Anomaly Detection in SIGMA Data

Kes Ward
STOR-i Doctoral Training Centre,
Lancaster University, UK

Caroline Pyke, Luke McGarry
National Nuclear Laboratory, UK

Presented at the NUSEC Technical Workshop
October 9th, 2023

Abstract

The Poisson Functional Online Cumulative Sum (Poisson-FOCuS) method is a method for solving the likelihood ratio test of Poisson(λ) null against Poisson($\mu\lambda$) alternative where $\mu > 1$, i.e. searching for an increase in count. This can be thought of as equivalent to testing all possible anomaly start points $\tau \leq T$ at each timestep T, giving a computationally efficient way to analyse count anomalies that occur over intervals of time. We run the Poisson-FOCuS method on SIGMA data, with an additional adjustment to remove anomaly tail traces, and report the results.

1 Data description

The SIGMA data consists of gamma radiation (high energy photon impacts) on radiation detectors placed at different locations around London, UK. We want to design a system to monitor this data in real-time and search for threat profiles, for example a person smuggling a backpack of weapons-grade material with a Uranium-235 signature. At the same time, we want to be able to identify and discount anomalies in the data that are not threat profiles, such as a patient leaving a hospital after a radionuclide thyroid scan with an iodine-123 signature. The SIGMA data is from multiple different sensors, however as they are located far from each other it is assumed that any radiation threat profile would only show up in one sensor at a time.

The data exists in 4096 energy band bins. The file containing each day's worth of data is separated into approximately 55800 time bins, giving a sample rate of 1 bin per second.

There are approximately four months' worth of data, from 2018-08-06 until 2018-12-18 (135 days). This gives a total of approximately 31 billion total data bins (time by energy band) per sensor. In our analysis we will primarily consider data from sensor **D3 SGM101427**. All data plots are taken from the 6th and 14th August, 2018.

2 Problem setup

Our data signal $(\vec{x}_1, \vec{x}_2, ..., \vec{x}_T, ...)$ is a multivariate signal evolving through time. Each $\vec{x}_t := (x_t^1, ..., x_t^p)$ is a p-dimensional object, which represents the energy spectrum (for our data $p = 4096$).

We denote by T the present time, such that at time T only the signal $\vec{x}_t : t \leq T$ has been observed. We are interested in algorithms that perform well when $T \to \infty$, i.e. we have been observing a signal for a long time, or the signal is high-velocity. In the data we have, we assume batch processing each day's worth of data independently with $T \to 86400$ and this computation repeated up to 135 times to process the whole available dataset.

An anomaly with start time τ affecting some subset $P \subset \{1, ..., p\}$ of coordinates is such that for $t > \tau, i \in P$ there has been a change in the underlying process used to produce the measurements. We want to identify τ and P at the soonest possible point $T > \tau$ we are able to observe sufficient evidence. We are also only interested in anomalies where their length $h := T - \tau + 1$ is relatively short, e.g. $h \leq 300$ (up to five minutes). We are only interested in anomalous increases in count, rather than decreases.

We will consider each x_t^i to be the realisation of a random variable X_t^i. As radiation counts can be modelled well as a Poisson process, we will say that under our null hypothesis of no anomaly each $X_t^i \sim \text{Poisson}(\lambda^i)$. We can estimate the λ_t^i well using previous data, noting that in the absence of anomalies our data stream does not change much over time.

By additivity of Poisson processes, we have that

$$\sum_{i \in P} X_t^i \sim \text{Poisson}\left(\sum_{i \in P} \lambda^i\right).$$

Initially we will work with $P := \{1, ..., 4096\}$ the whole signal trace and define $x_t := \sum_{i=1}^{4096} x_t^i$ and $\lambda := \sum_{i=1}^{4096} \lambda^i$, noting that we can estimate $\lambda \approx 28$ although there are a few mild fluctuations in the data. In section 7, we will outline ways of defining subsets that may be useful for detecting the radiation signatures of different isotopes.

3 Theory and method

3.1 Likelihood ratio testing

In general our significance (how surprised we are) is a function only of what we expect to see, and what we actually see.

$$\text{significance} = f\ (\text{expected count}, \text{actual count})$$

When working with count data, we use Poisson random variables. Denoting the actual count x_t and the expected count λ, we have

$$\text{significance p-value} = \mathbb{P}(\text{Poisson}(\lambda) \geq x_t)$$

However, this can be computationally inefficient. Therefore we use Wilks theorem (Wilks, 1938) to approximate twice the Poisson log-likelihood ratio by a χ^2_1 random variable. This is a very accurate approximation for any appreciable value of λ, certainly so for our problem. We have that

$$\frac{(\text{significance sigma-value})^2}{2} = x_t \log\left(\frac{x_t}{\lambda}\right) + \lambda\left(\frac{x_t}{\lambda} - 1\right)$$

We use thresholds of 4.5 for a 3-sigma event, 12.5 for a 5-sigma event, etc. In general, a k-sigma event needs a threshold of $k^2/2$.

To find expected and actual counts for an interval $[\tau, T]$, you just add up the actual counts $\sum_{t=\tau}^{T} x_t$ and expected counts $\lambda(T - \tau + 1)$ for each time point in the interval, and use these in the above method to calculate your significance statistic. However, a signal of length T generates $T + 1$ new intervals when a new point is added. Even if we only check the final h intervals, this can be computationally costly.

3.2 Page and FOCuS

The Functional Online Cumulative Sum (FOCuS) method (Ward et al., 2023) is a quick method for solving the likelihood ratio test of Poisson(λ) null against Poisson($\mu\lambda$) alternative where $\mu > 1$, i.e. searching for an increase in count. This can be thought of as equivalent to testing all possible anomaly start points $\tau \leq T$ at each timestep T. Imposing a constraint of maximum length $h_{\max}$ of anomaly is equivalent to imposing a constraint on minimum intensity $\mu_{\min}$, as less intense anomalies are only detectable over longer timescales. This is linked to the sigma significance k and the background rate λ as follows:

$$\mu_{\min} \log(\mu_{\min}) - (\mu_{\min} - 1) = \frac{k^2}{2h_{\max}\lambda}.$$

For a statistical threshold of $k = 5$ (a "five-sigma event"), a background rate $\lambda \approx 28$ and $h_{\max} = 300$ (five minutes) this solves to give $\mu_{\min} \approx 1.055$, i.e. an anomaly of average magnitude less than 5.5% of the background level is statistically undetectable on this timescale. Anomalies that only occur over shorter timescales will have to be of greater magnitude in order to be detectable: for a rough estimate see Table 1. In particular, an anomaly present in just one time bin would have to more than equal (108.4%) the background radiation rate in order to be statistically detectable. By considering intervals

rather than points, we are able to substantially improve on this power and detect the presence of less intense anomalies.

maximum time	$h_{\max}$	$\mu_{\min}$	relative magnitude	absolute magnitude
5 minutes	300	1.055	5.5%	1.54 counts/sec
1 minute	60	1.124	12.4%	3.48 counts/sec
10 seconds	10	1.313	31.3%	8.77 counts/sec
1 second	1	2.084	108.4%	30.36 counts/sec

Table 1: How large an anomaly needs to be, as a relative proportion of the background signal and as an absolute size assuming $\lambda = 28$, to be detected over different timescales at a 5-sigma threshold.

In order to search for anomalies with a length exactly $h_{\max}$, we could use an iterated form of the Page-CUSUM statistic (Page, 1955; Lucas, 1985) for Poisson data $S_T(\mu_{\min})$, defined as follows:

$$S_0(\mu_{\min}) = 0, \quad S_T(\mu_{\min}) = [S_{T-1}(\mu_{\min}) + x_T \log(\mu_{\min}) - \lambda(\mu_{\min} - 1)]^+$$

Here, the notation $[]^+$ is used to denote the maximum of the term in brackets and zero. The last time $\tau \leq T$ that $S_{\tau-1}(\mu_{\min})$ was zero is the estimated start point for any anomaly. When $S_T(\mu_{\min})$ resets to zero, this indicates that it is more likely that no anomaly is present than an anomaly of intensity $\mu_{\min}$ is. Under a null hypothesis of no anomaly present, this should occur frequently, moreso the larger $\mu_{\min}$ is. Values of $S_T(\mu_{\min}) \geq k^2/2$ indicate a sigma significance k over the interval $[\tau, T]$.

All anomalies that would be picked up by running a window of size $h_{\max}$ over the data will be picked up by using $S_T(\mu_{\min})$. One advantage of using $S_T(\mu_{\min})$ over the use of a window of size $h_{\max}$ is that the method is one-scan: it does not need the separate storage of points inside the window that is required to remove the $T - h_{\max}$th point at time T. This makes it fast. One disadvantage of this algorithm is that it is not well-targeted for the detection of anomalies of $\mu > \mu_{\min}$, i.e. shorter, more intense anomalies.

The FOCuS method calculates $\max_\mu S_T(\mu) : \mu \geq \mu_{\min}$ in a similar timescale and is therefore able to efficiently detect anomalies of all sizes $h \leq h_{\max}$.

For both Page's method and FOCuS, anomalous *intervals* $[\tau, T]$ in the raw signal correspond to anomalous *points* in the significance trace $S_T(\mu)$. This means that it's a lot easier to identify interesting intervals in the significance trace than in the raw signal. For an example see Figure 1.

Because FOCuS can be thought of as an expanded form of Page's method, a similar set of intuitions apply, for example:

- For a given $\mu_{\min}$, the FOCuS trace for that signal must be at least the Page trace for that signal.

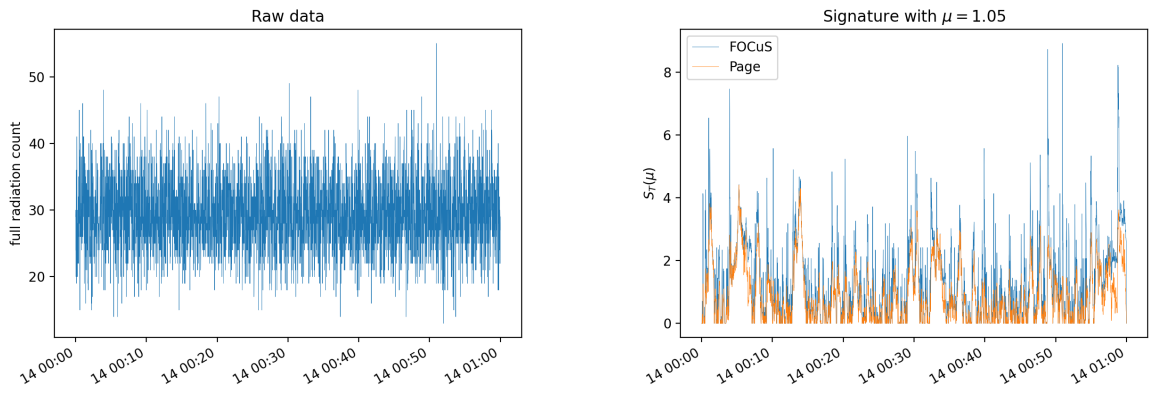

Figure 1: The first hour of data on the 14th August 2018, as raw data and as siginifance trace.

- The FOCuS trace tends to be fuzzier at low significance levels as it captures the normal fluctuations associated with individual time bins $h = 1$ and smaller intervals $h << h_{\max}$.
- Reading from the signature traces, the estimated start point for an anomaly giving a FOCuS trace at a high significance is approximately where the FOCuS trace started climbing out of this fuzzy state. (The actual start point is available within the algorithm).

4 Dealing with background fluctuations

The first six hours of data of 14th August, 2018 contain a slight deviation above baseline background rate, but not enough to be considered anomalous. This demonstrates the necessity of choosing an appropriate $\mu_{\min} > 1$. Too small, and the background fluctuation is captured in the signature trace, as in Figure 2. Here, a choice of $\mu_{\min} = 1.015$ means that the slight deviation above baseline is recorded as a six hour long anomaly.

Even with $\mu_{\min} = 1.05$ we can see that the signature we receive differs from what we would expect from a simulation of independent Poisson random variables (shown in the right of Figure 3). There are periods of up to half an hour where there is a slight deviation above baseline. However because it no longer builds over long stretches of time, the statistical significance of this is not strong enough to trouble us.

It should be noted that an alternate way to handle this problem would be assuming a nonconstant λ and performing a rolling estimate of the background rate λ_t, using e.g. a sliding window or exponential smoothing. This may be needed if we were interested in detecting longer anomalies and distinguishing them from background. However bias in the

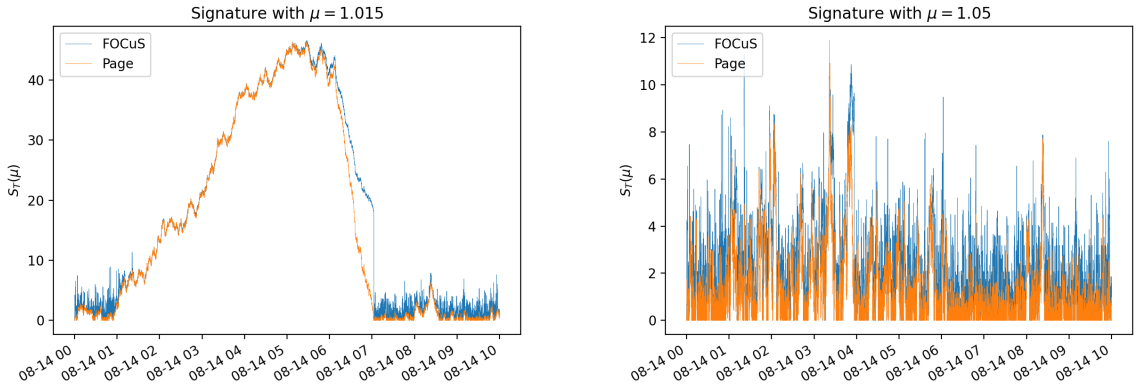

Figure 2: Comparing two different values of $\mu_{\min}$ for their ability to filter out small, long background fluctuations.

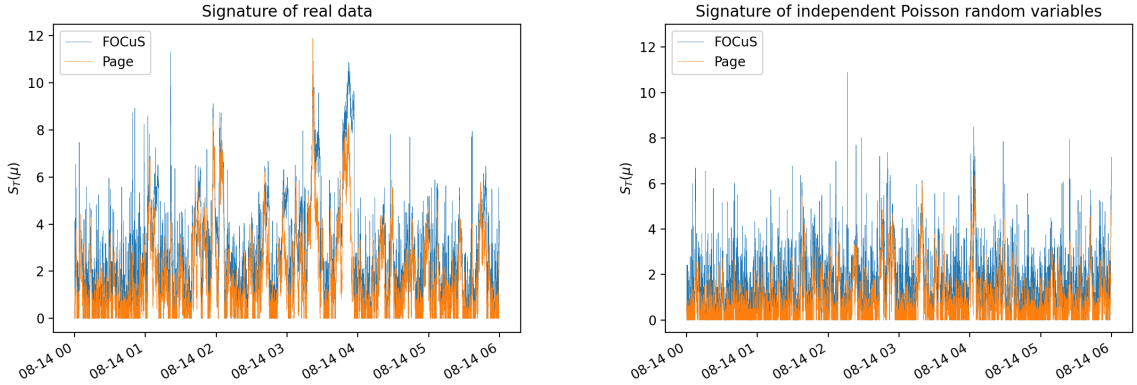

Figure 3: The difference between the data containing a small upwards fluctuation and independent Poisson random variables.

background rate estimate can be difficult to remove and in either case a choice of $\mu_{\min} > 1$ will help to mitigate this bias.

5 Resetting after large anomalies

From 11:10 to 11:15 on 14th August 2018, there is a large radiation anomaly clearly visible in the raw signal, as shown in Figure 4.

Large anomalies can leave traces in our significance trace long after they have ended. See Figure 5 for the effect of this large anomaly on our significance traces. Page's method tends to drop linearly and FOCuS polynomially (from a higher starting point), but both

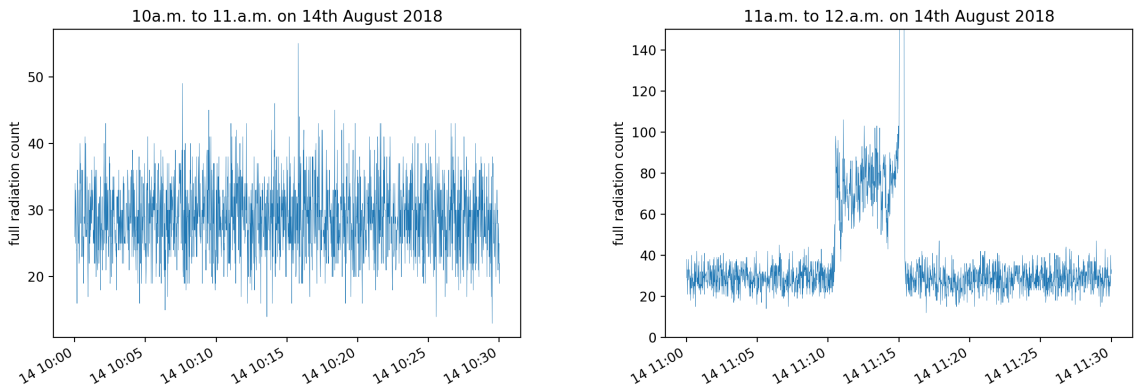

Figure 4: Two half-hours of data from 14th August 2018.

leave the same length of tail, which is in this case approximately five hours.

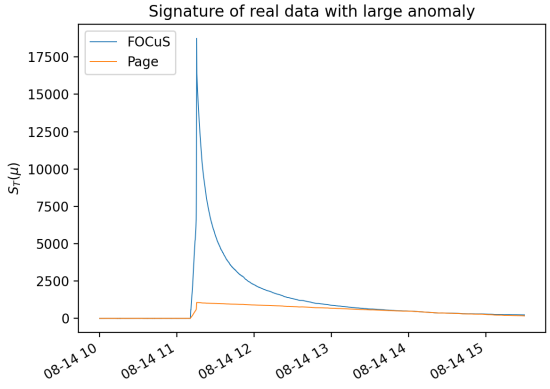

Figure 5: The signature of the large anomaly shown in Figure 4.

In order to get around the tail trailing behaviour, we institute a parameter h_{clear} that will remove any start points further in the past than h_{clear} if in the interval $[T - h_{\text{clear}}, T]$ contains no possible start points for an anomaly of size at least $\mu_{\min}$. This says that whatever anomaly is present is deemed to have ended, and should no longer be recorded in the current signal.

There is no positive evidence for an anomaly of intensity $\mu_{\min}$ beginning anywhere in the interval $[T - h_{\text{clear}}, T]$ and ending at T precisely when

$$S_T(\mu_{\min}) = \min_{t \in [T - h_{\text{clear}}, T]} S_t(\mu_{\min})$$

This can be calculated quickly using the ascending minima algorithm (Harter, 2009)

with computational cost not dependent on h_{clear}.

Because we know that positive evidence for an anomaly of size $\mu > \mu_{\min}$ on an interval necessitates positive evidence for an anomaly of size $\mu = \mu_{\min}$ on that interval, we can calculate Page's statistic and use it to reset FOCuS using the same condition. However, it can be more advantageous to reset FOCuS using the signal output from FOCuS directly, that is if

$$\max_{\mu>\mu_{\min}} S_T(\mu) = \min_{t\in[T-h_{\text{clear}},T]} \max_{\mu>\mu_{\min}} S_t(\mu).$$

In most cases this will be very similar to resetting using Page's statistic. Where it differs is when passing over a large anomaly, the FOCuS statistic will drop more quickly and will continue to drop even if a small anomaly continues to be present. However, because the FOCuS algorithm would either way not store the start point corresponding to the small anomaly, resetting this way can be preferable.

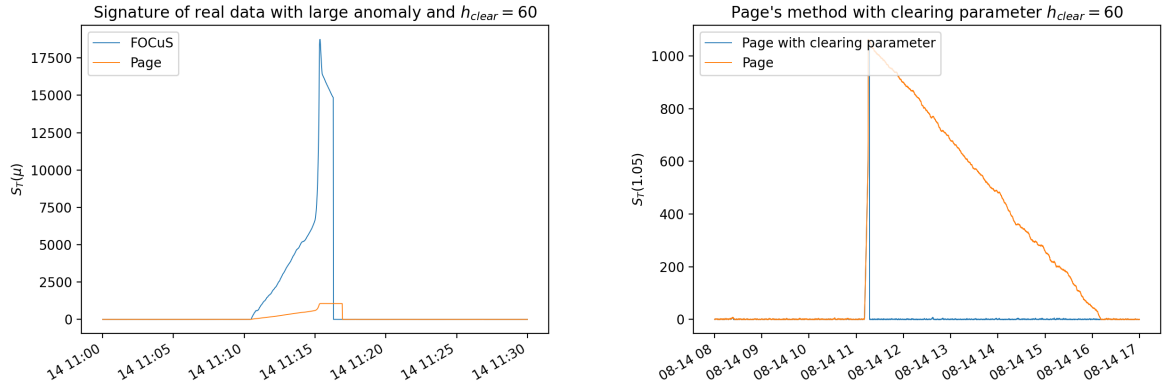

Figure 6: By using a clearing parameter h_{clear}, we can reset our method after large anomalies have ended.

Figure 6 shows the effect of this resetting strategy. Arbitrarily choosing $h_{\text{clear}} = 60$, we find that after passing over a large anomaly the algorithm resets within about a minute. Page takes a little longer than FOCuS as it takes a little longer to be sure of no evidence at all within the last minute, but even Page resets quickly compared to the five-hour tail of without a clearing parameter (see Figure 6).

6 Finding a threat

Figure 7 gives an artificially generated example of the kind of threat profile we would like our algorithm to be able to detect. A threat source approaches the detector, stops, and then leaves, over the total course of three minutes. With a radiation count mean of up to 3

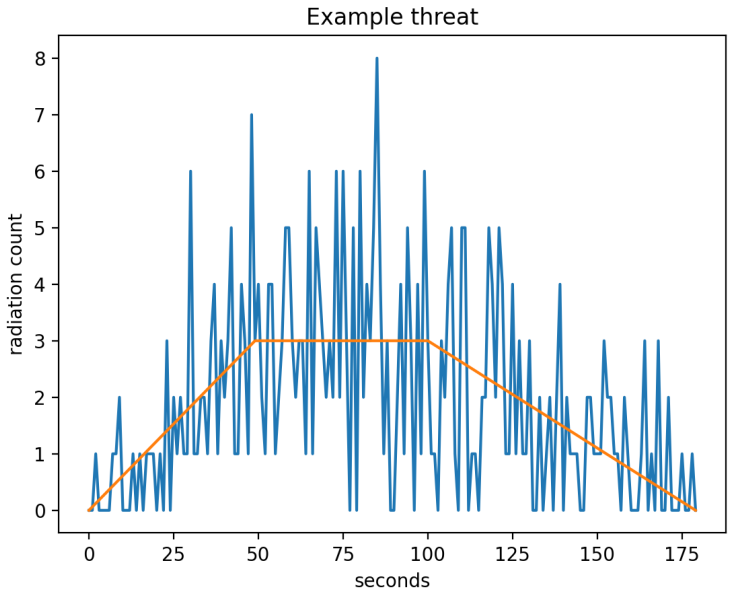

Figure 7: An example threat with intensity $\mu \approx 1.1$.

counts/sec compared to a background count mean of 28 counts/sec, this means it is barely visible to the naked eye when added to the SIGMA data (see Figure 8). Here our threat is about 10% of the size of the background signal, giving $\mu \approx 1.1$.

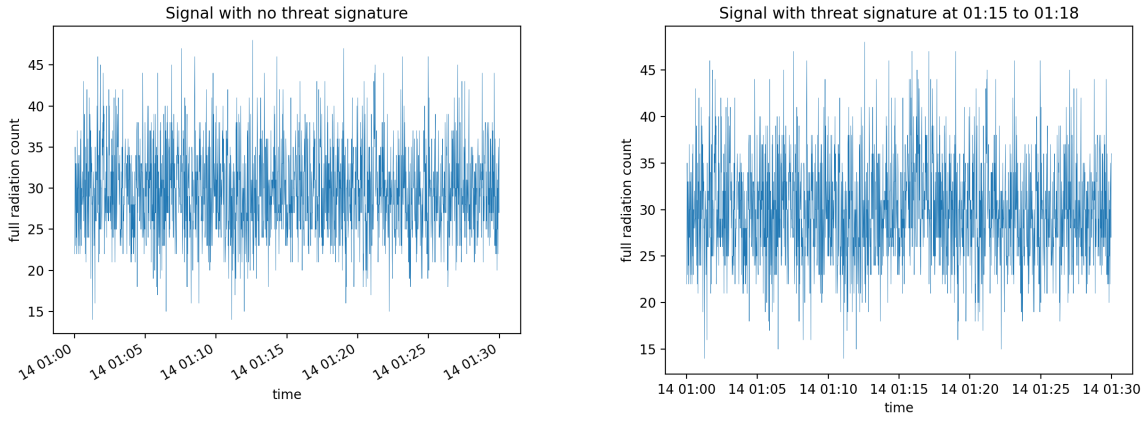

Figure 8: The example three minute threat incorporated into the SIGMA data at 01:15 to 01:18, barely visible to the naked eye.

The threat is detectable by FOCuS as shown in Figure 9 and is clearly visible in the significance trace. It crosses the 7-sigma significance line. The advantage over Page's method is also apparent: as FOCuS correctly targets the intensity of the anomaly it records

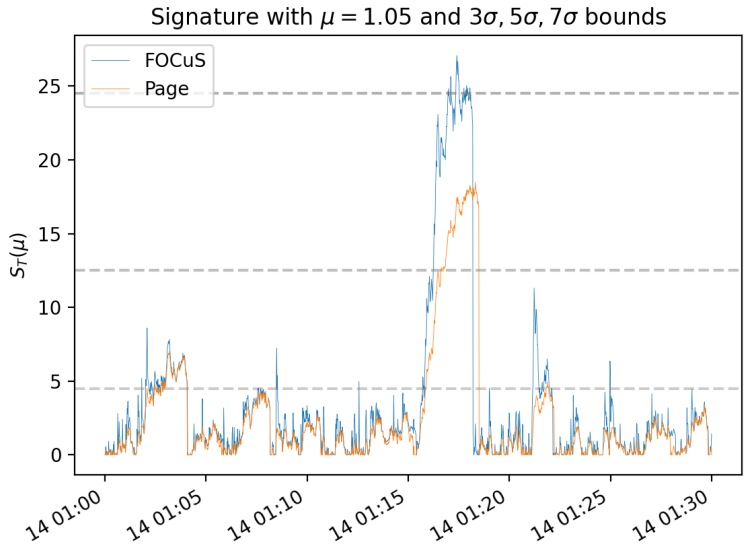

Figure 9: The signature trace of the threat with dashed lines showing 3σ, 5σ and 7σ significance levels.

a higher significance. The location of the anomalous interval can be easily read from the graph as from wherever FOCuS began to increase to where it attained its maximum value: here from 01:15 to 01:18.

7 Future work

Future work should incorporate the 4096 energy bands. Here, the average radiation count varies greatly by energy band. The raw energy band counts for one hour's worth of data on the 6th August 2018 are shown in Figure 10. The 4096 bands are represented individually on the left, and are grouped into logarithmic multiples of $2^{1/8}$ on the right to give a spectral graph that's easier to read. Most of the background radiation count occurs in energy band 100 to 1000.

The idea is that each of these energy bands may provide a different utility for identifying a threat, depending on the relative mix of background and anomalous radiation held by the band. The bands with the most utility for finding anomalies would have low background radiation, but a high amount of anomalous radiation if an anomaly was present. For an example of this, see Figure 11 of the log energy spectrum of a large anomaly on the 6th August 2018 plotted against nearby background traces. Note that although the anomaly is numerically greatest in the centre, it has a greater ratio of anomaly to background rate in the spike off to the right.

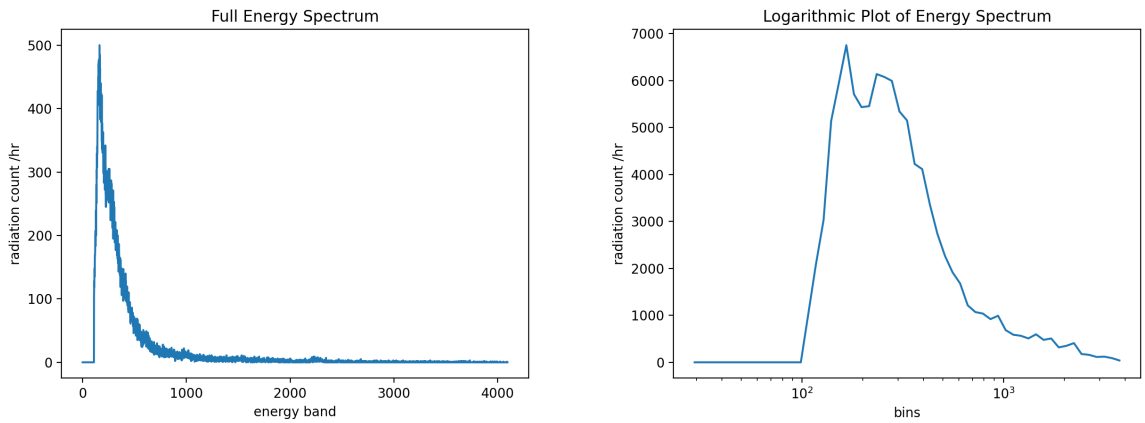

Figure 10: Energy band counts for one hour's worth of data

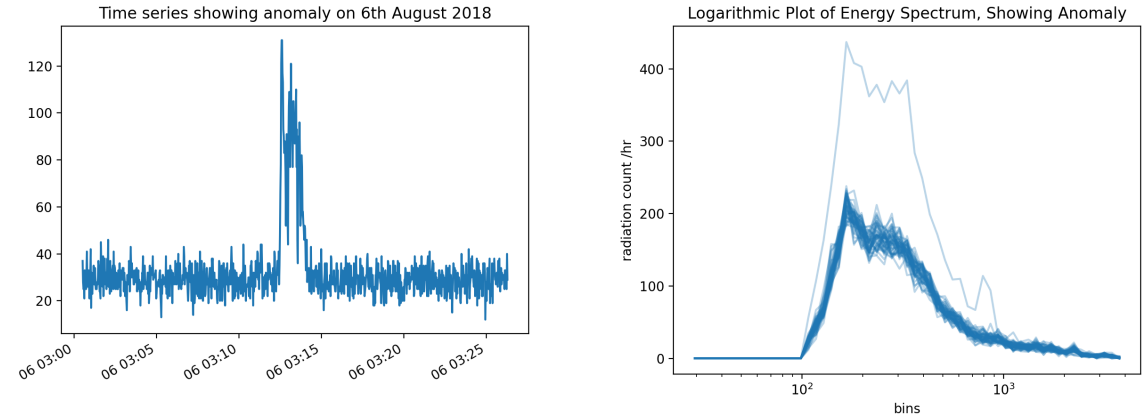

Figure 11: Graph showing the time and spectral structure of a large anomaly present in the SIGMA data on 6th August 2018.

Previously we have been using all the bands and combining them into one (by summing up counts) before calculating significance. Other options include:

1. Only use one subset of the bands with the most utility and combine before calculating significance.

2. Use multiple distinct subsets of bands, adding significances across bands.

3. Use multiple increasing subsets of bands that contain each other, taking the maximum significance across bands.

7.1 Utility criteria

Suppose that we have a subset of the signal included in our anomaly detection algorithm and we are looking into whether it would make sense to expand it.

Already included elements have in total anomaly amount A and background amount B. We are looking to see if it increases overall significance to add an energy band with anomaly amount a and background amount b. Working on the assumption that a and b are small compared to A and B, it can be shown that this binary yes/no question is a function only of the ratio a/b. For example, consider the alternate question of whether to add a category with anomaly amount $2a$ and background amount $2b$. If we assume the amounts are sufficiently small to be able to discount second-order effects, then the answer to both questions should be the same.

A good ranking measure of utility of a spectral band should be the ratio between the (normalised) rate a of the threat profile in that band and the background rate b in that band. We want to include all and only spectral bands with utilities above a set threshold.

Exactly how this utility threshold should be chosen for different threat profiles is not clear and may vary according to our desired false positive rate, so we may wish to track multiple increasing subsets. This difficulty may be complicated by the fact we do not know exactly the radiation signature of the threats we are hoping to find. This is because, for example, a sample of Uranium-235 will contain both the signature of U-235 and the signature of the decay products of U-235, in differing proportions depending on the age/purity/etc of the sample. Therefore we cannot use just one isotope when calculating a threat profile.

8 Conclusion

8.1 What we have done

We have constructed a fast algorithm to run on the SIGMA data and report possible anomalies for further consideration. To summarize, the detection procedure with specified parameters $\mu_{\min}$, h_{clear}, k is as follows:

1. Set a sensible $\mu_{\min}$ based on your upper time limit for an anomaly that removes long fluctuations from the data (we suggest $\mu_{\min} = 1.05$ here but higher values of $\mu_{\min}$ may be sensible if the data contains more fluctuations)

2. Run FOCuS with the clearing parameter h_{clear} in order to easily reset after passing over large anomalies. The algorithm is not particularly sensitive to exactly what h_{clear} is, but in this report we have arbitrarily chosen $h_{\text{clear}} = 60$ in order to reset the algorithm after a minute.

3. Alert all instances of intervals giving a significance trace greater than $k^2/2$ (which indicates a k-sigma significance level) for further checking as a possible threat.

8.2 What we could do next

In order to specifically look for a particular threat profile, we could compute what the threat profile should be based on an isotope mix, and then include only a subset of energy band categories based on the utility ratio, choosing a utility threshold as appropriate. This would likely improve the accuracy of detection of specific known threat profiles.

We could also pair FOCuS as a preliminary method with a more accurate but more computationally expensive method, as follows:

We deliberately choose a low sigma significance level k and therefore high false positive rate in order to ensure that FOCuS when run as preliminary accurately picks up all anomalies of interest. For example, if our overall desired false positive rate is one in eight hours requiring human input checking, we run FOCuS with a false positive rate of one in every ten minutes (600 seconds) and only feed positives highlighted by FOCuS into our more computationally expensive algorithm. This cuts down the amount of computation needed for the more expensive algorithm by at least 600 times, if not more as FOCuS can accurately estimate the start point of anomalies so each positive only requires the checking of one interval. This approach is summarised in Figure 12.

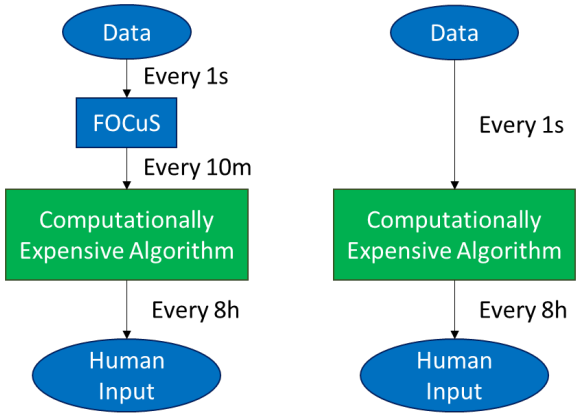

Figure 12: Flowcharts showing the computational processing comparison for with and without FOCuS as a preliminary method.

Acknowledgements

My PhD work is supported by the Engineering and Physical Sciences Research Council grants EP/N031938/1 and EP/R004935/1. Thank you to my PhD supervisors Idris Eckley and Paul Fearnhead at Lancaster University for helping to source this collaboration. Thanks to the AWE SIGMA Data Challenge for project-specific funding and data access.

The SIGMA Data supporting this paper are not publicly available for security reasons. Please contact the AWE SIGMA Data Challenge team at `info@nusec.uk` if you want more information or to request access.

References

Harter, R. (2009). The minimum on a sliding window algorithm. *https://richardhartersworld.com/slidingmin/*.

Lucas, J. M. (1985). Counted data CUSUM's. *Technometrics*, 27(2):129–144.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3-4):523–527.

Ward, K., Dilillo, G., Eckley, I., and Fearnhead, P. (2023). Poisson-FOCuS: An efficient online method for detecting count bursts with application to gamma ray burst detection. *Journal of the American Statistical Association*, 0(0):1–13.

Wilks, S. S. (1938). The Large-Sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62.