Statistical Noise Baseline in AC6 Data

Mykhaylo Shumko

2019-02-04

This report describes the procedure taken to calculate the statistical noise baseline in the AC6 data. The motivation for this analysis is to estimate how often (or fraction of the time) that a cross-correlation (CC) above a certain threshold is due to random chance is not physically connected. This will help identify where in the Cumulative Distribution Function (CDF) plots the coincident events are just due to random chance and not anything truly physical.

**Count Binning Procedure**

The general approach here is to first bin AC6-A (since it has more 10 Hz data) dos1 counts at a similar magnetospheric location and condition. The binning was implemented by looping over every day with AC6A 10 Hz data, binning the counts by Lm\_OPQ, MLT\_OPQ, and AE. The AE data was appended to the file using minute cadence AE files. The binned counts were saved into its own file in the data/binned\_counts/ folder. Here is an example filename that designates what bin’s counts are in contained whithin “AC6\_counts\_4\_L\_5\_9\_MLT\_10\_300\_AE\_400.csv”. I defined bins in L, MLT, and AE. Currently, the bin edges are:

* L = {4, 5, 6, 7, 8},
* MLT = {0, 1, 2, 3,… 14}
* AE = {100, 200, 300, 400, 500}

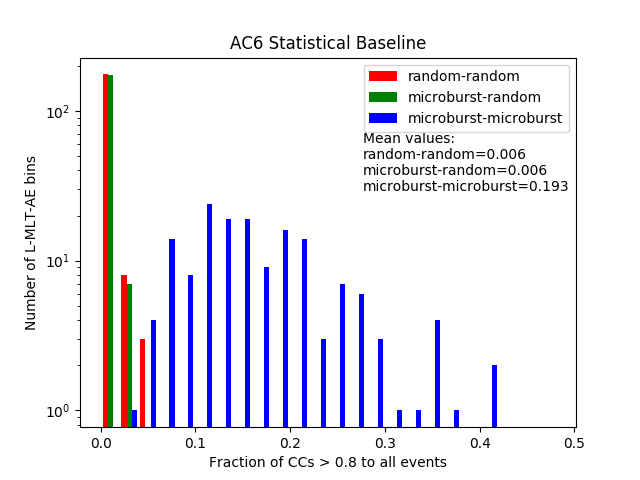
These files were created for convenience sake. It is much easier to loop over these smaller files than loop over all of the AC6A count files for each bin.

**The Statistical Baseline**

The overall approach here is to calculate a range of baselines given their own assumptions about what is CCd to give a range of plausible baselines. One such baseline “**random-random**” is calculated with a CC of random counts to other random counts in each L-MLT-AE bin. The CC windows are 1 and 1.2 s with the *mode=’valid’* to account for timing and Poisson uncertainty. This should result in the lowest fraction of valid events for the CDF (fraction of random CC realizations above 0.8).

Another baseline considered here is the “**microburst-random”** which is calculated in a similar maner to random-random, with CC of randomly picked microburst detections and randomly picked data in each bin. This gives us knowledge of how often a microburst CCd with random data will give us a valid event. Note that the random data could include a microburst!

The last baseline I’ve considered here is the “**microburst-microburst”** which is calculated in a similar maner to random-random and microbust-random, with CC of two randomly picked microburst detections in each bin. Each bin must have at least 10 unique microbursts to CC or it is skipped to avoid CCing a handful of microbursts against themselves to bias the fraction towards higher values. This gives us knowledge of the fraction of valid events that are due to CC of two unrelated microburst detections.

The plot on the right shows a histogram of the number of L-MLT-AE bins as a function of the fraction of valid events in that bin.

It is interesting to note that random-random and microburst-random have a very similar distribution that has a mean value of ~ 1 % chance of a valid event. I would expect the microburst-random to go further out in the x-axis.

The microburst-microburst distribution goes out to higher fractions as I would expect with a mean value of 0.2. Thus we can say that 20% of the time two unrelated microbursts will have a CC > 0.8 in valid AC6 data (values with flag != 0 and dos1rate=-1E31 removed).

**Interpretation**

I need to relate these values back to the CDF distribution that I’ve calculated to estimate what is the statistical background that is present through this distribution. Having a good grasp of this is essential to determine the error on the CDF as a function of separation for least squares regression and Bayes modeling soon after.

I am currently interpreting this as for example, if microbursts occur 50% of the time on average (take Paul’s Fig. 5 in O’Brien et al., 2003 JGR paper), a random microburst will appear to be coincident roughly 1% of the time with random data and 20% of the time with other random microbursts so the overall baseline would be 1/2\*1% + 1/2\*20% = 11%. But if we choose a smaller microburst duty cycle e.g. 10%, then 10% of the time we will CC a microburst against an another microburst and 90% of the time against non-microbursts. If I am doing my math correctly, this would correspond to 9/10\*1% + 1/10\*20% = 2%.

~~One interpretation is a baseline defined by the mean of the microburst-microburst distribution. By making this assumption, we are essentially saying that the CDF is a superposition of some number of true coincident microbursts, out of which a 0.2 fraction of them are valid events from microbursts that are unrelated.~~

~~20% of the events when we have a CC > 0.8 in the CDF~~

~~One thing to consider is that~~

~~in the scenario where I am CCing only microbursts against other unrelated microbursts.~~