Further analysis of the cross correlation method for the detection of North Korea tests

Gareth Bird
gareth.bird@sjc.ox.ac.uk

University of Oxford

February 20, 2018

This is not as complete as I would hope as I was planning on discussing this draft before finishing.

1 Introduction

With data from lower altitude satellites, seismic activity has been linked to particle bursts in similar L shells. [1]. In currently unpublished work, the methods for matching particle bursts has been modified to be used on public access gps electron count data with reasonable success. Further to this, Wach observed that by taking cross correlation data for the raw signal between energy channels of the satellites results in anomalous values compared to background data (see Figure 1) but there was not any discussions as to why.[2]

This work was undertaken as a continuation of Wach's work with main objective of taking and testing these analysis methods further. The project also aimed to detect the events not previously detected if the method could be replicated.

All code for the project insofar has been created using Python 2.7 with Jupyter notebooks with simple wget commands in ubuntu. There also exists a parallelised version of this method but they are currently not compatible with one another. All branches of the code can be found here.

2 Methods

2.1 Previous methodology modifications

After tracing through the code from the previous project and discussing with group, a couple of a minor issues arose.

- When downloading data using the meta-search class, the data returned retrieved all file names between the given time interval, but did not then cut according to the dates provided. This seemed problematic as the background data did impose an exact time interval. As discussed later, this does have a effect on the cross correlation spectra (See Section 3.1) but does not invalidate the method.
- To modify the satellite number for downloading, the source code needed to modified and was not clearly labelled. This has been fixed in both new branches, with Jack's having more options.
- By inspection of the shape/size of the histogram, the background data sample size could have been larger

As a result the first job was to make the previous code replicable and clarifying the method further.

2.2 General Procedure

All of my code involves analysing the data from the returned python dictionaries when calling the meta-search class method. Further, the majority fo the code is finding and manipulating cross correlation values. This was done by splitting the data into equal time periods, calculating the cross correlation values and storing the information about the given time intervals alongside it. This allows me to generate the plots outlined in Section 2.4 .

2.3 General Object Structure

In an attempt to make my work easier to merge with other people's efforts, all my work is split into classes and methods within *xcorrobj.py*. There are two main classes:

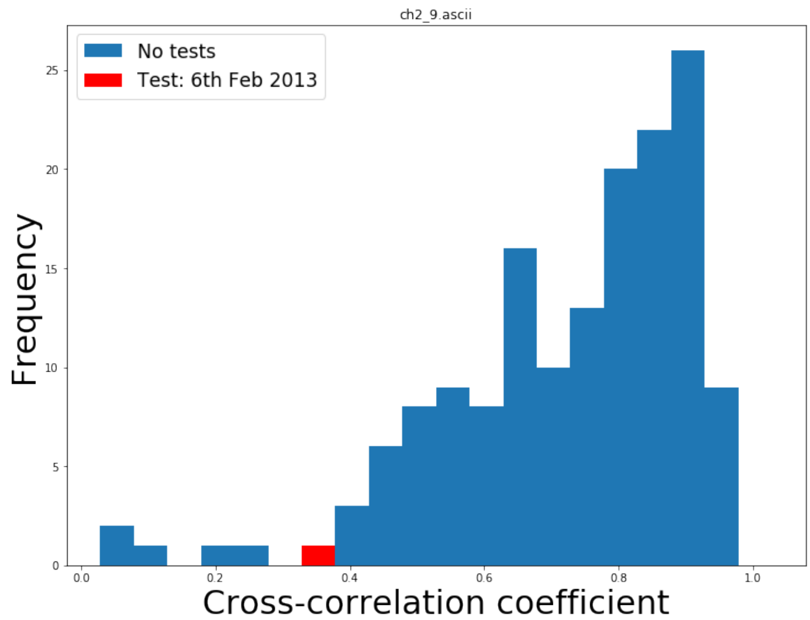


Figure 1: An example plot of an outlying cross correlation value during a north korea test, taken from [2]

1. **crosscorrelator**: Its main function is to calculate cross correlations of raw electron count signals over given time intervals. Stores data and dates for a given interval.
2. **plotgenerator**: Takes the satellite number and two instances of the correlator. Built to compare North Korea to Background data and save and repeat drawing instances of the data used in the plots. [1]

2.4 Plot procedures

The most instructive way to see all the plots generated and to get an idea of structure is to look at the *routine demonstrations* python notebooks for given satellites. The intended output of the plots methods are described in Table 1 . For specific implementation information/ input arguments, please refer to the commented code within *xcorrobj.py*. To save any particular plots to file instead of displaying, change the filename from an empty string.

Method Name(s)	**Description of output plot(s)**
crosscorrelator.create_NK_plot	Recreates cross correaltion spectra plots in the same way as Figure 1 but with the fixed time intervals.
crosscorrelator.createscatters	A redundant method, replaced with generate_signal_time_plots
plotgenerator.generate_signal_time_plots	Generates scatters of plots with a range of cross correlation values for two channels with matching time series plots. Can also create matching fourier transforms and labels the scatters if they include bad data.
...tor.generate_bad_data_demonstration	Creates histograms of comparing the cross correlation values of intervals containing bad data labels in the sample interval file and
fulldataconstruction or eventfinding	Generates a full set of cross correlation values for given intervals alongside finding when the minimum cross correlation value occurs in the North Korea data set and it's associated p value (proportion of background data smaller than the value)

Table 1: Descriptions of the plot methods within *xcorrobj.py*

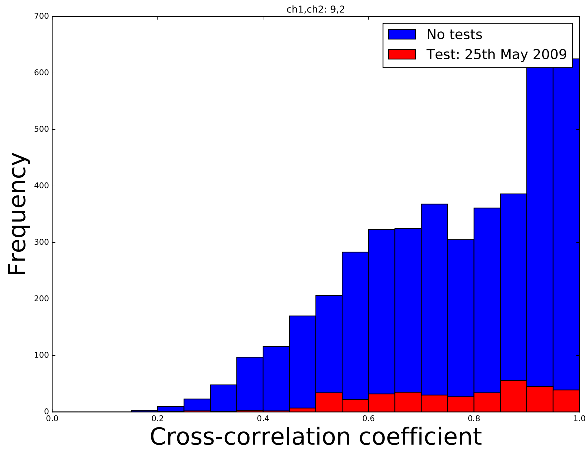
3 Results

This section is far from a complete set plots/deductions that can be found within the data files. Most of what is discussed in Sections 3.1 and 4 arose from Skype discussion of the presentations. If you are continuing investigating these phenomena I'd recommend scanning through the extent of these plots yourself here alongside generating your own.

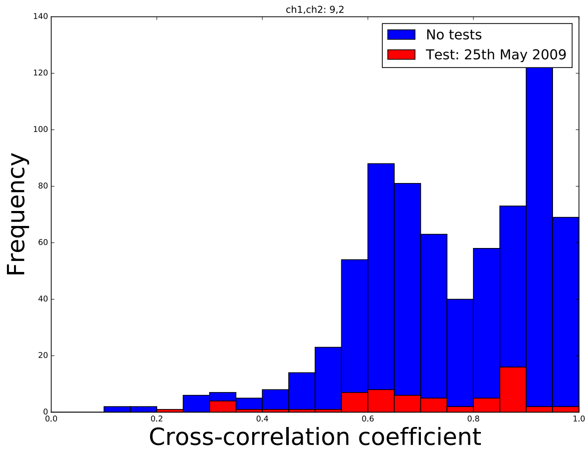
[1] A caveat is that not all the plot procedures are currently within plot generator although the code could easily be modified so that is the case.

3.1 Plots and Trends

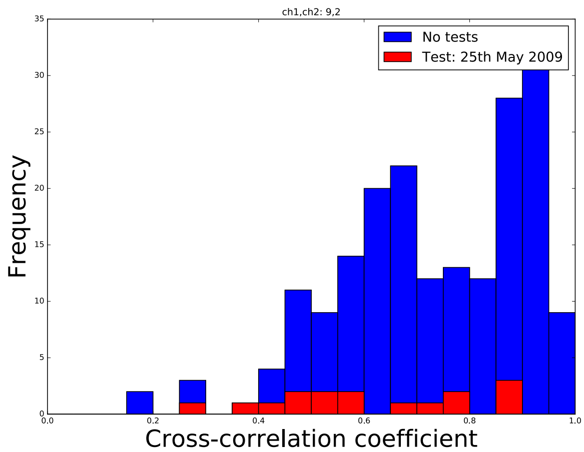
3.1.1 Further Cross Correlation Spectra Plots with timestamps (*eventfinding*)



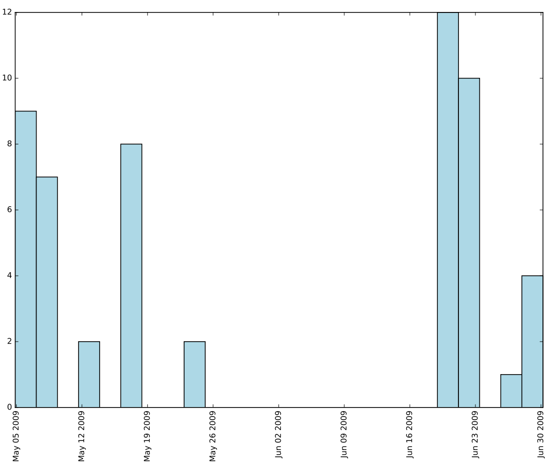
(a) 4 Hours



(b) 24 Hours



(c) 90 Hours



(d) Date spread of lowest cross correlation value for all channels with 90 Hour time intervals

Figure 2: North Korea Cross correlation spectra for satellite ns_56 with different

Once the adjustments in Section 2.1 had been made, we replicate Wach's results by working over a month time period around the North Korea test date to see if the anomalous value intervals clustered around the test date. To observe this specifically, we find the date interval which corresponds to smallest cross correlation in the.

From this you can observe the following:

- In Figure 2 and in the majority of the data it can be seen that taking time intervals significantly longer than the expected particle burst mechanism seems to make the North Korea spectra more distinct compared to the background.
- The dates of the smallest correlation value intervals don't cluster around the test date used, which is a concern.
- Especially with longer time intervals, the histograms of background seem to have unusual structure larger than a naïve estimate of the errors on the bins. [2]
- For two high energy channels, the cross correlation coefficients tended to be more likely to be close to 1 than the background data.

[2]This may be less important

3.1.2 Scatter, Time Series and Fourier transform plots

Even though there is no obvious temporal link between the tests and the correlation anomalies (other than reduced values around the month of test), the next logical step was to check the nature of data for the intervals with different correlation coefficients. This allows us to see if smaller/larger cross correlation coefficient values are caused by less correlated data, outliers or non-linear relations between the signals.

Additionally, by creating signal time plots alongside these we can see how structured scatter plots form over time. An example of this is shown in Figure 3

This revealed a large variety of exotic structures which has not been fully analysed but a few things are of note:

- Practically all of our plots have a periodic nature and some even seem to have a second periodicity convolved with it. This period seems to be roughly 6 hours but a FFT did not produce as clean a signal as one would hope. [3]

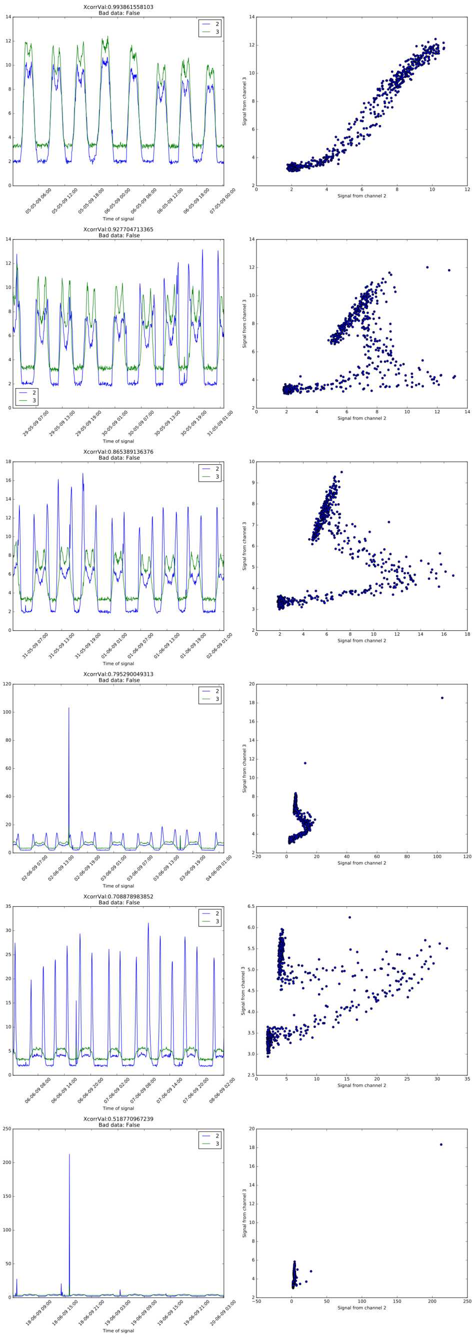
- There often seems to be 'gaps' in the scatter plot where there is no combination of signals and sometimes form ring-like attractor shapes

- There are 'bursts' of very large signals that can last an hour or so which are not always labelled as bad data.

- Bad data labels often do not break the data as much you'd expect because often bad data just seems to mean that there is no signal (i.e set to 0) for a short period of time

Furthermore, there is clearly quite a lot still to be analysed in the raw signal alone. However, it seems reasonably clear that just using a cross correlation coefficient to try and classify nuclear test behaviour from background seems unlikely without further metrics and a better understanding of the data.

Figure 3: Sample 48 hour correlation intervals between channels 2 and 3, two low energy channels on satellite ns_56



3.1.3 Bad Data Spectra

An additional set of plots were created to see if using bad data changed the cross correlation spectra. It seemed that that this didn't have a drastic effect on the overall set of cross correlation values.

4 Suggestions For Continuation

- Analyse data points with respect to position of satellites especially whether the satellite is within the van allen belt with the aim the to ascertain the signal periodicities.

- Consider completely classifying outlying data with more detailed analysis outlined T.-P. Li and Y.-Q. Ma which is already implemented within *gps_particle_data.py*. [3]

- Merge the methods described here with the method's used by Jack so that his methods can be used to further analyse any potential anomalies with the North Korea test data

[3] Checking this would be a good way to learn the code

References

[1] S. Y. Aleksandrin, A. Galper, L. Grishantzeva, S. Koldashov, L. Maslennikov, A. Murashov, P. Picozza, V. Sgrigna, and S. Voronov, "High-energy charged particle bursts in the near-earth space as earthquake precursors," in *Annales Geophysicae*, vol. 21, no. 2, 2003, pp. 597–602.

[2] F. Wach, "A study of van allen belt signatures of nuclear weapon tests for future ctbt technologies." [Online]. Available: https://github.com/fw14863/SP_2017/blob/master/FW_report.pdf

[3] T.-P. Li and Y.-Q. Ma, "Analysis methods for results in gamma-ray astronomy," *The Astrophysical Journal*, vol. 272, pp. 317–324, 1983.