Applying matrix factorisation methods to improve specificity of fast anomaly detection algorithms

Caroline Pyke[1], Luke McGarry[1]

[1] National Nuclear Laboratory

Overview

This poster presents the application of the fast change point detection algorithm FOCuS [1] and a computationally efficient implementation of Sparse-Principal Component Analysis (SPCA) [2] to the SIGMA data. The purpose of this is to detect anomalies in the Gamma Counts data and attribute the detected anomalies to specific regions in the spectra through analysis of loadings. Application of FOCuS to the SIGMA gamma counts data was undertaken using Python [3] and the matrix factorisation of SIGMA gamma spectra data was completed with R [4].

Outline of Method

Step 1

Apply FOCuS algorithm to data and identify anomalies

Step 2

Determine regions of spectra responsible for anomalies using SPCA on spectra within local neighbourhood of anomaly

Step 3

Compare similarity between detected anomalies by analysing the loadings

Step 4

Analyse all anomalies found using non-negative matrix factorisation (NMF) and compare with SPCA *(ongoing)*.

Future work

Convert between spectrometer channel and energy bin and treat data as functional rather than finite dimensional. Will enable comparison across the detector types. Compare loadings with Gamma spectra templates of radioisotopes.

Results from steps 1 to 3

Detecting anomalies

FOCuS was applied to the observed gamma counts for a detector against a prediction of the background count from an exponential smooth of preceding gamma counts. Anomalies needed filtering from the data used to estimate the background otherwise they would influence the smooth of the data and reduce the limit of detection in the neighbourhood of large anomalies. This was done using a rolling window filter. An example of anomalies detected for a single detector and a single day are shown in Figure 1 and the comparison between the raw gamma counts and FOCuS response are given in Figure 2.

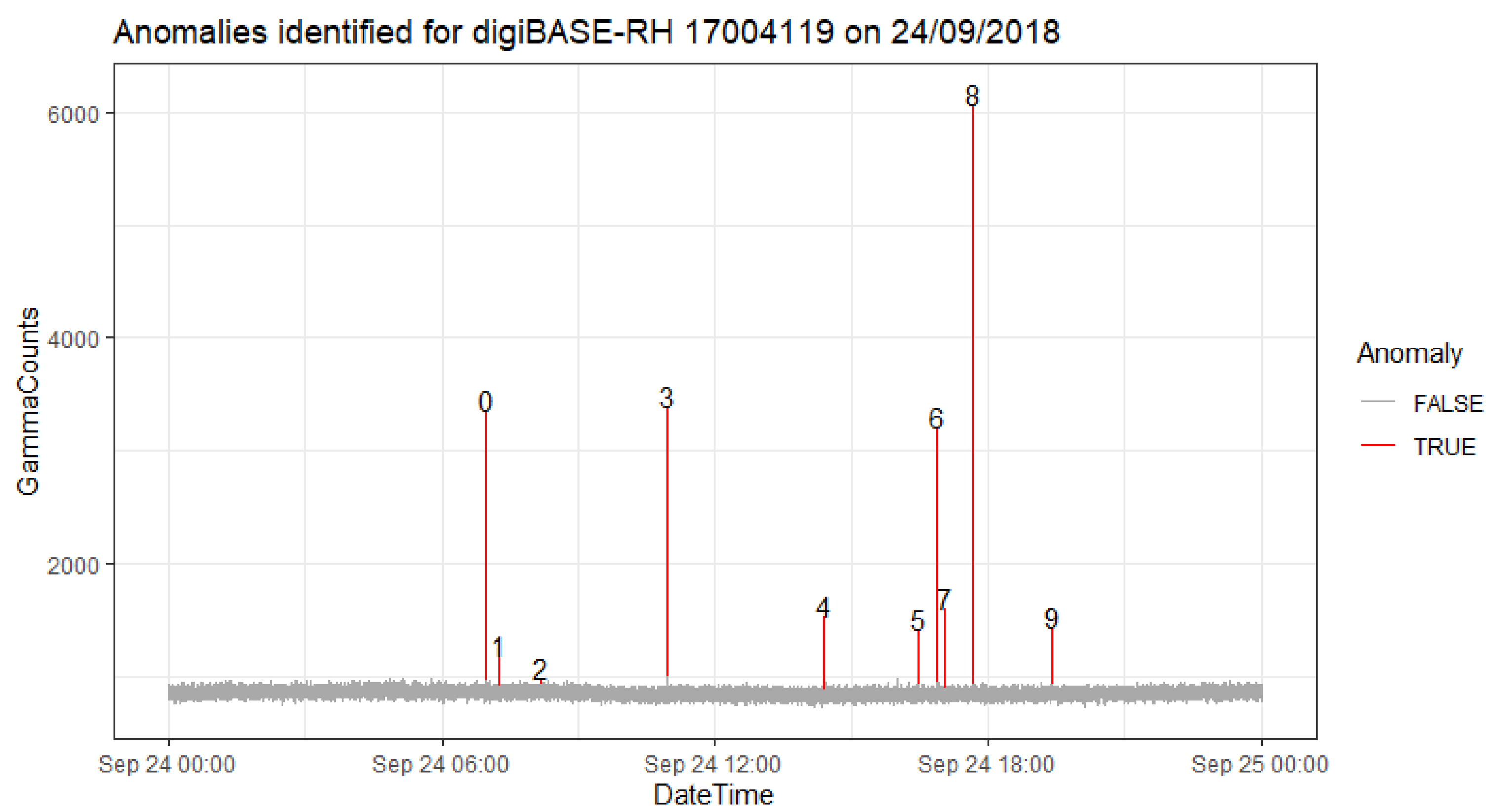

Figure 1 *All anomalies detected by FOCuS for detector digiBASE-RH 17004119 on date 24/09/2018.*

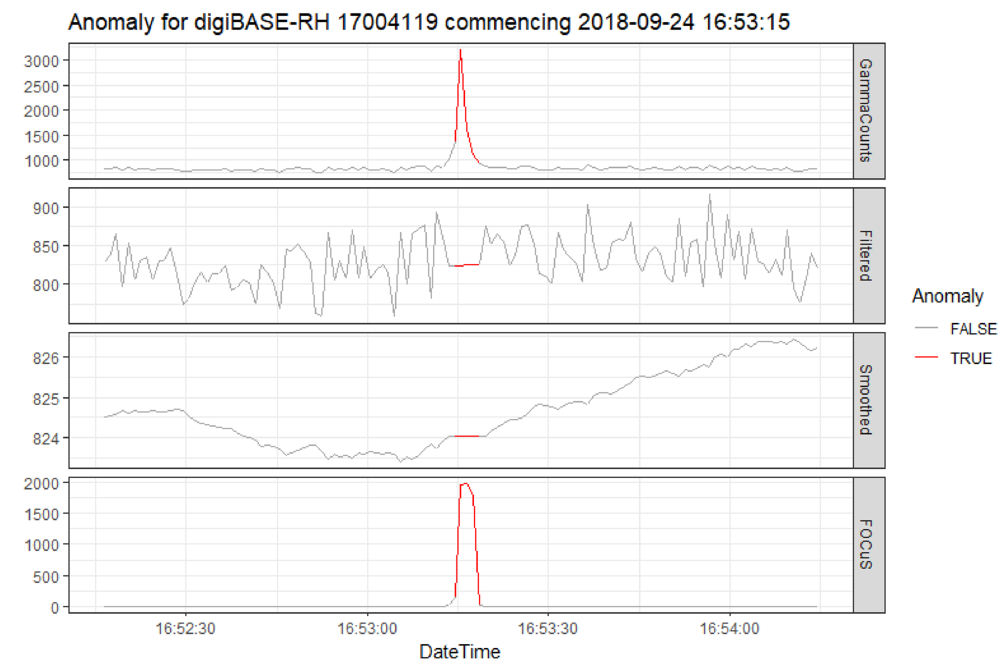

Figure 2 *Anomaly 6 from Figure 1. Shown are the raw gamma counts, the filtered gamma counts, smoothed gamma counts and the FOCuS response.*

SPCA of the gamma spectra

The Gamma spectra for all time-points for a FOCuS anomaly and an adjacent time-period of 2 minutes (containing background data) were analysed using SPCA. The spectra and SPCA loadings for data in this interval around an anomaly are shown in Figure 3.

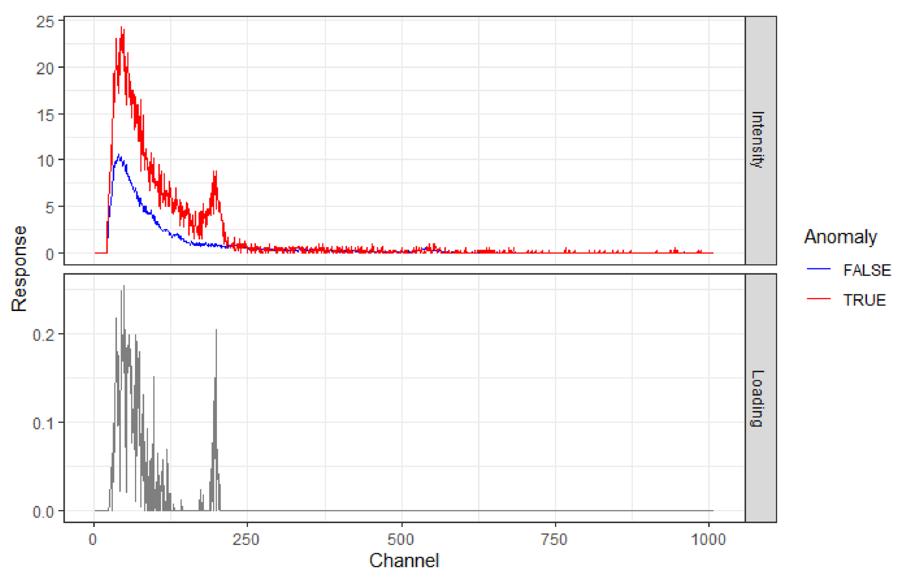

Figure 3 *Top row of figure shows the raw counts vs channel number. Red trace is the mean intensity of the anomaly, the blue trace is the mean intensity of the background. Bottom row shows the primary SPCA loadings where SPCA is penalized heavily to ensure a high degree of sparsity.*

For the example above, the interval of spectral data subject to SPCA resulted in a data matrix (*n,p*) of approximate dimension n = 130 and p = 1024. This was due to the anomalies during this time-period being relatively short lived lasting <10s. The primary principal component explained ~25% of the variance for most of the anomalies and the loadings were strictly positive upon correction for orientation. In Figure 3 the loadings are equivalent to the first column of the matrix B defined in the Appendix. Analysis of higher order loadings and scores determined that the first sparse principal component was sufficient to distinguish the anomaly from the background using the scores.

Comparison between the SPCA loadings

To determine if the detected anomalies were related the primary loadings were compared. The SPCA for each anomaly were subject to the same sparsity constraints and were approximately the same dimension data-matrices. The level of sparsity between the primary component varied approximately between 60 non-zero loadings and 170 non-zero loadings for each anomaly.

Visual analysis of the loadings for each anomaly suggested Anomaly 2 was a false-positive since the spectra was indistinguishable from the background signal albeit slightly elevated, uniformly across the channels. The loadings for all anomalies were compared using the cosine similarity metric (Figure 4). This showed a high level of similarity between all anomalies apart from anomaly 2, further reinforcing that this may be a false positive.

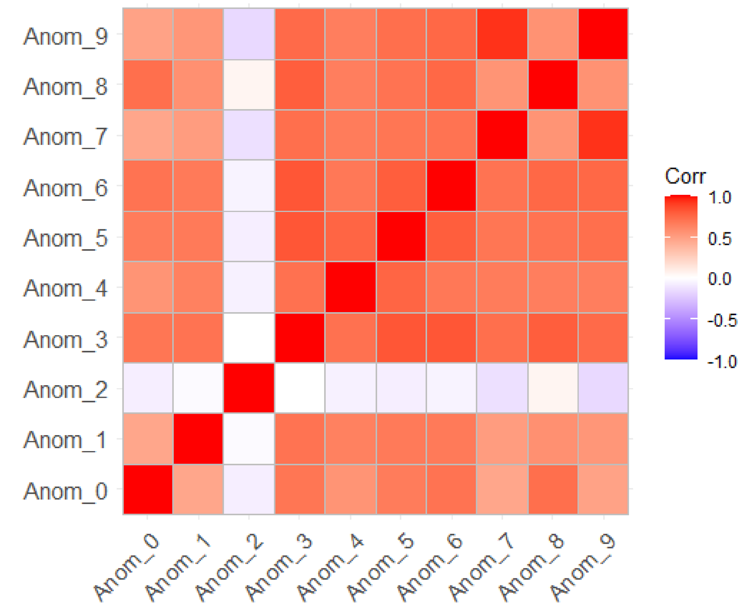

Figure 4 *Heatmap showing the cosine similarity between the loadings of all identified anomalies. No clear structure in the loadings was found for Anomaly 2 and this was subsequently deemed to not be an anomaly.*

Further methods to reduce false-positives

- SPCA was not sufficient to reduce the anomaly set alone. Work is ongoing using NMF to refine FOCuS further.
- It is expected that a move to functional data analysis will be required to compare signals between detectors
- Refining using expert judgements of standard radioisotope signals should also improve specificity.

Appendix

Sparse Principal component analysis

SPCA can be formulated as minimizing the following objective function for a given (n,p) data matrix X, orthogonal matrix A and sparse loadings matrix B:

$$f(A,B) = \frac{1}{2}\|X - XBA^{\top}\|_F^2 + \psi(B)$$

Here ψ denotes a sparsity inducing regularizer such as the LASSO or elastic net. The principal components Z are formed as follows

$$Z = XB$$

References

[1] K. Ward, G. Dilillo, I. Eckley & P. Fearnhead (2023) Poisson-FOCuS: An Efficient Online Method for Detecting Count Bursts with Application to Gamma Ray Burst Detection, Journal of the American Statistical Association, DOI: 10.1080/01621459.2023.2235059

[2] Sparse principal component analysis via variable projection, NB Erichson, P Zheng, K Manohar, SL Brunton, JN Kutz, AY Aravkin, SIAM Journal on Applied Mathematics 80 (2), 977-1002

[3] Python

[4] R statistical software

NATIONAL NUCLEAR LABORATORY