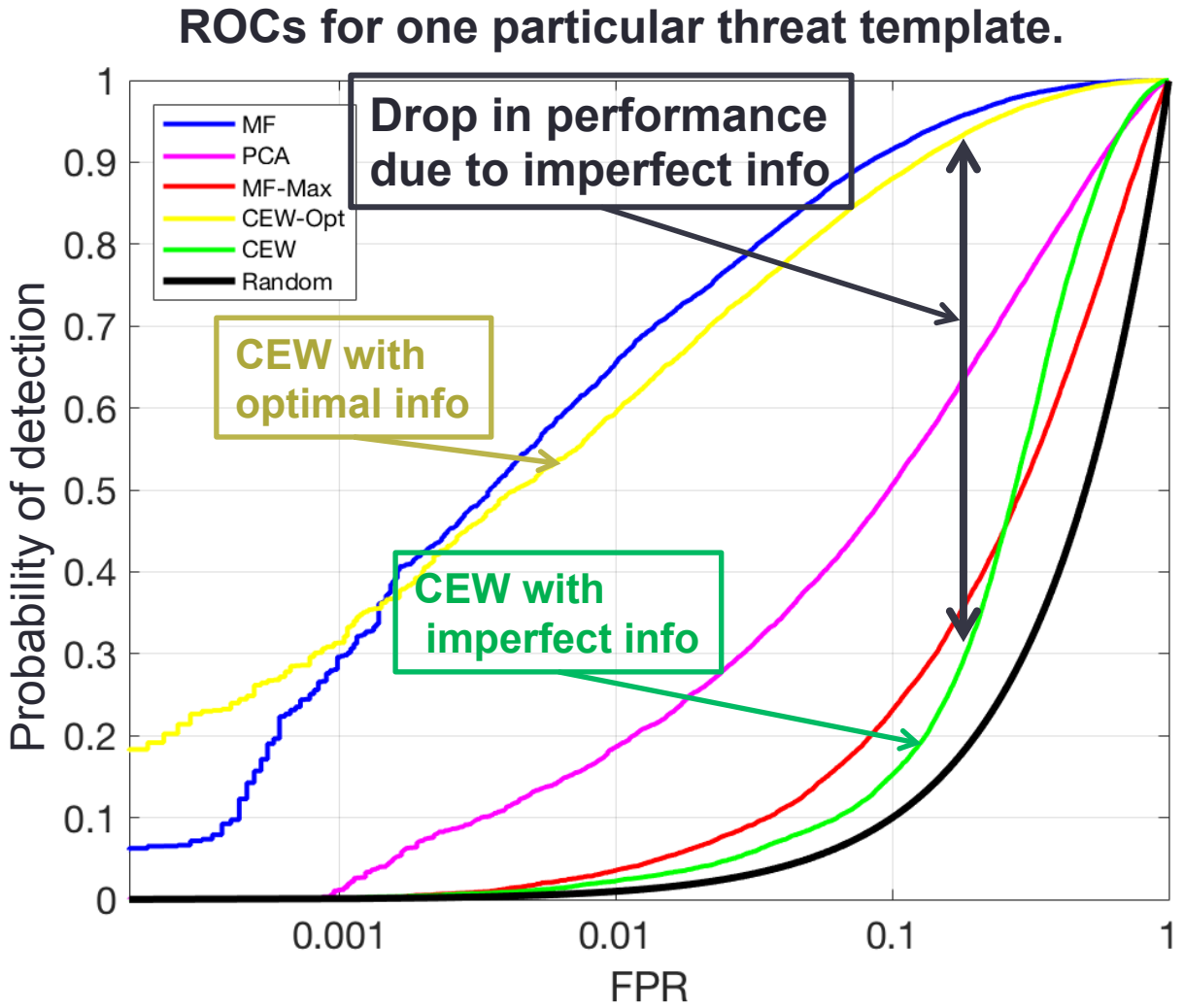
Alternative Circumstances of Threat Detection

- If we can predict the variety of possible threat templates and form a library of threat templates, we can use marginalized version of Matched Filter, i.e., MF-Max.
  - It would work as well as MF if marginalization always correctly picked the right threat template to use.



Alternative Circumstances of Threat Detection

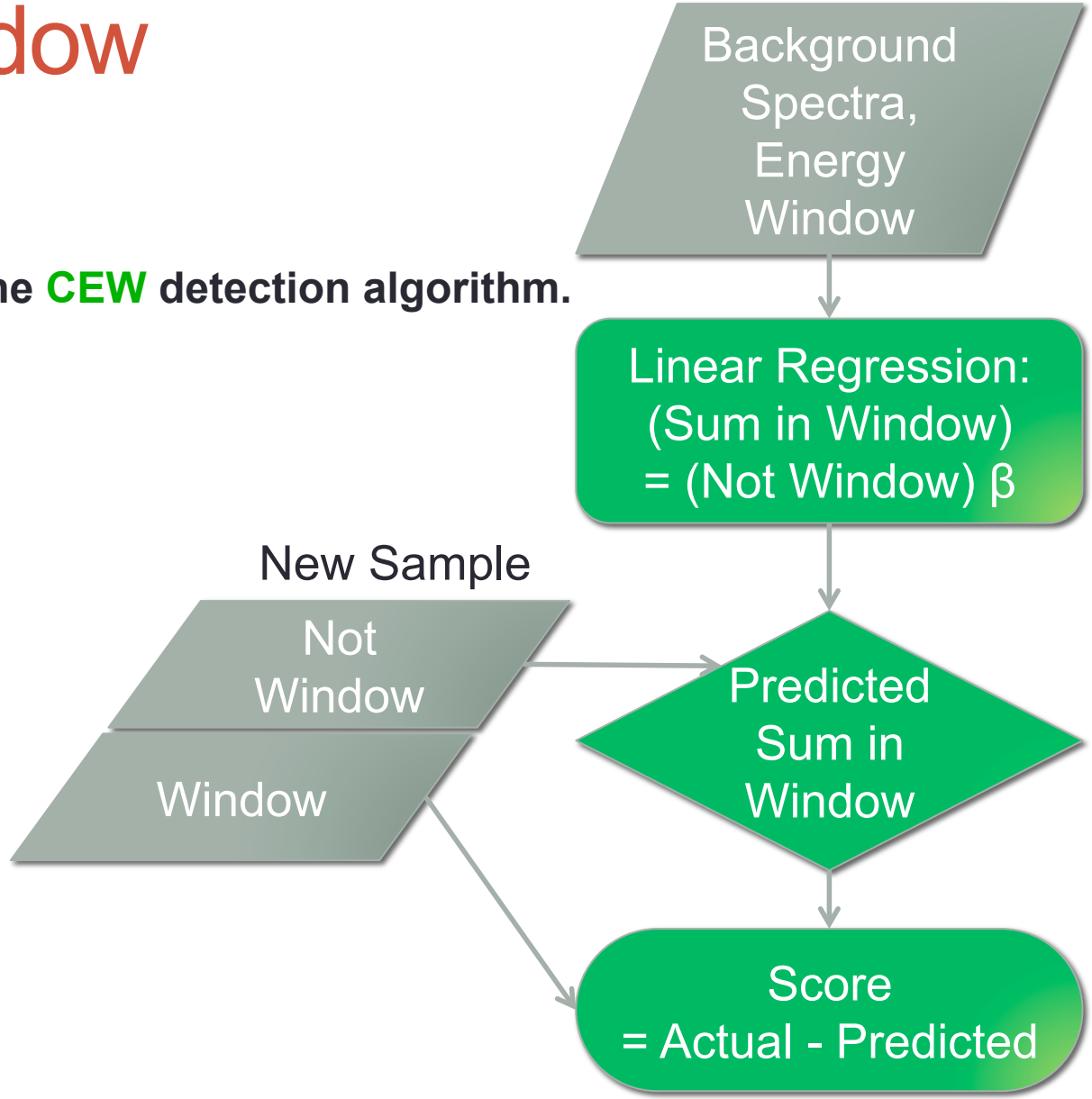
- If we do not know the exact shapes of a threat spectrum but know which ranges of energy it is most likely going to affect, we can use the Censored Energy Window (CEW) algorithm.



Censored Energy Window

- The **Censored Energy Window** for a threat given known distribution of background (threat-less) measurements:
  - Is the range of energy bins in which the threat is expected to be seen most clearly.
  - Can be computed from the known threat spectrum and estimated background distribution by maximizing expected SNR.
- The **CEW** detection algorithm predicts the sum of photon counts to be seen in the window using bins of counts outside the window as predictors.
  - Excess observed photons indicate threat.
  - Typically implemented using single-output multiple-input regression.
  - Performs well if the energy window is estimated accurately.

The CEW detection algorithm.



The Issue with Uncertain Energy Windows

- Threat spectra given to current methods are handpicked, but they could vary due to specifics of design and shielding.
- Consequently the expected energy window can be inaccurate.

Existing Method: CEW

- Sums counts in window.
- Single-output prediction.
- Often sensitive to window quality. Bad information can be worse than no information.

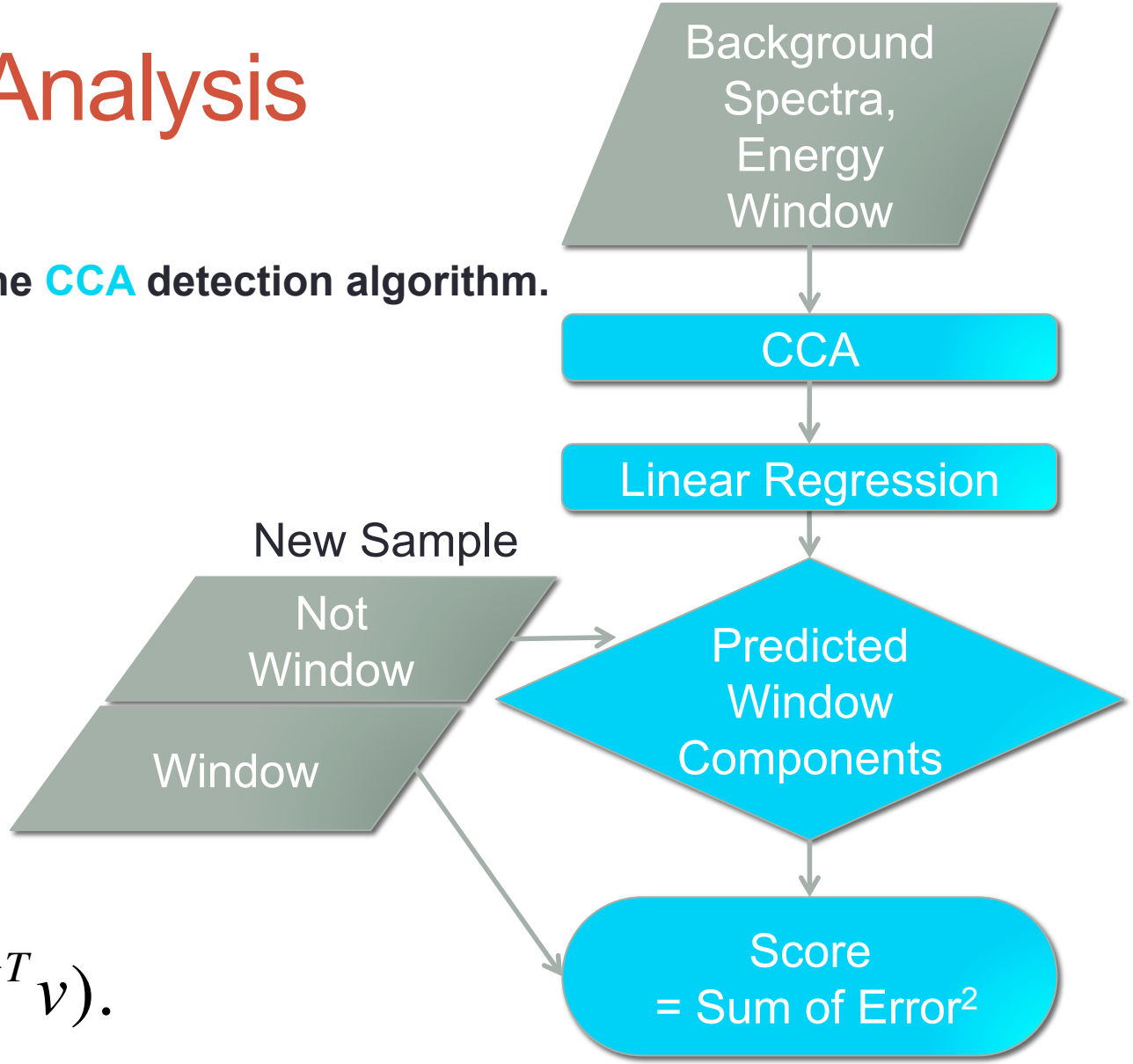
New Method: CCA Detection

- Uses full spectrum in window and predicts for all bins in it.
- Bridges the gap between SAD (using no threat information) and source-type-aware methods.
- More robust to imperfect knowledge of source spectra.

Canonical Correlation Analysis

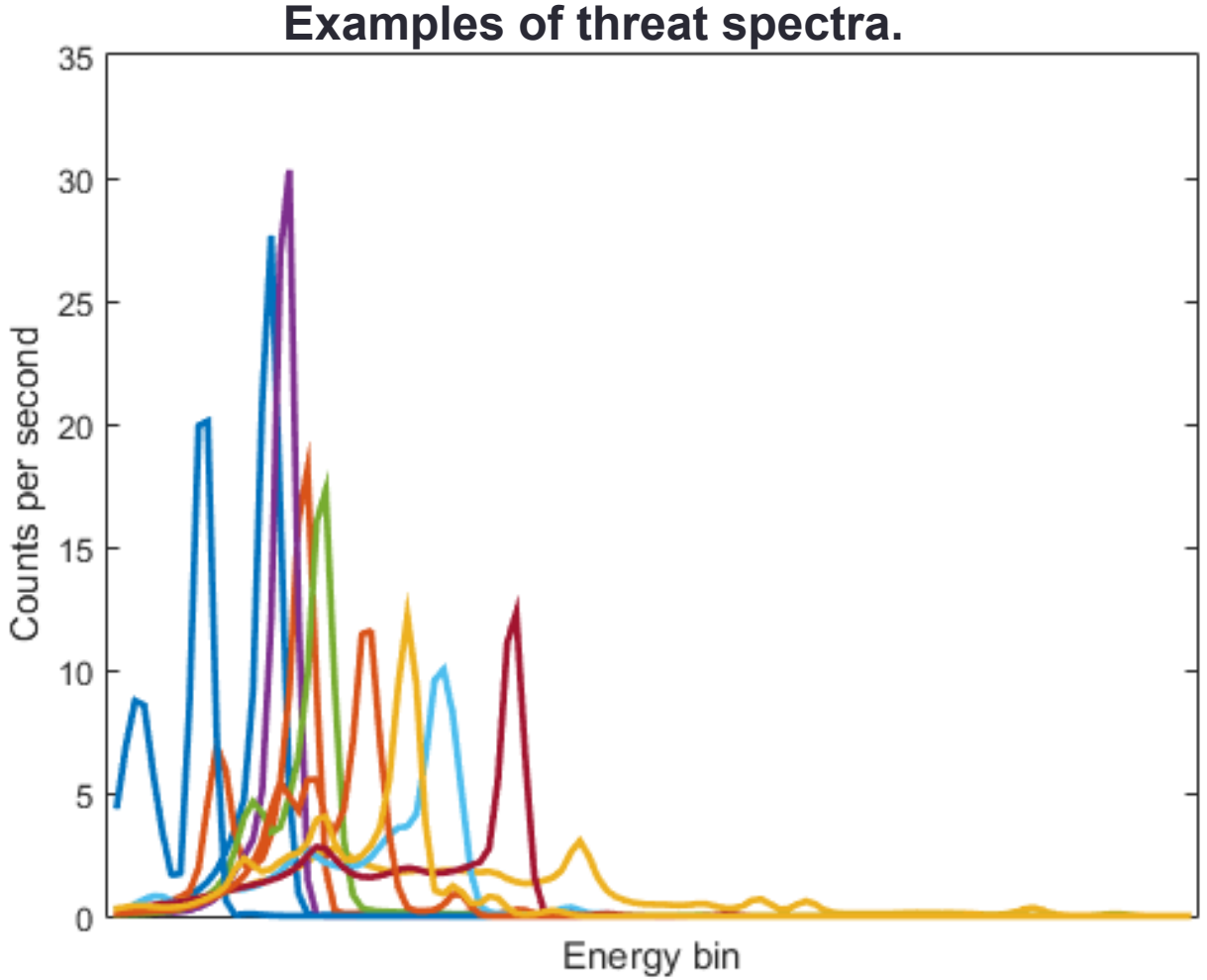
- **Canonical Correlation Analysis (CCA)** is a well-known statistical method for finding structured correlations between two sets of variables, *X* and *Y*.
- Here, *X* and *Y* are photon counts inside and outside the energy window.
- CCA solves $\max_{u,v} corr(X^T u, Y^T v)$.

The CCA detection algorithm.



Data Set and Experimental Setup

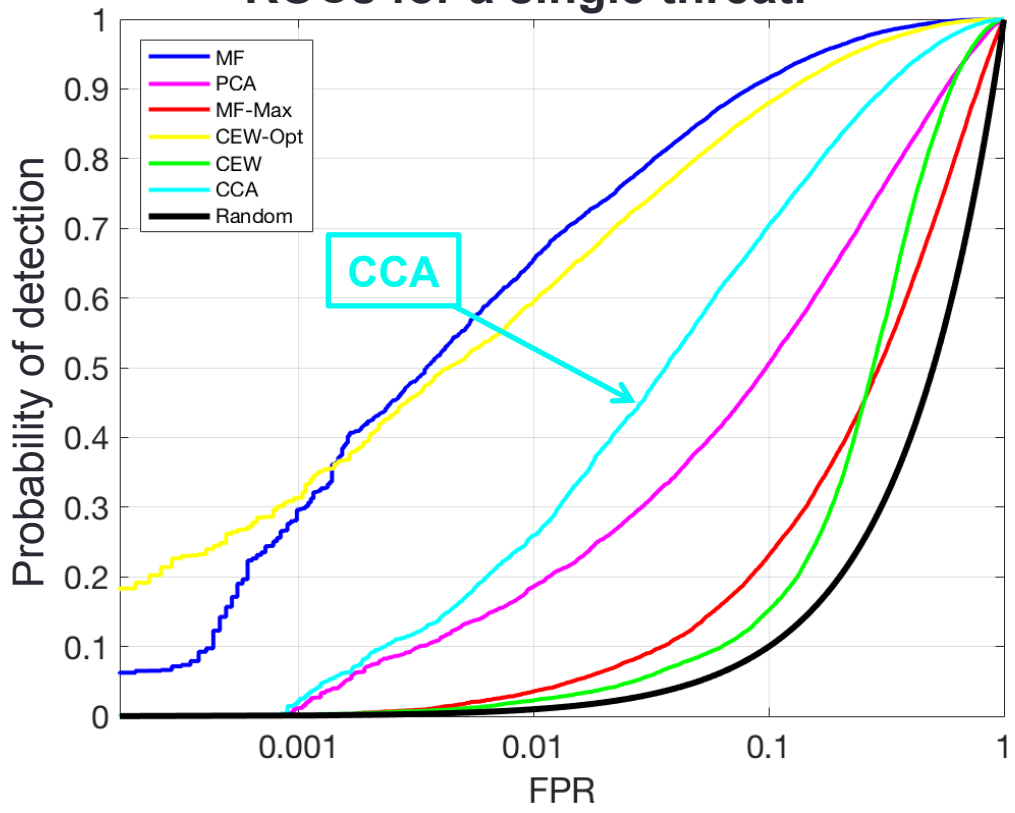
- 86,000 measurements:
  - Assumed to be background data.
- 67 threat templates:
  - Simulations of different configurations of material and shielding.
  - Signal-to-noise ratio of 2.
- Produce synthetic positive measurements by adding Poisson samples from the threat spectra to the background.
- Cross-validate by leaving a threat out of the training procedure and only using it in testing.



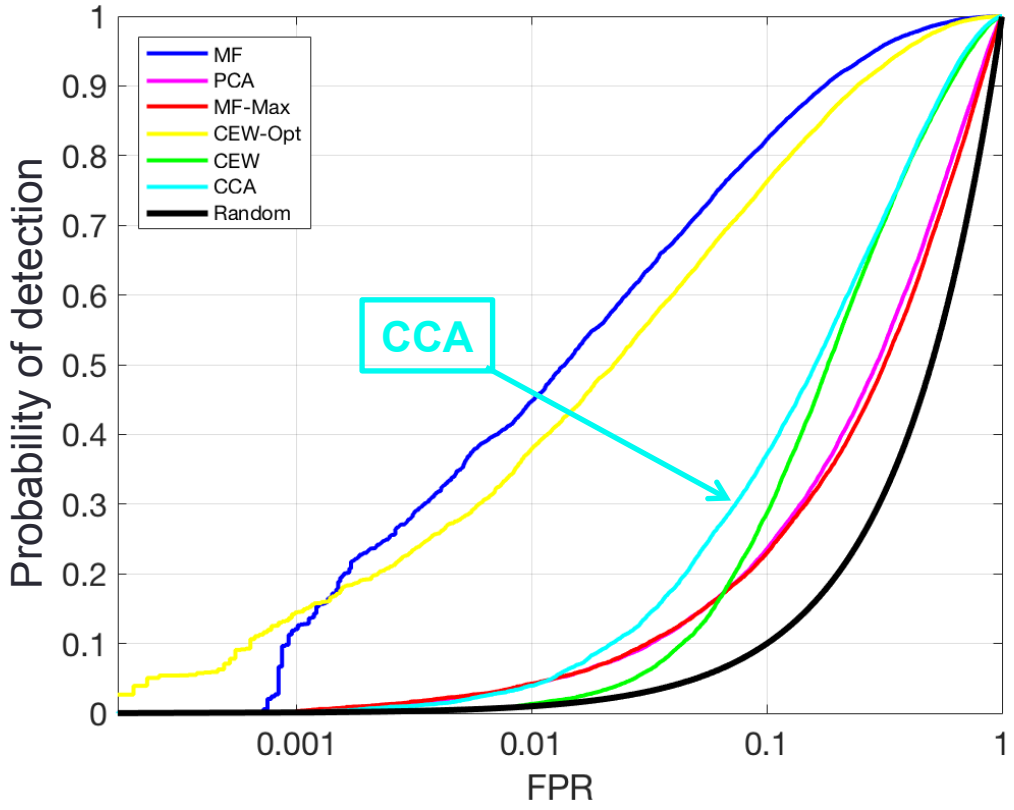
Simulations with Imperfect Information

- We compare MF-Max, CEW, and CCA where we marginalize over a threat library that does not contain the actual threat.
- Our CCA method yields improved performance closer to the optimal information case.

ROCs for a single threat.



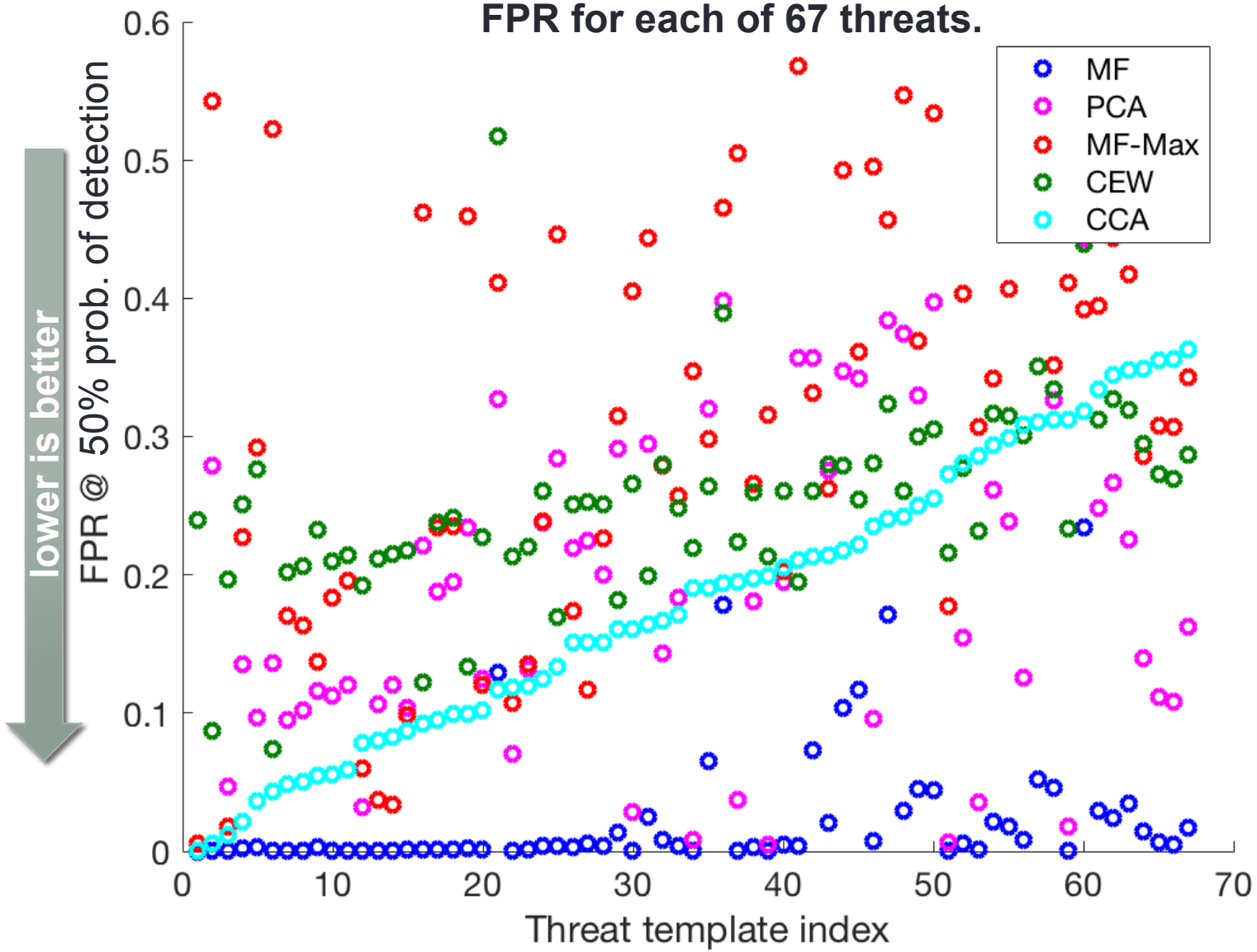
ROCs for a different threat.



Simulations with Imperfect Information

We sort all threat templates by their CCA performance to compare the methods across threats.

CCA usually performs better than CEW, MF-Max, and SAD.



Changing the Energy Window Quality

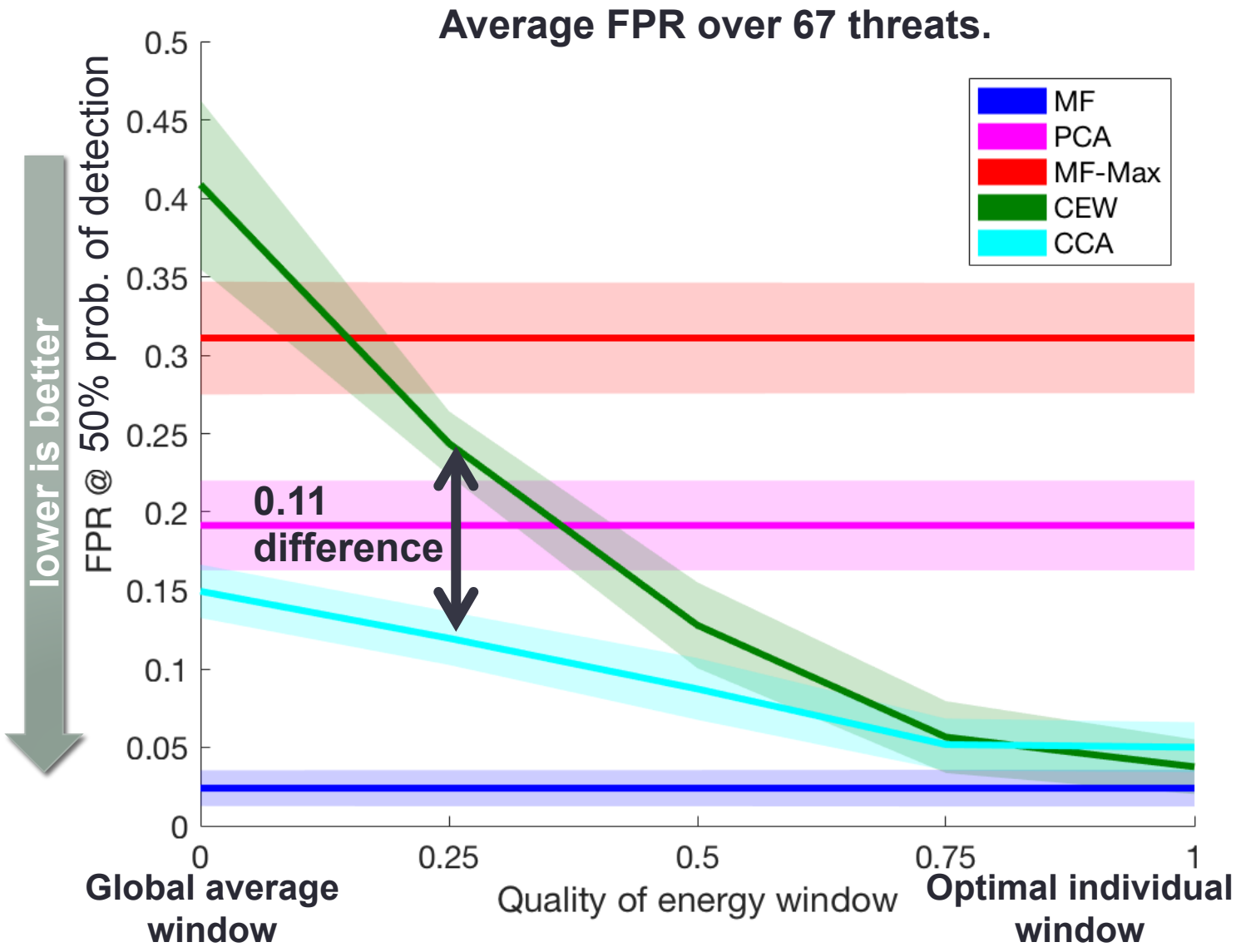
1. Compute global average threat template.
2. Compute convex combinations of average template and actual template.
3. Find energy windows of combination templates and pass to CEW and CCA.

Changing the Energy Window Quality

FPR for each method as the window changes from low-quality to optimal.

As information about the threat spectrum decreases, the performance of CEW degrades and becomes much worse than CCA.

(Other methods do not use a window).



Conclusion

- Perfect information about threats or energy windows may not be realistic in application scenarios.
- We proposed a method (Canonical Correlation Analysis) that remains sensitive when prior knowledge of the expected types of sources deteriorates.
- It bridges the gap between methods that do not require any source knowledge (Spectral Anomaly Detection) and those that work well when reliable prior knowledge is at hand (Matched Filter, Censored Energy Window).
- The CCA detection algorithm pays attention to details of spectra, while CEW aggregates counts in the source-type-specific energy window of interest.
- Our results can be useful in practical applications whenever designs of sources of harmful radiation are not precisely known.

References

Nelson, Karl, and Simon Labov. “Aggregation of Mobile Data.” Lawrence Livermore National Lab Technical Report 1.2.2 (2012): 2-3.

Hotelling, Harold. “Relations between two sets of variates.” Biometrika 28.3/4 (1936): 321-377.

Tandon, Prateek. “Bayesian Aggregation of Evidence for Detection and Characterization of Patterns in Multiple Noisy Observations.” No. CMU-RI-TR-15-23. Robotics Institute, Carnegie Mellon University (2015).

How to Compute the Optimal Energy Window

To compute the optimal energy window for each source, we rely on the following algorithm, which takes the source signature as input, represented as a vector of Poisson rates over energy bins. We compute the ratio of the rates to the mean background sample and sort the energy bins in descending order of this ratio. For each integer k from 1 to the number of bins, we obtain the sum S_k of the top k bins and the sum B_k of k corresponding mean background bins. We find $k^* = \arg\max_k S_k/\sqrt{B_k}$ and output as the energy window the top k^* bins according to the ratio of source rates to mean background.

$$SNR = sum(template) / \sqrt{sum(mean(background))}$$

CCA Detection Algorithm

- Training procedure:
  - ➢ Input: background spectra; and a target energy window.
  1. Apply CCA to find weights (u, v).
  2. For each pair (u, v) of weights,
     - i. Fit a linear regression of $X^{\mathsf{T}}u$ on $Y^{\mathsf{T}}v$.
     - ii. Compute residuals for all samples and fit a univariate normal distribution.
- Prediction procedure:
  - ➢ Input: new measured spectrum.
  1. For each pair (u, v) compute regression residuals and their z-scores.
  2. Output the sum of squared z-scores

Data Set Details

- 86,000 gamma-ray measurements (John Hopkins):
  - Photon counts over a spectrum partitioned into 120 energy bins, 2600 counts/sec average.
  - At one-second intervals.
  - Recorded by a sodium-iodide detector moving around an urban area (Baltimore).
  - Assumed to be background data.
- 67 threat templates (Lawrence Livermore National Laboratory):
  - Each threat a spectrum.
  - Threats are normalized at 100 counts/sec.
  - For each threat, create synthetic positive samples by drawing independent samples from the Poisson rates and adding them to the background data.
  - Templates are simulations of different configurations of material and shielding.
  - Clustered into 10 groups using *k*-means. For each threat, its cluster was excluded from known information.
  - SNR of 2.

TPR at fixed probability of detection