EXTENDING THE RUN04 BAD AREA CUT

Kr-83m injection data

Rachel Mannino

7 November 2017

RUN04 ^{83M}KR DATASETS

- Using single-scatter ^{83m}Kr injection data from Run04:
  - kr83minjections_TB1.mat
  - kr83minjections_TB2.mat
  - kr83minjections_TB3.mat
  - kr83minjections_TB4.mat
    - Evan created these with filter code
- Plan: Use ^{83m}Kr data to extend the bad area cut as this provides high statistics dataset of 32.1 keV + 9.4 keV IC electrons.
  - May merge within an event to look like a 41.5 keV signal.

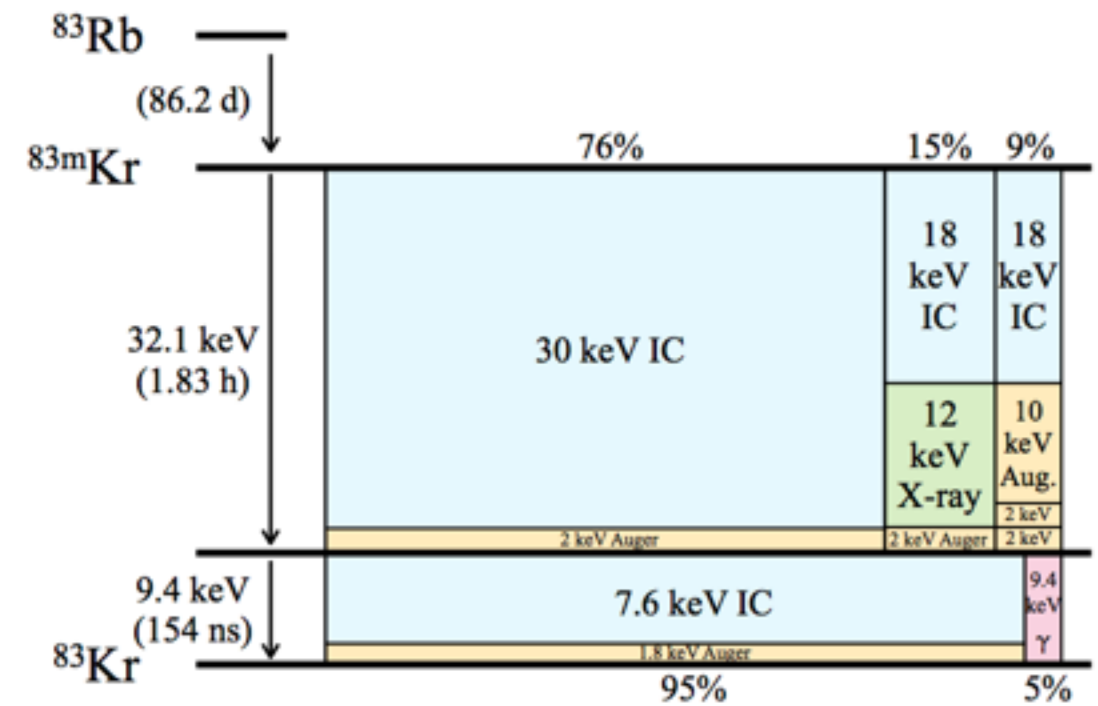
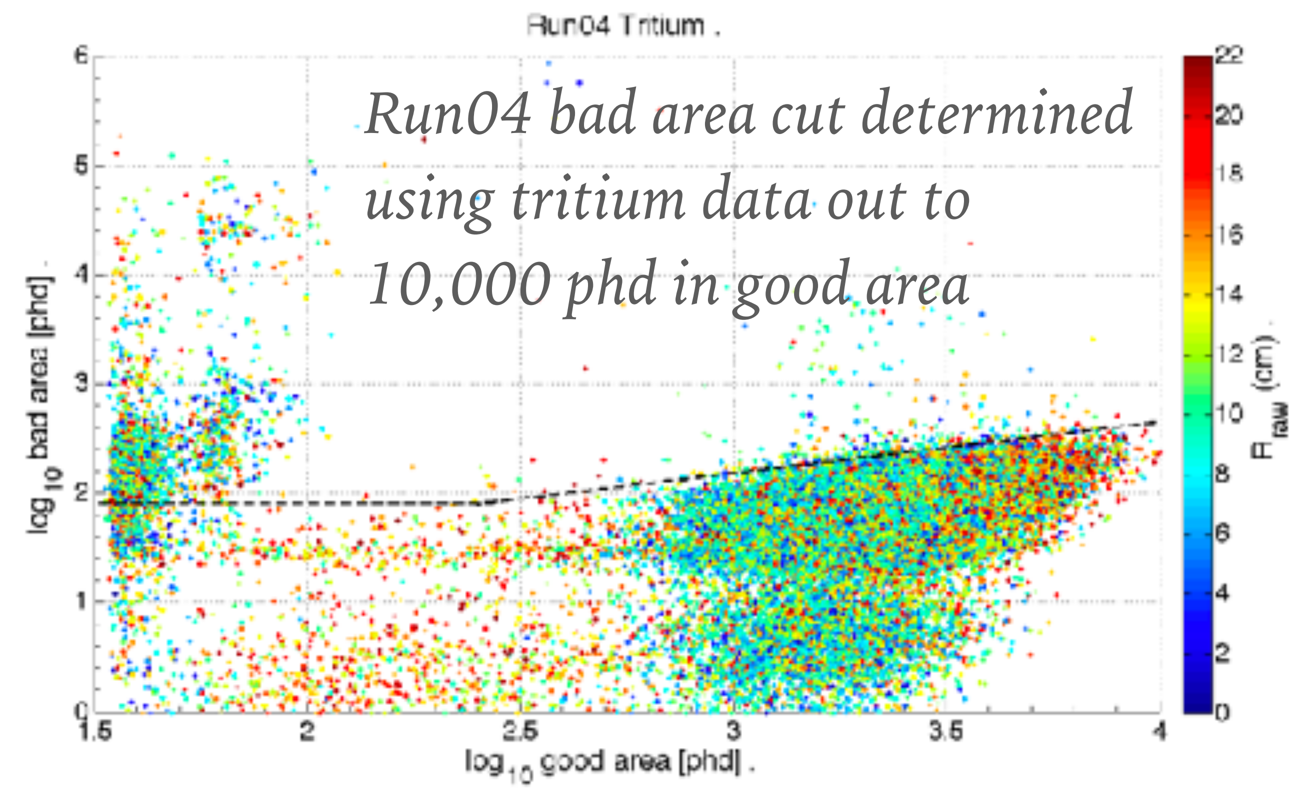


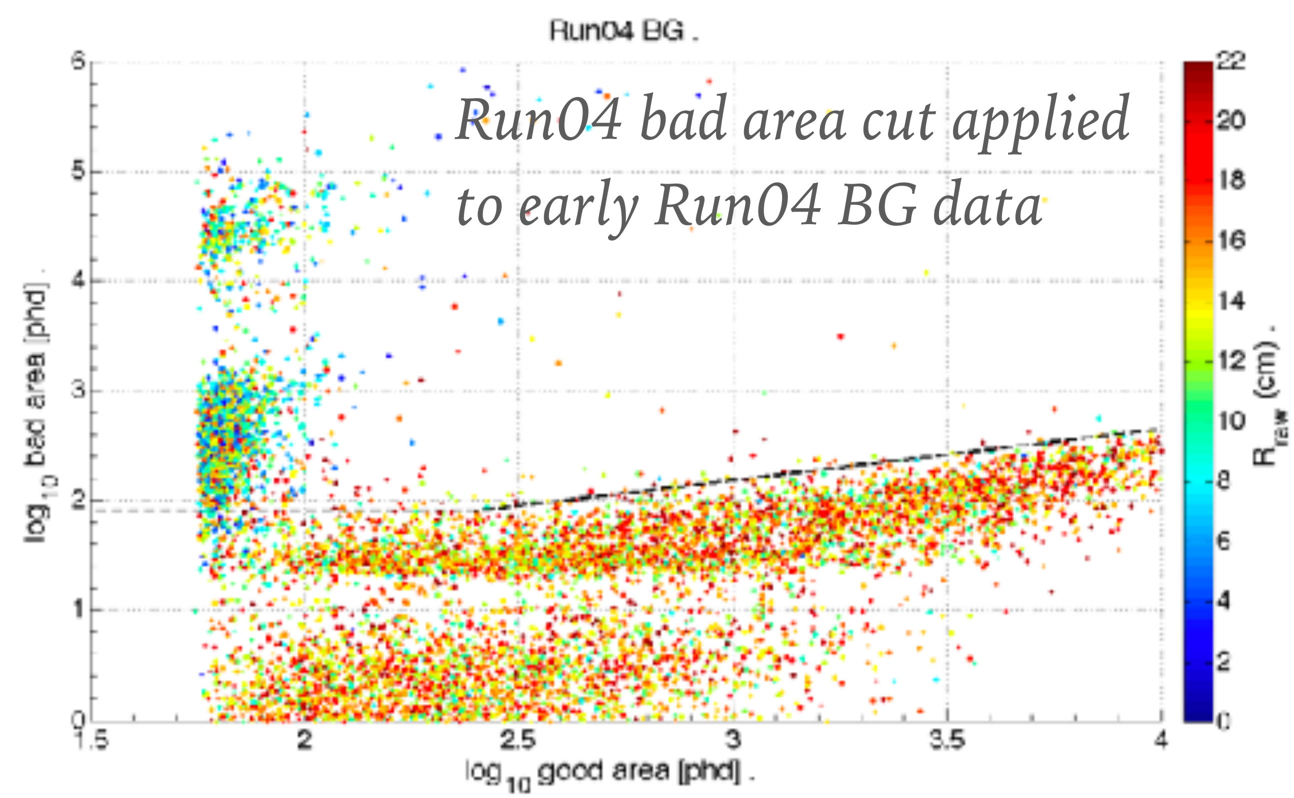
FIG. 1. Decay schematic of ^{83m}Kr . The width of each column is proportional to the branching fraction of that decay mode, the vertical divisions are proportional to energy partitioning among internal conversion electrons, Auger electrons, x-rays, and gamma-rays. Numerical values from Reference [2].

arXiv:0905.1766

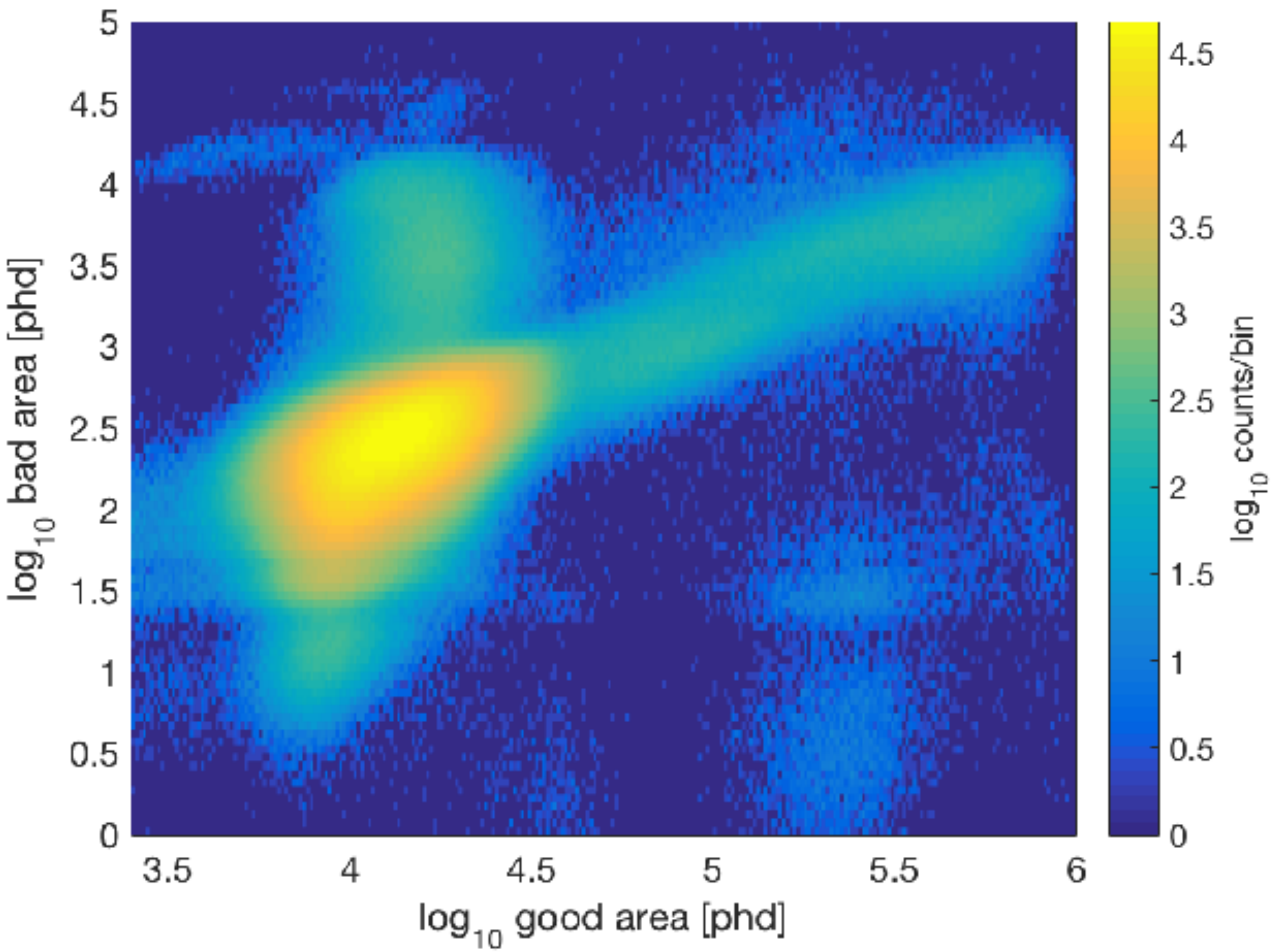
GOOD AREA AND BAD AREA

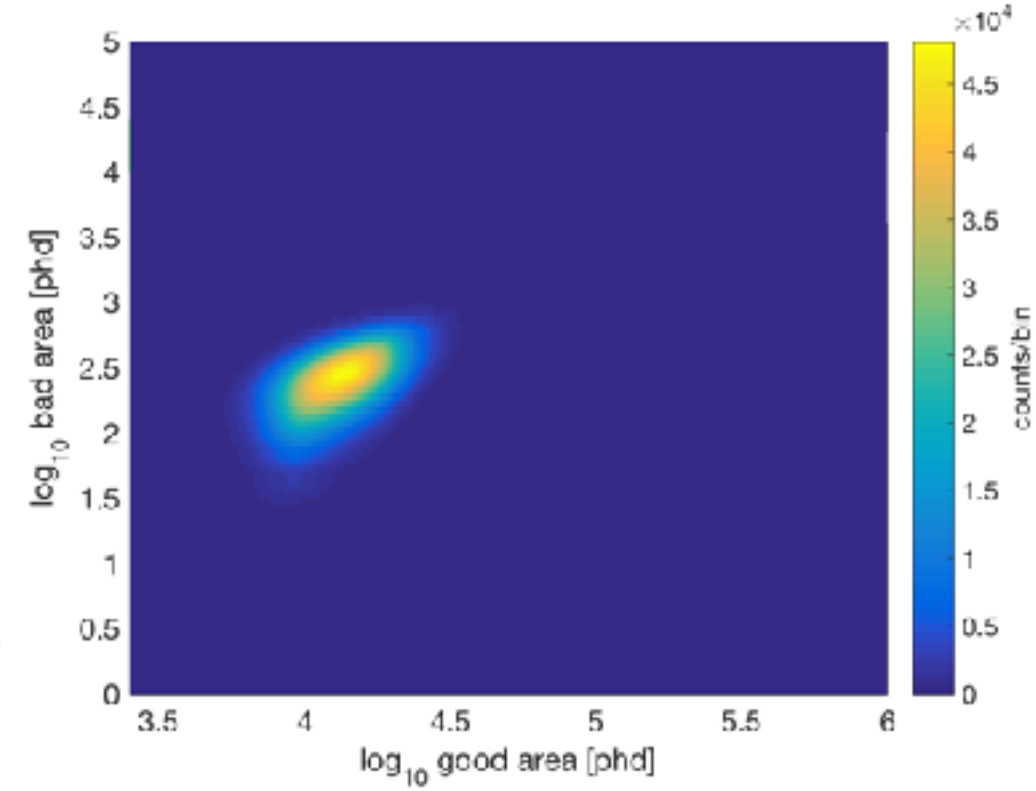
- Good area = S1 + S2;
- Bad area = full event area - good area;
- Bad area cut removes events where the event window has anomalies such as electron trains, glow, etc.
  - LUX only keeps 10 pulses/event, so using the full_event_area_phe RQ captures the area of all signal area above baseline, even if the PulseFinder did not classify it as a pulse.
  - Designed for single-scatter events.
  - Calibrate bad area cut using high statistics datasets such as tritium (*earlier incarnation of Run04 bad area cut*) or ^{83m}Kr (*now*).
- Filter code creates "goodarea" and "badarea" RQs using uncorrected, raw S1 and S2 areas.





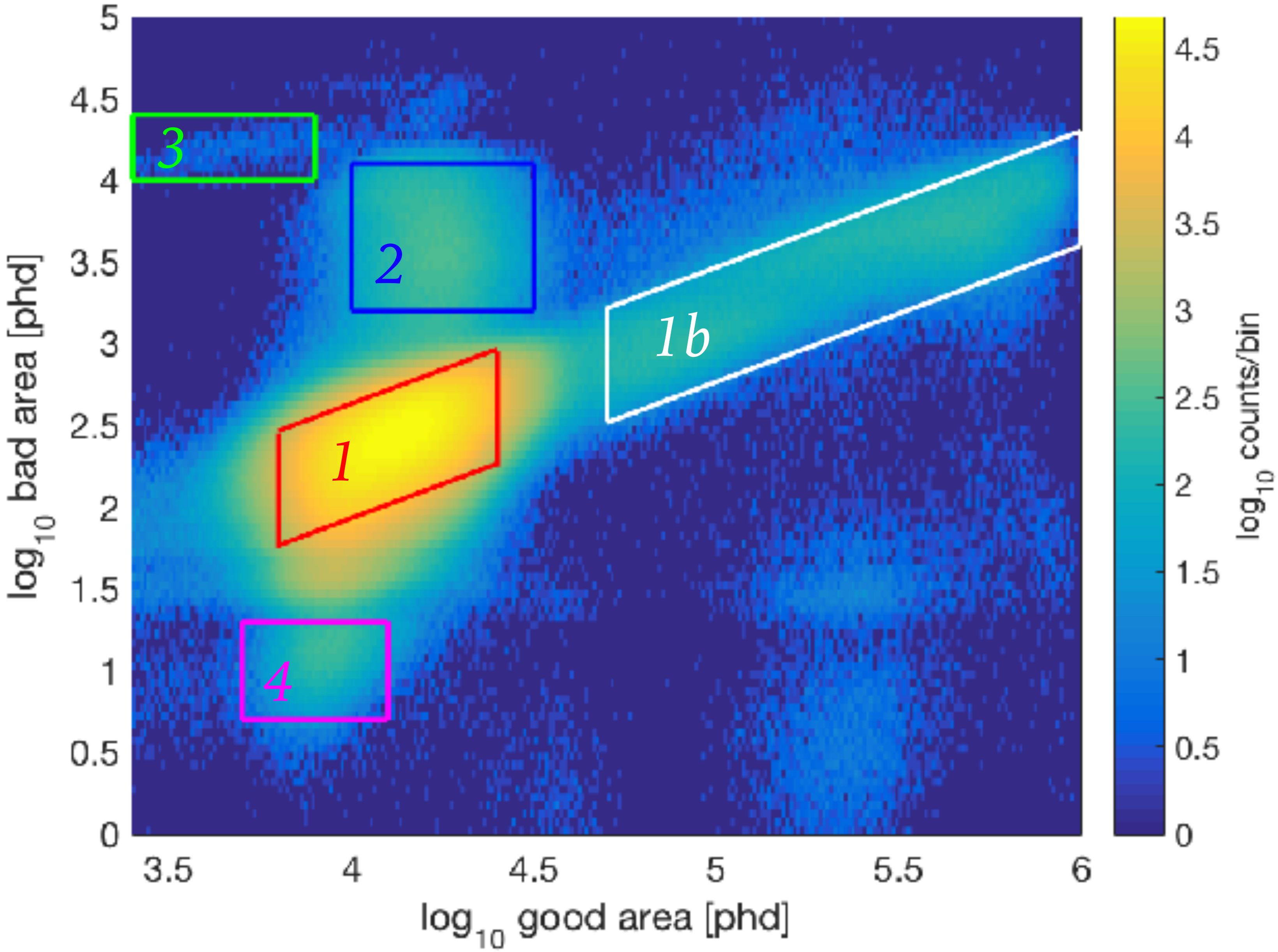
LOG$_{10}$(BAD AREA) VS. LOG$_{10}$(GOOD AREA)





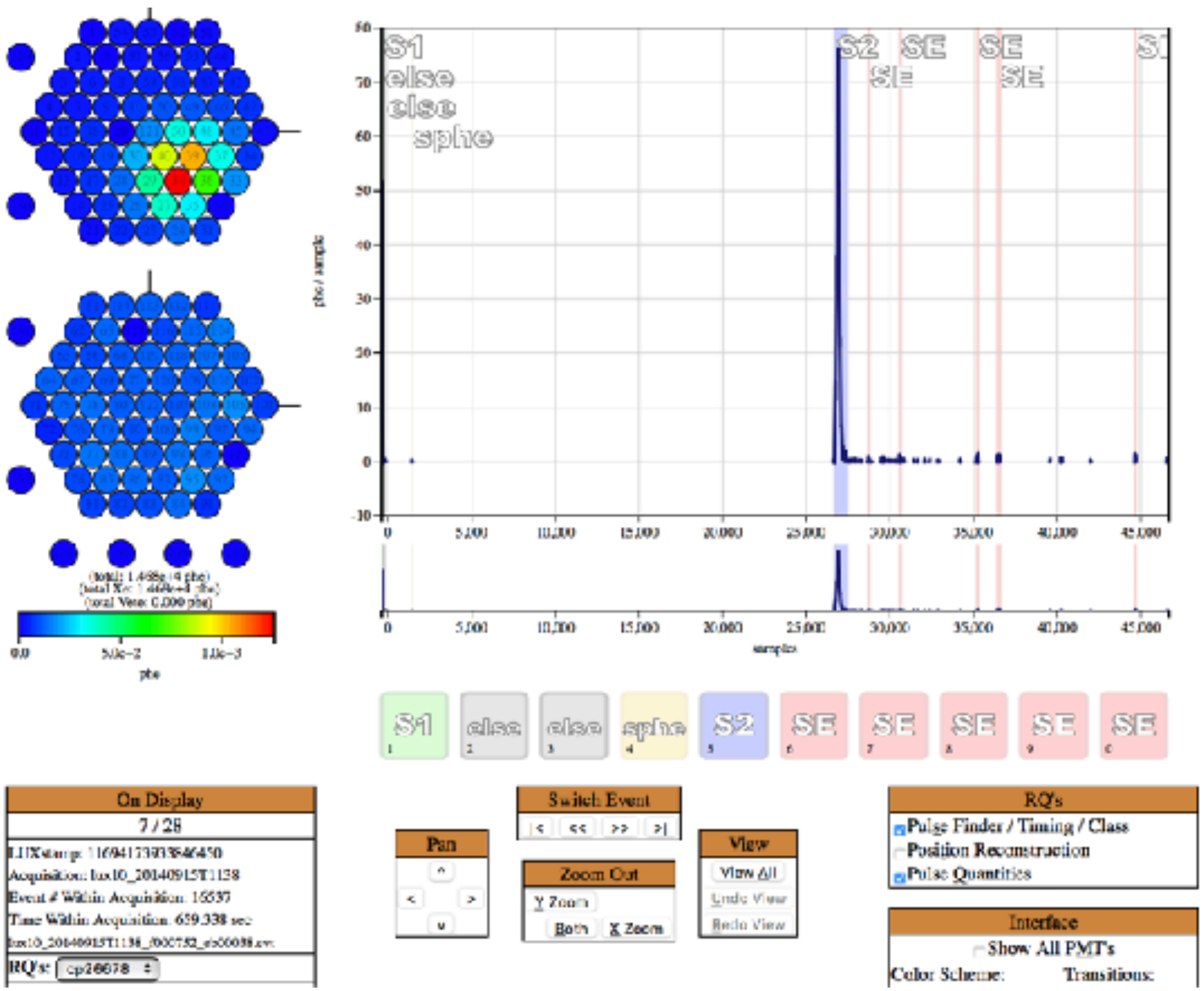
- Plot log$_{10}$(bad area) vs. log$_{10}$(good area) for the Run04 Kr-83m injection data.
  - *(left)* Color scale is in units of log$_{10}$(counts/bin).
  - *(above)* Color scale is in units of counts/bin.

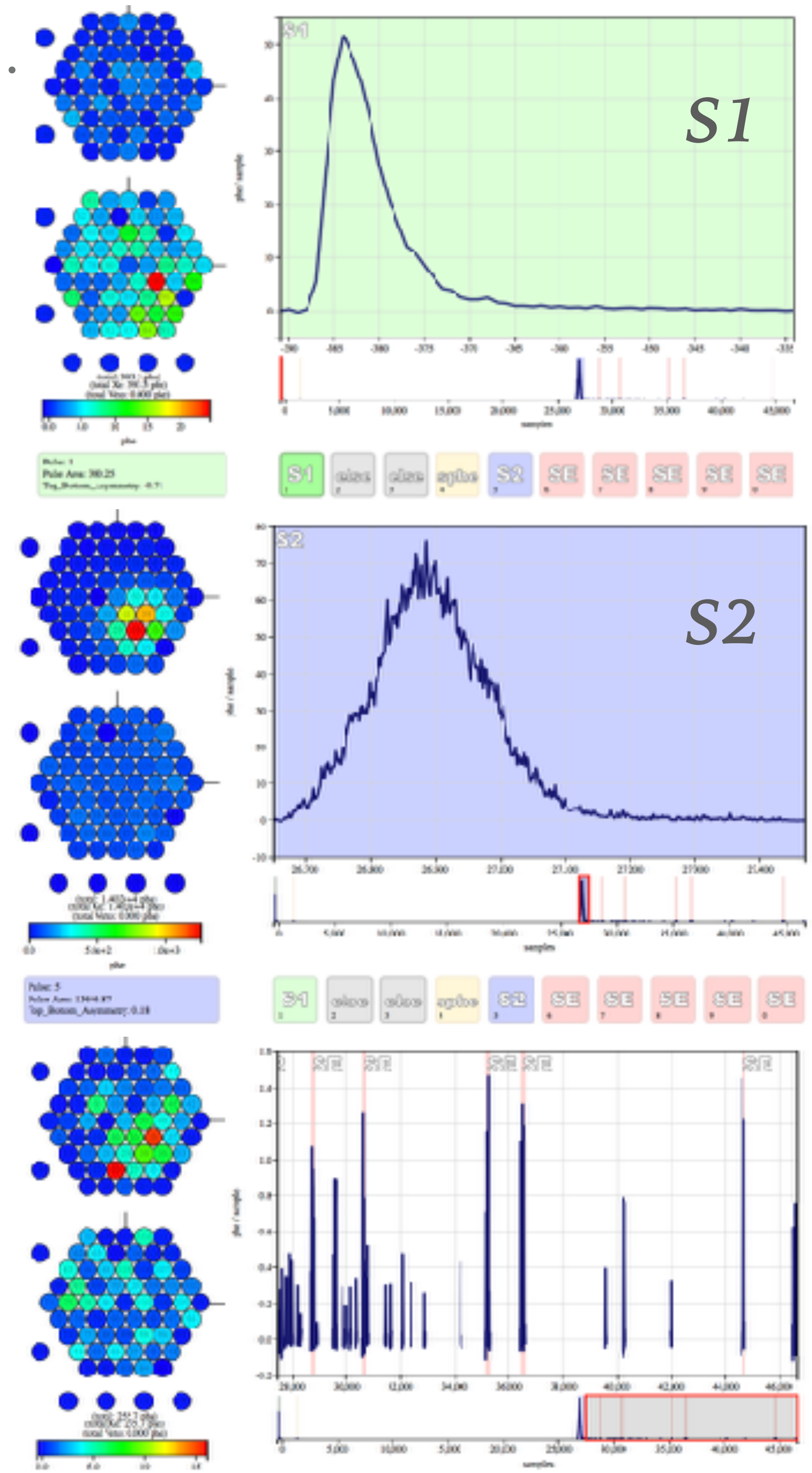
LOG$_{10}$(BAD AREA) VS. LOG$_{10}$(GOOD AREA)



- Roughly classify events into populations to study S1 & S2 areas, energies, and any anomalies in Visualux.
  - *(red)* Population 1
  - *(white)* Population 1b
  - *(blue)* Population 2
  - *(green)* Population 3
  - *(magenta)* Population 4

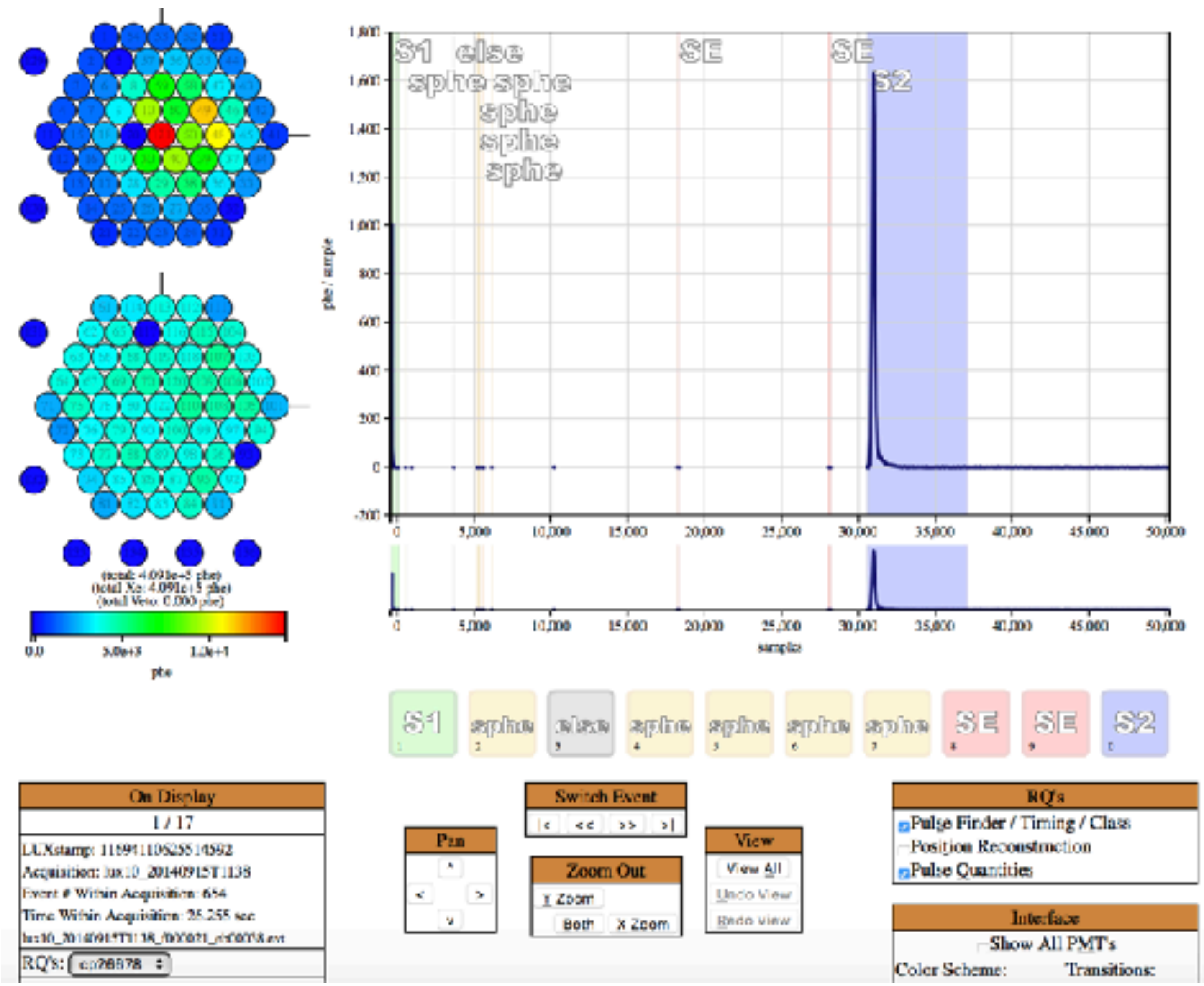
VISUALUX: POPULATION 1

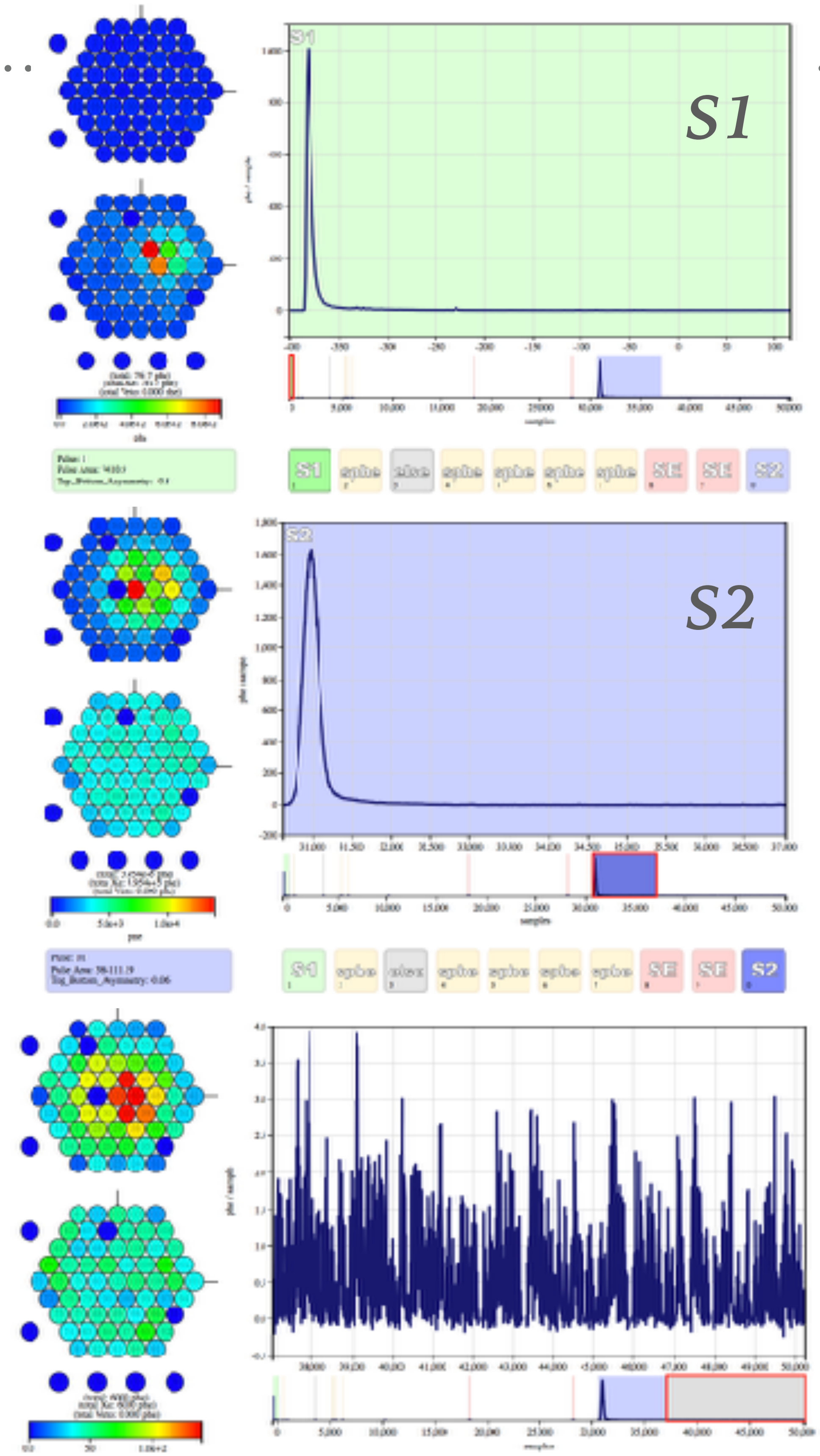




- Scanned 100 events of pop1 in VisuaLux. Events are single-scatter with the S2 followed by several-to-dozens of SE and SPE.
- *(lower right)* View of typical pulses following an S2. These don't qualify as e-trains, but there are likely to be >10 pulses/event.

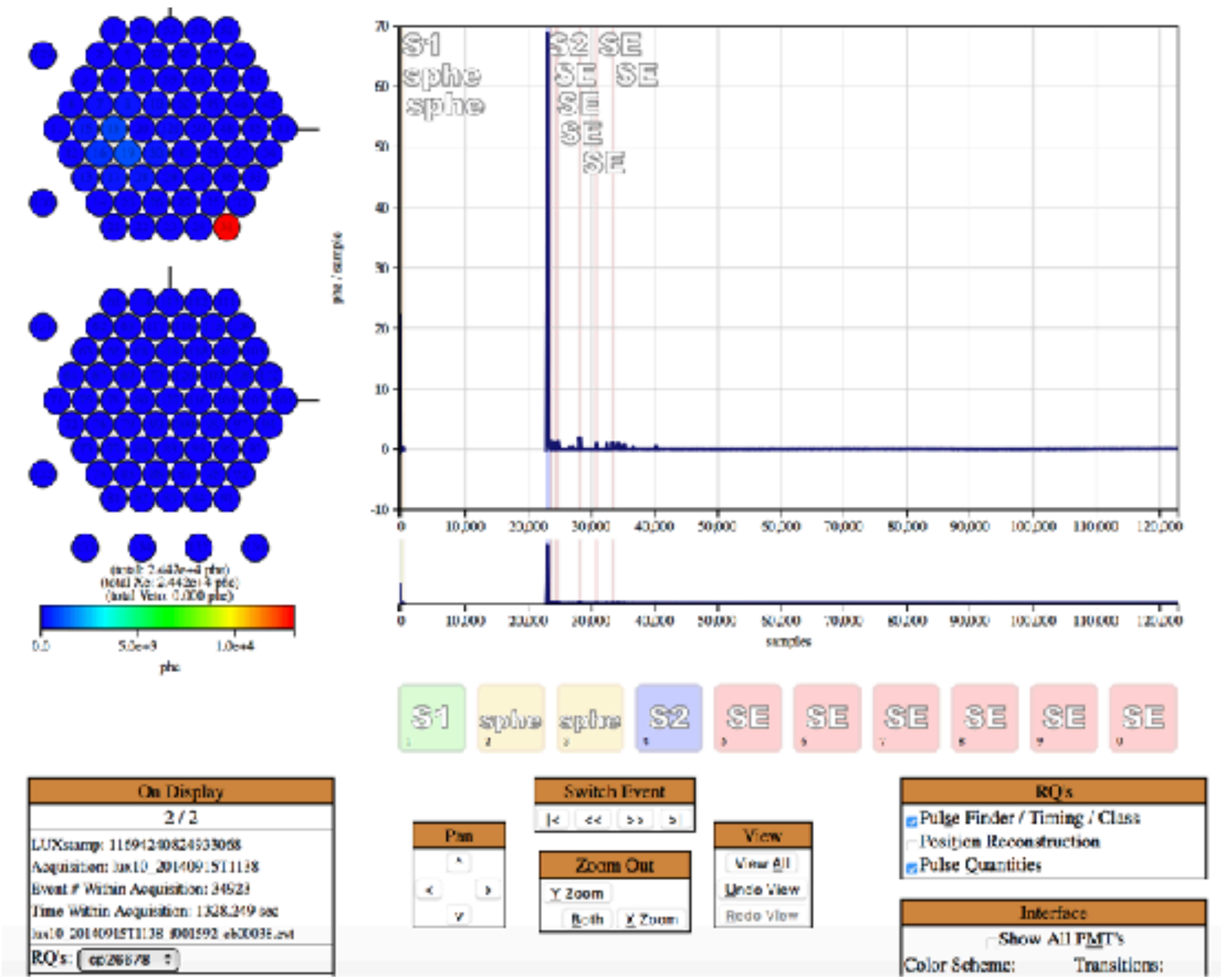
VISUALUX: POPULATION 1B

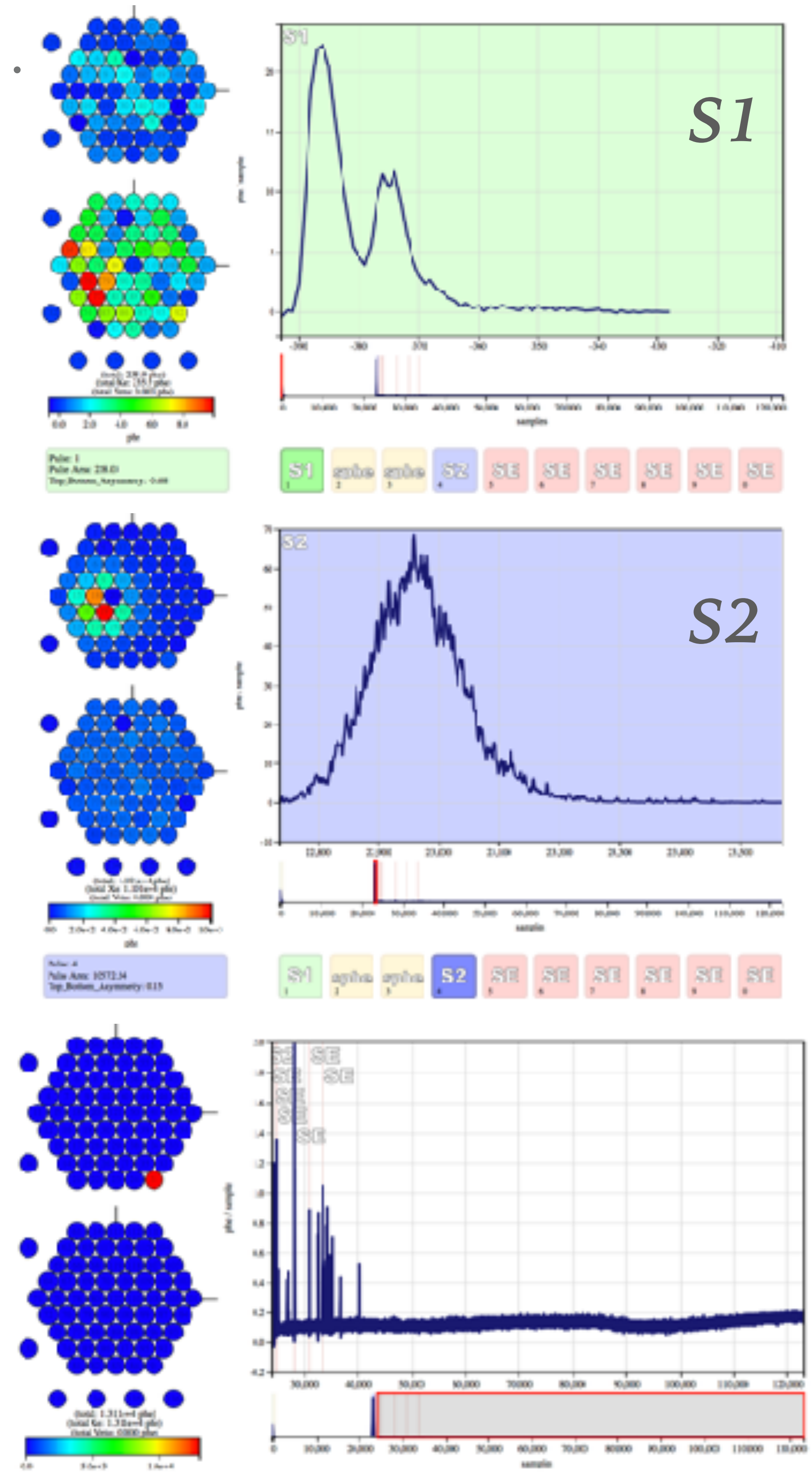




- Scanned 100 events of pop1b in VisuaLux. Events are largely single-scatter with the S2 followed by electron trains.
- *(lower right)* View of typical pulses following an S2. This is an e-train.

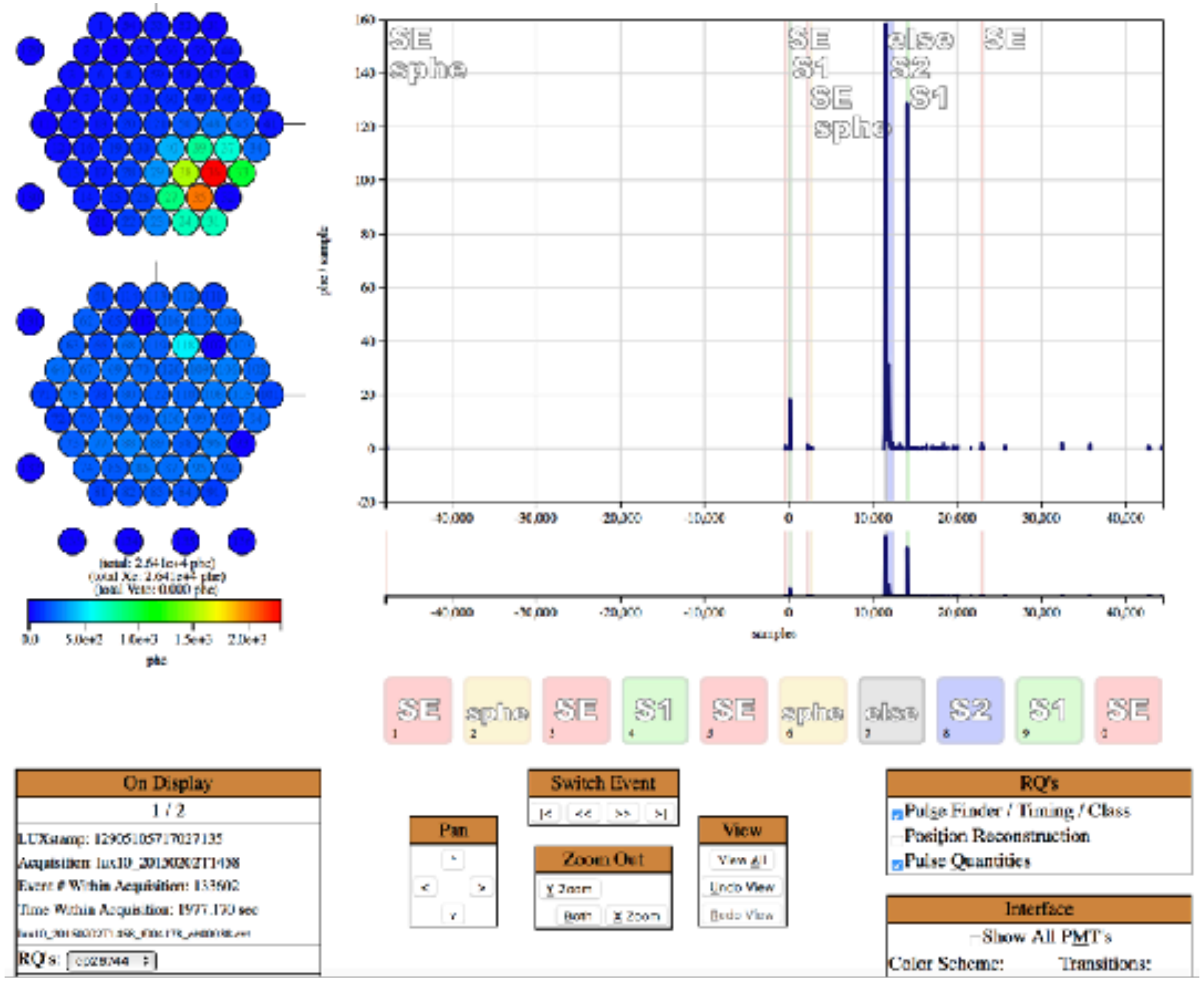
VISUALUX: POPULATION 2

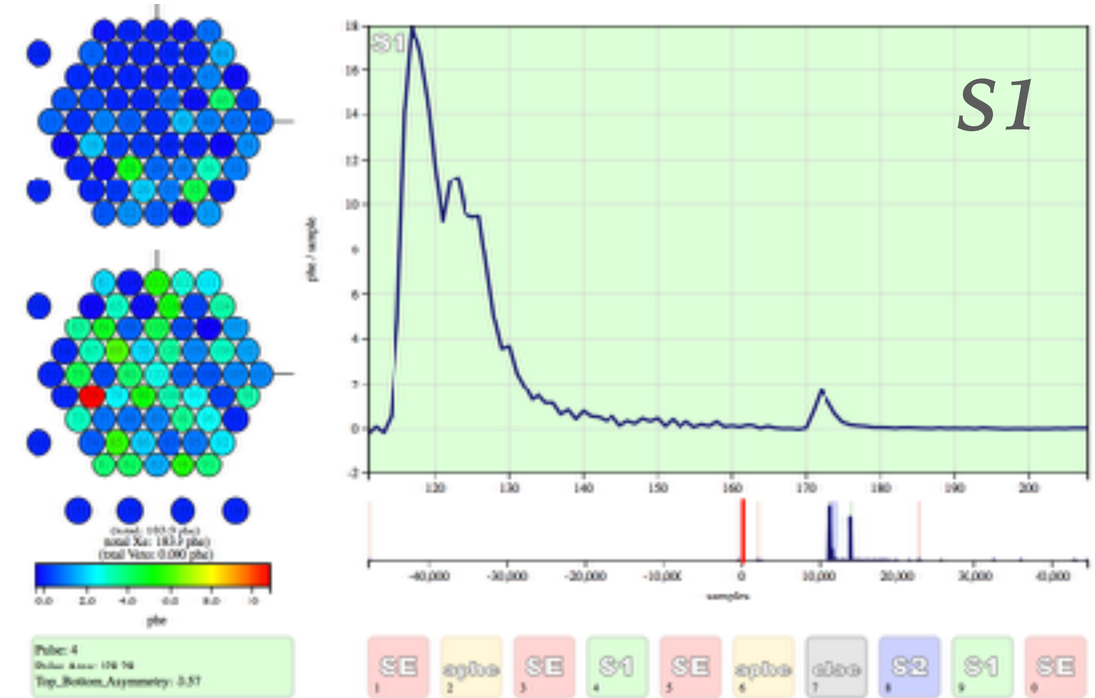


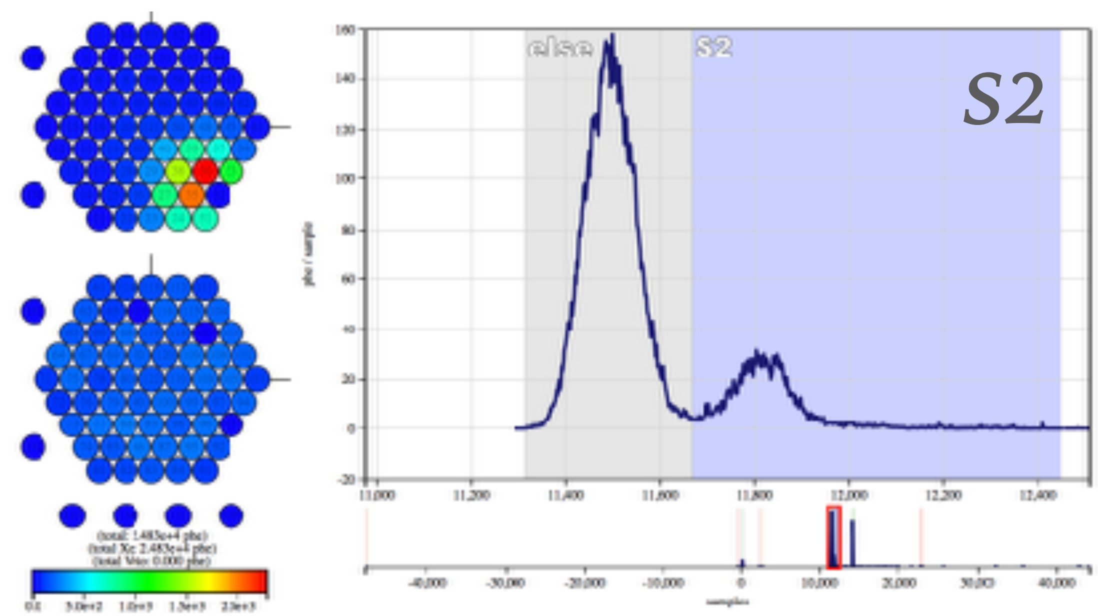


- Scanned 100 events of pop2 in VisuaLux. Events are largely single-scatter with the S2 followed by a baseline shift in the rest of the event window. 95% of scanned events had baseline problems.
- *(lower right)* View of typical pulses following an S2. This is a baseline shift.

VISUALUX: POPULATION 3

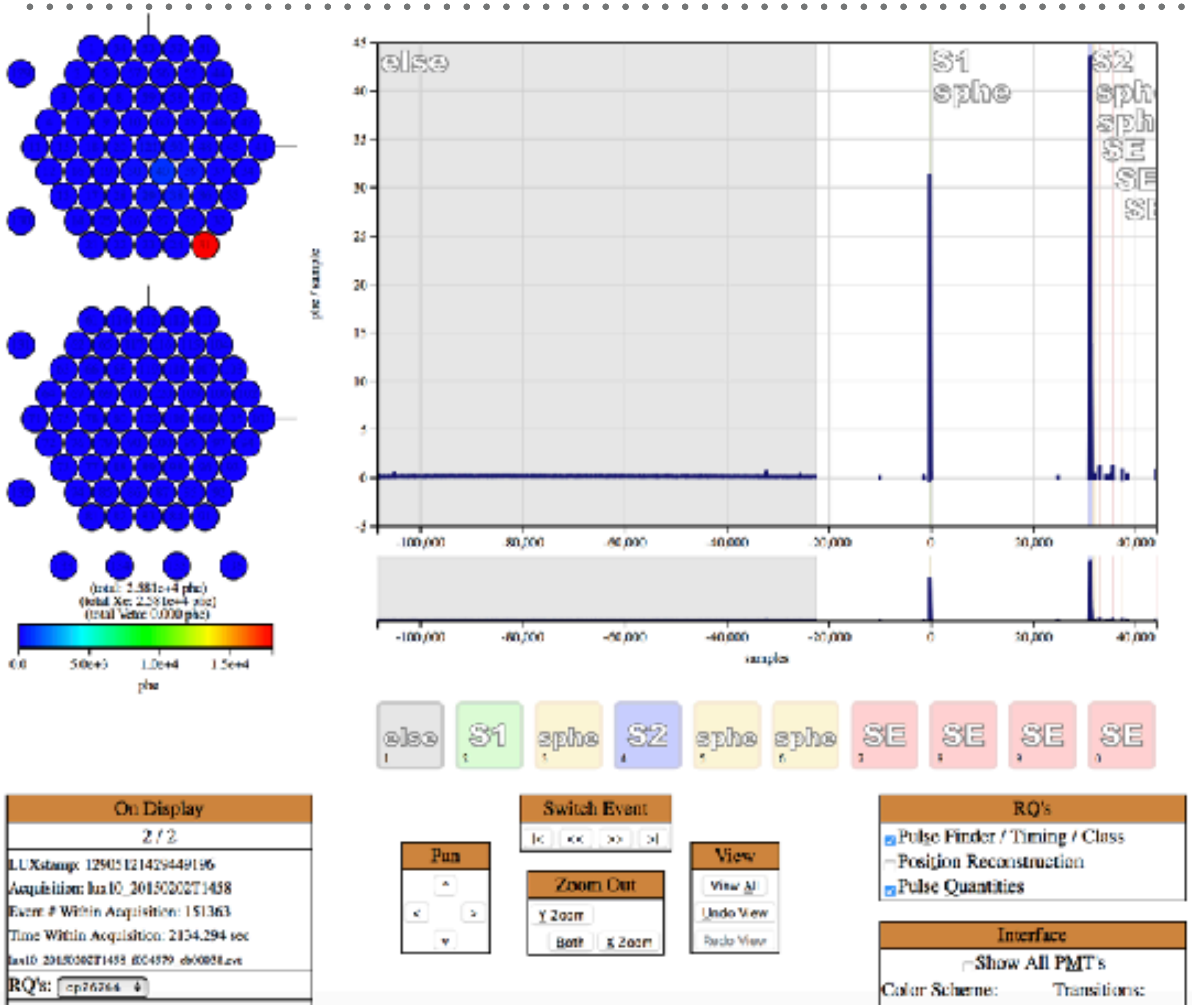


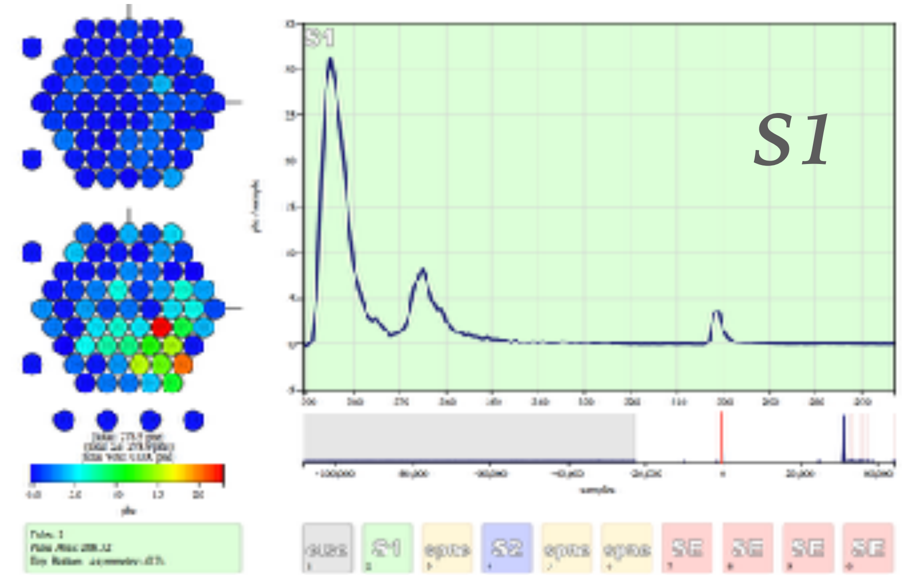


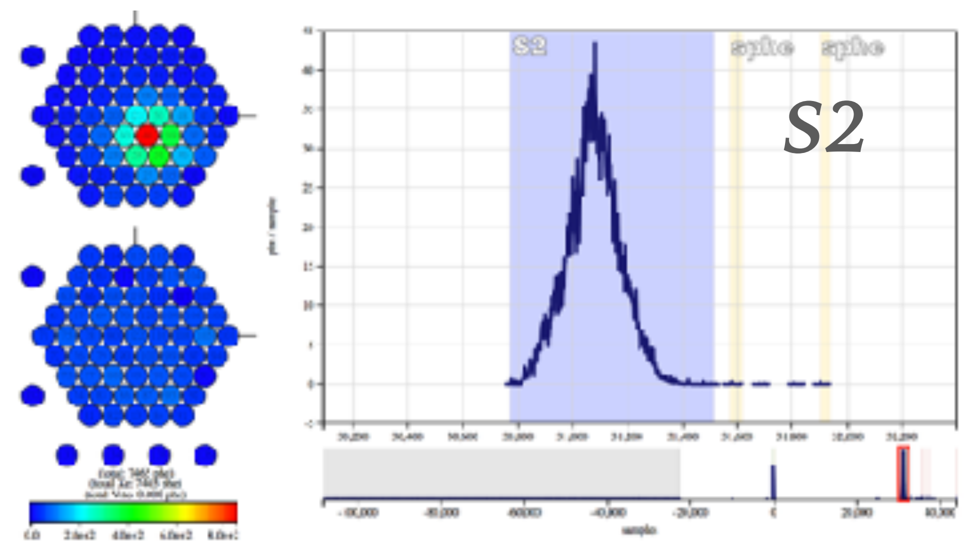


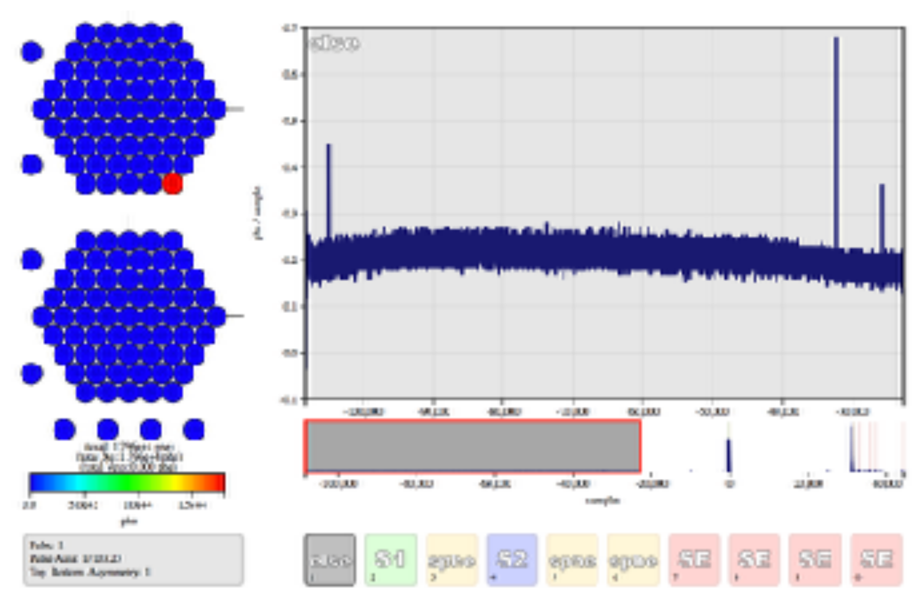
- ➤ Scanned 100 events of pop3 in VisuaLux. Events are *classified* as single-scatter, but 91% of the events looked like a double-scatter with the 1st S2 misclassified as an “else” or class 5 pulse. The S1 often, but not always, looked like 2 discernible S1 pulses.
  - ➤ The “else” pulse contributes to the “bad area!”

VISUALUX: POPULATION 3, CONTINUED



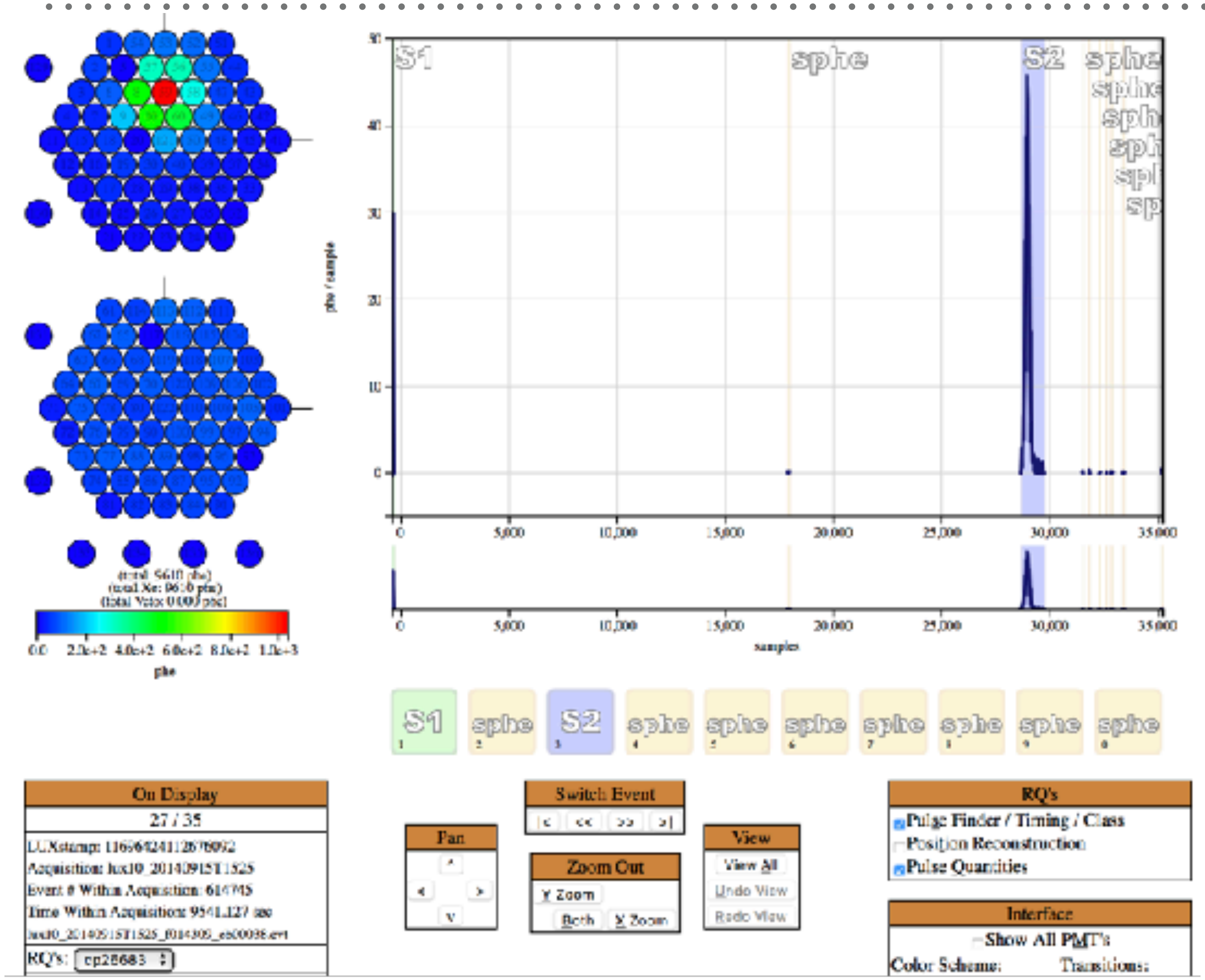


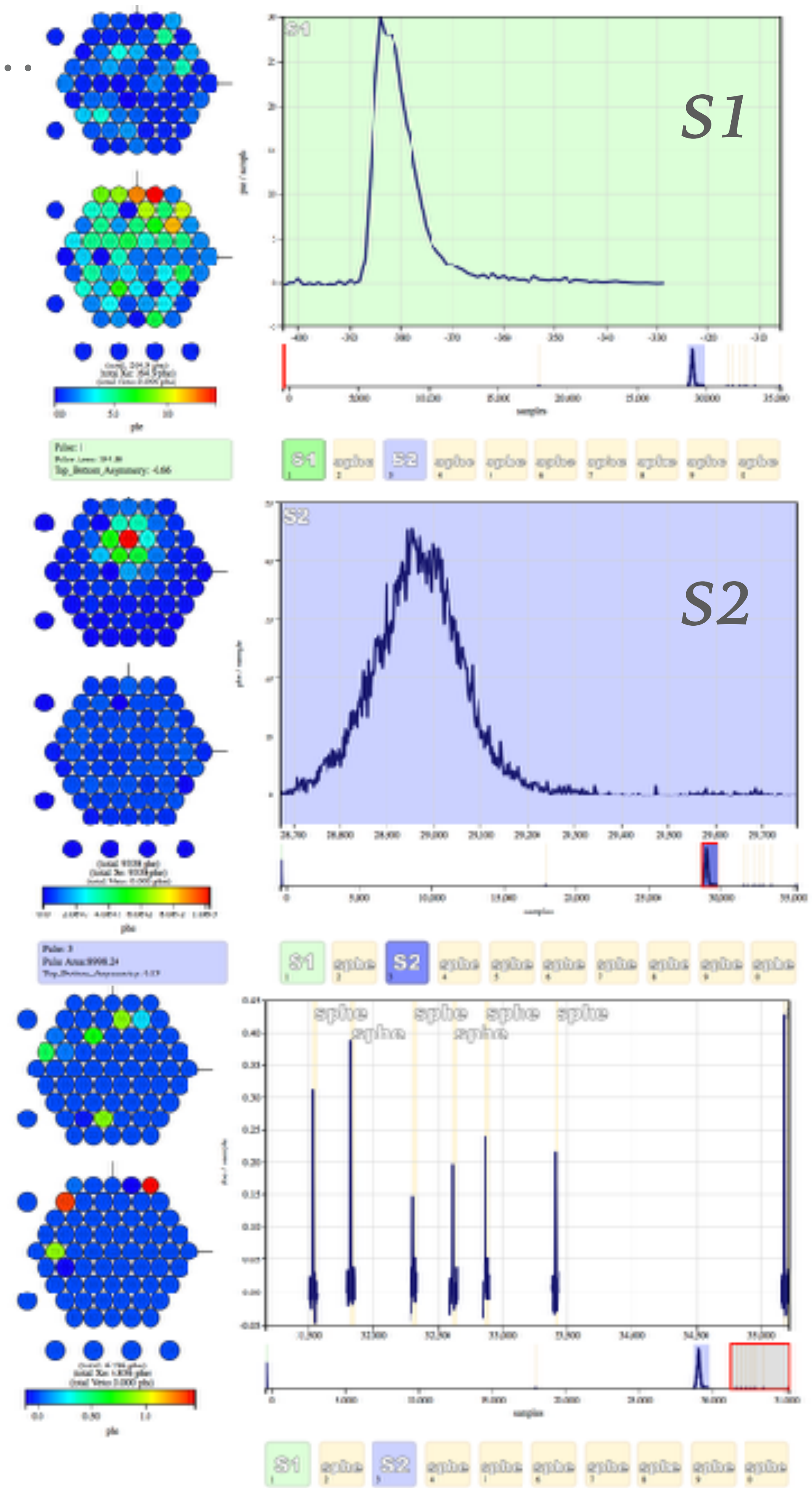




- Scanned 100 events of pop3 in VisuaLux. 7% of scanned events had baseline problems either before or after the S2 pulse.
- *(lower right)* View of a baseline shift early in the event window before the S1 pulse.
- 2% of events simply had a super large S1 pulse after the S2 contributing to the bad area. The pattern was S1, S2, big S1.

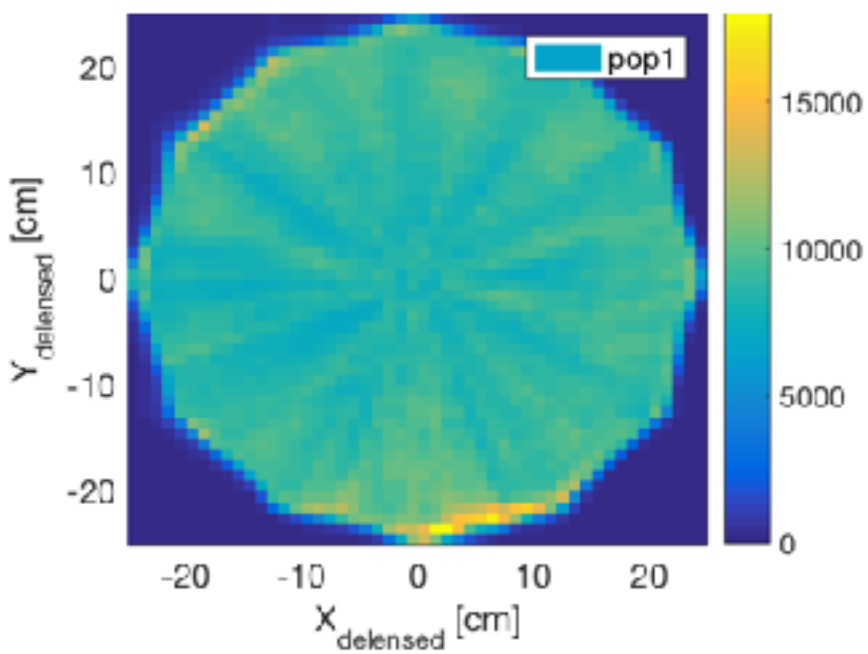
VISUALUX: POPULATION 4

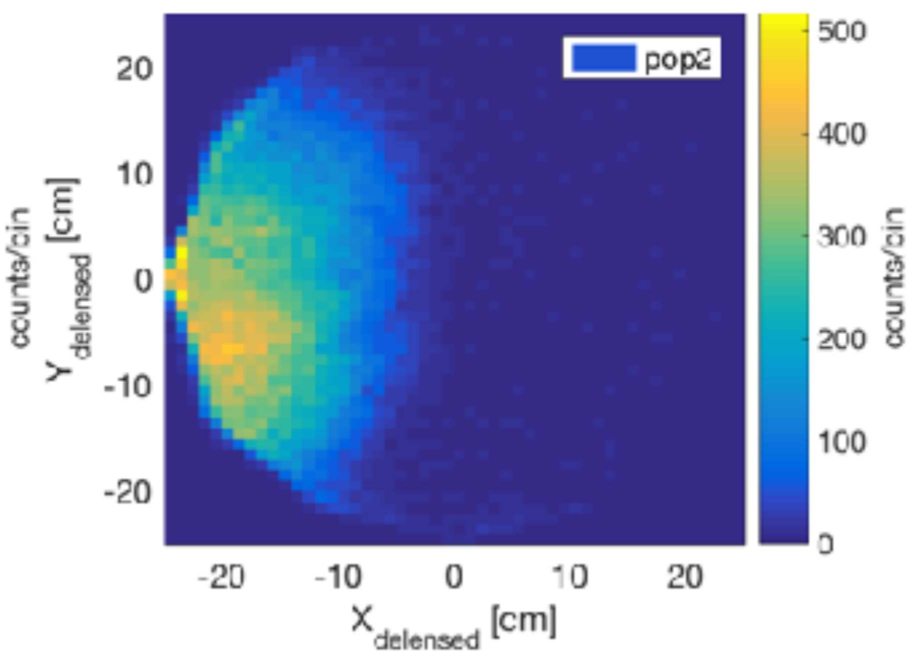


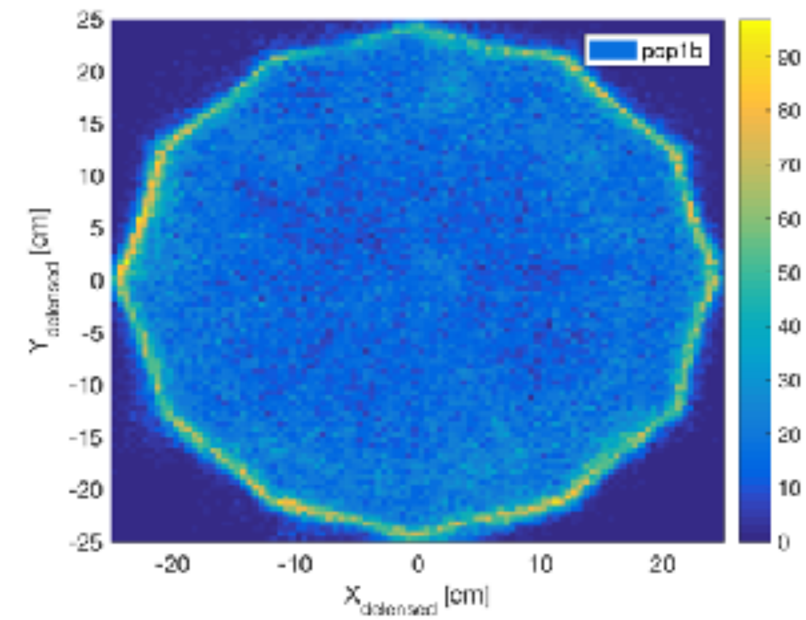


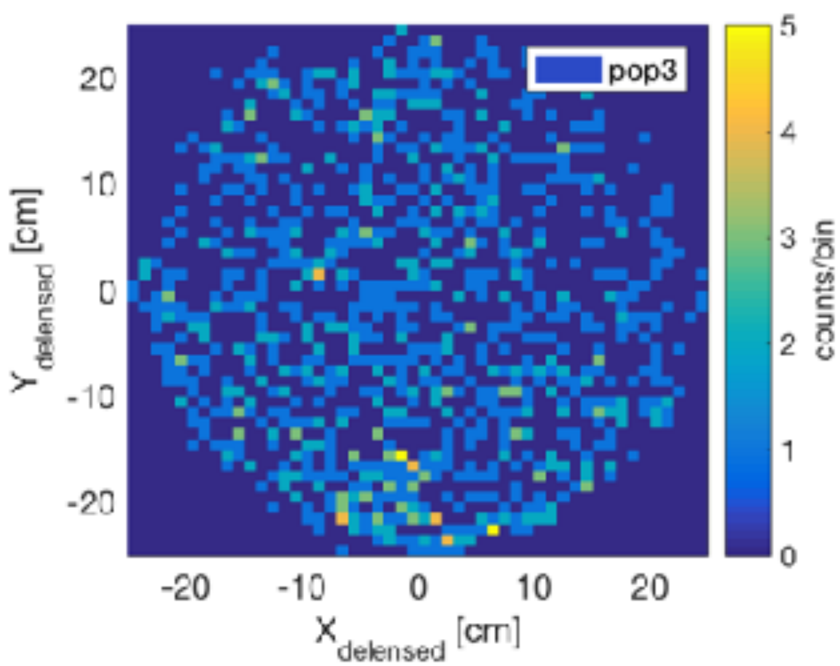
- Scanned 100 events of pop4 in VisuaLux. Events are largely nice-looking single-scatter events where the S2 is followed only by SPE. Occasionally (3-4 events scanned), a SE followed the S2, but this typically had a small area of ~10 phd.
- *(lower right)* View of the single photoelectrons (SPE) following an S2.

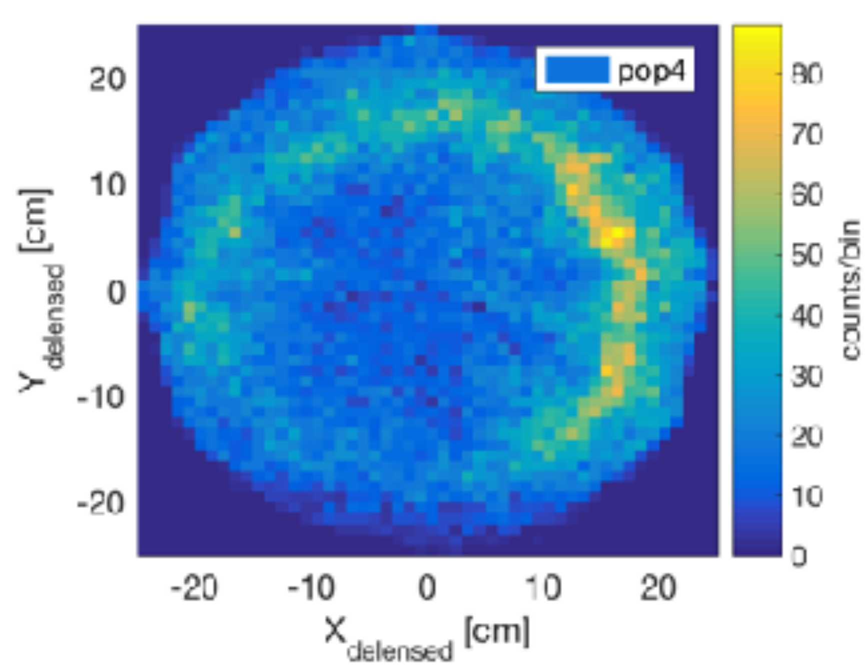
XY POSITIONS OF THE POPULATIONS





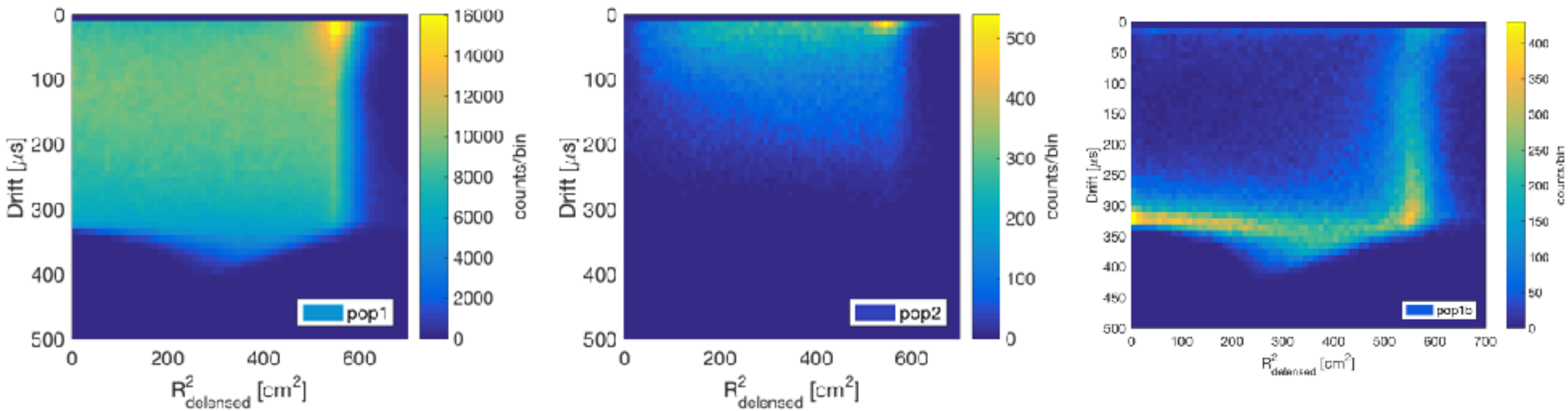


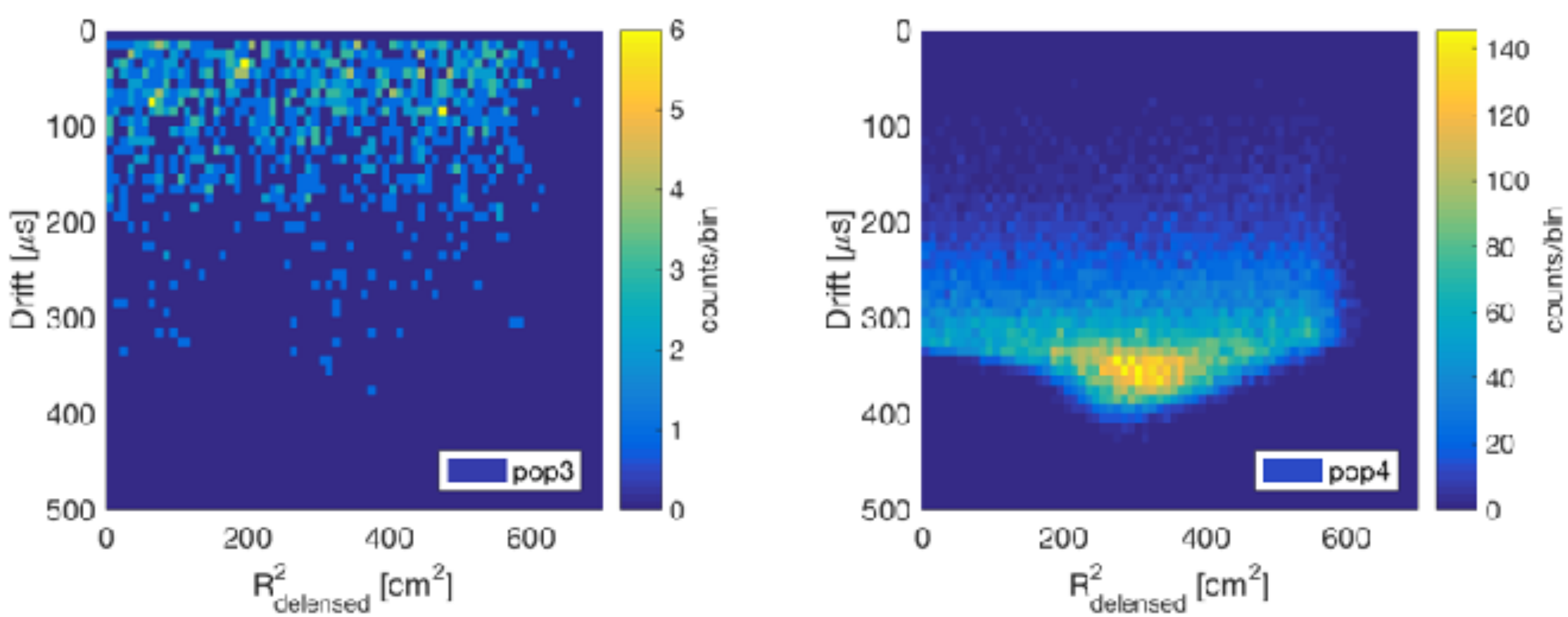




- *(upper left)* Population 1 has a uniform distribution in x-y which looks like Kr.
- *(upper center)* Population 2 is concentrated at -x near the 9:00 panel.
- *(upper right)* Population 1b is mostly uniform with an excess near the walls.
- *(lower left)* Population 3 is mostly uniform.
- *(lower right)* Population 4 is fairly uniform but may have a drift time dependence.

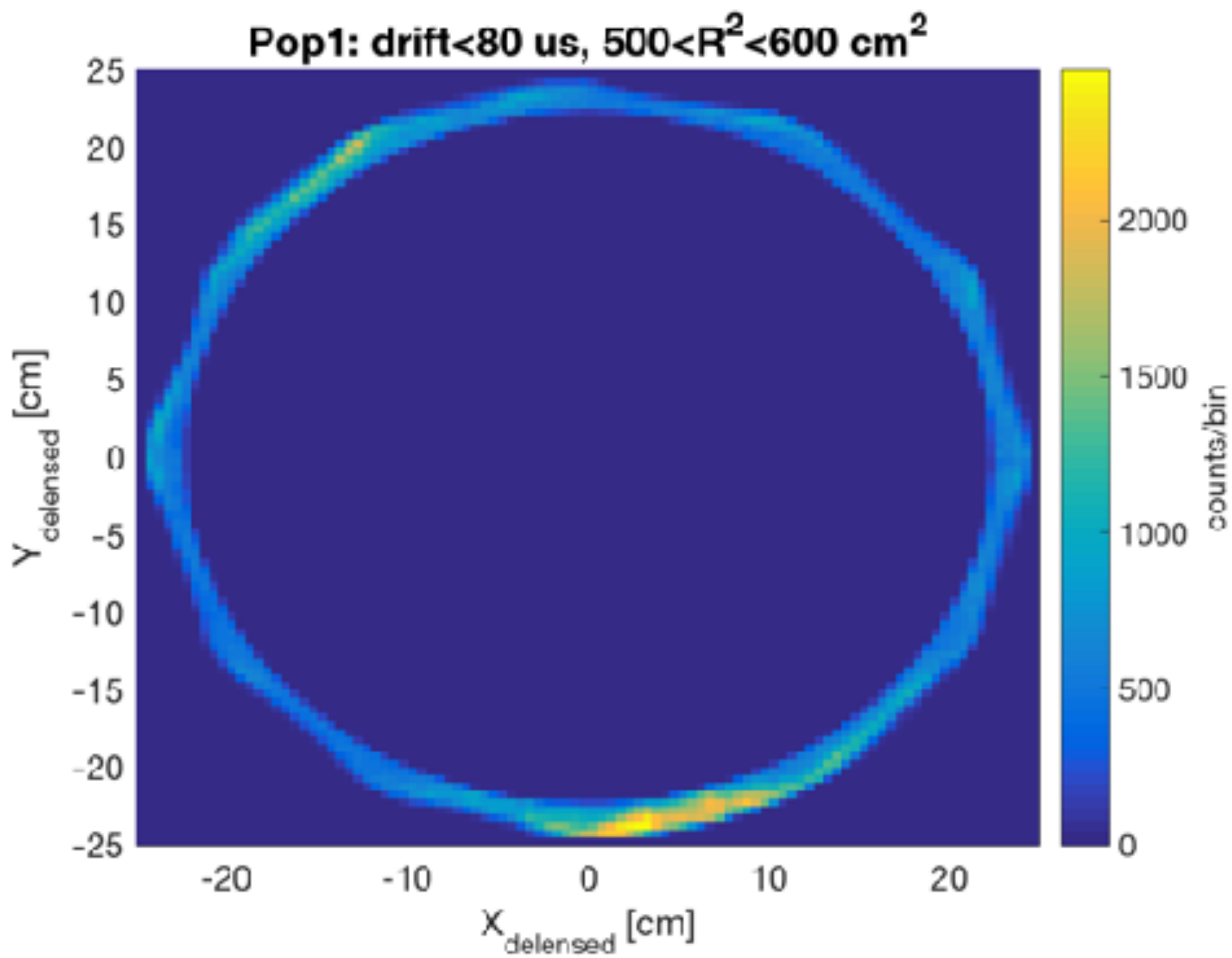
DRIFT POSITIONS OF THE POPULATIONS

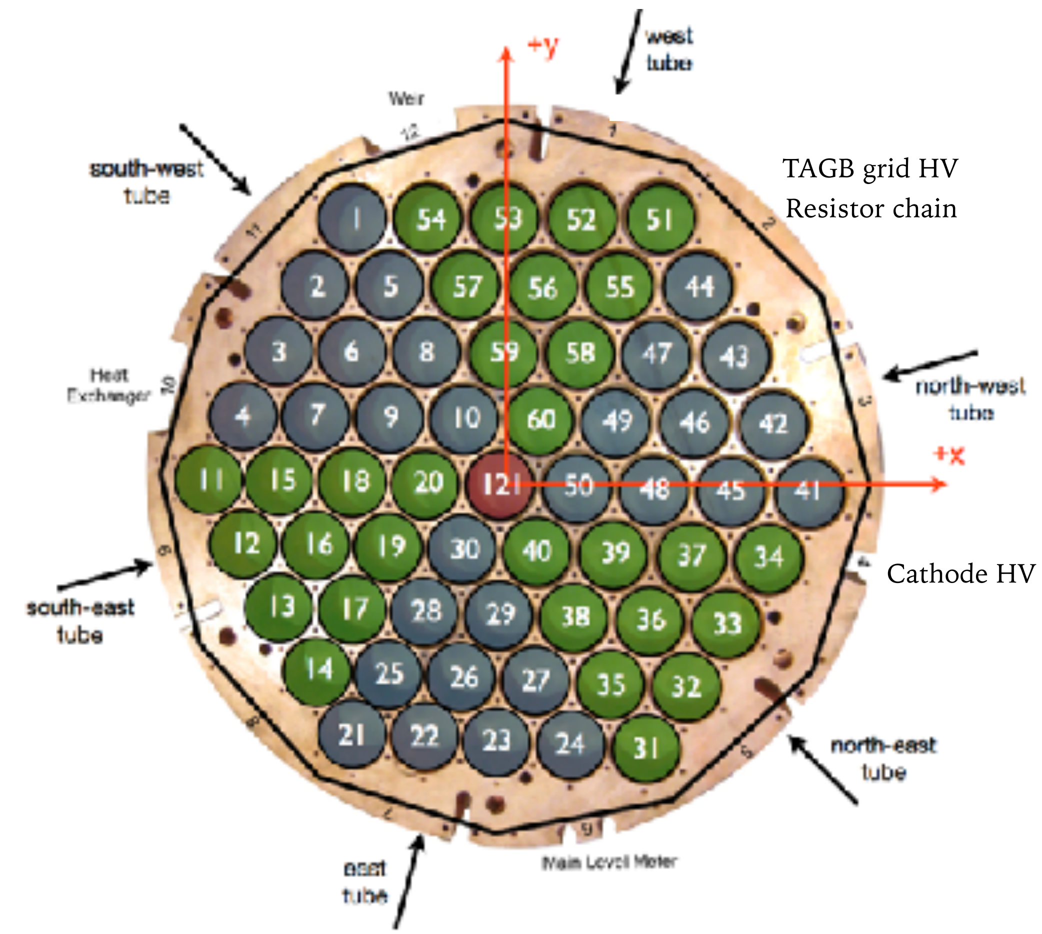




- *(upper left)* Population 1 is mostly uniform in drift vs. r^2.
- *(upper center)* Population 2 is concentrated toward the top of the detector.
- *(upper right)* Population 1b is concentrated near the bottom and walls of the detector.
- *(lower left)* Population 3 is near the top of the detector.
- *(lower right)* Population 4 is near the bottom of the detector.

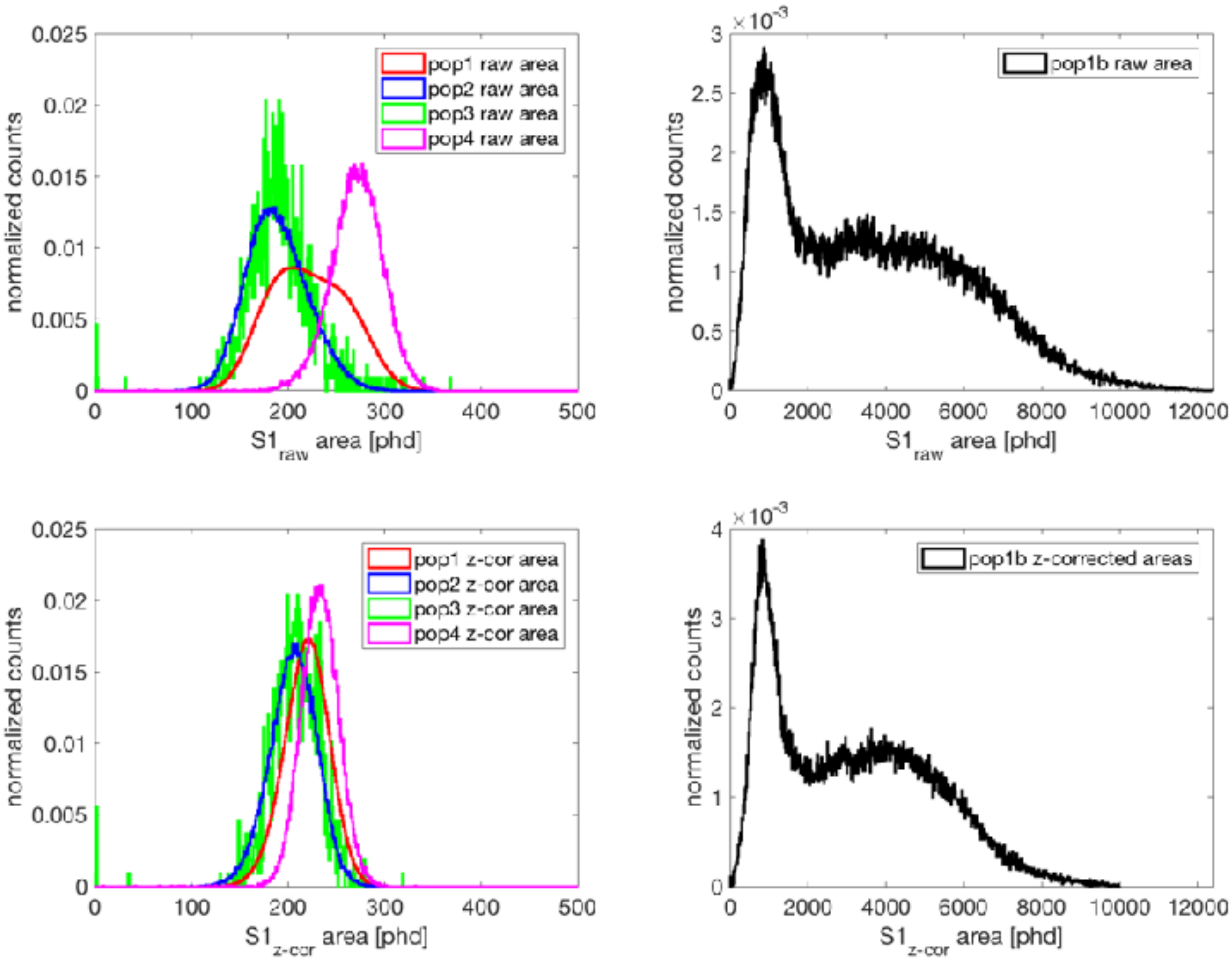
IDENTITY OF HOTSPOT IN POPULATION 1





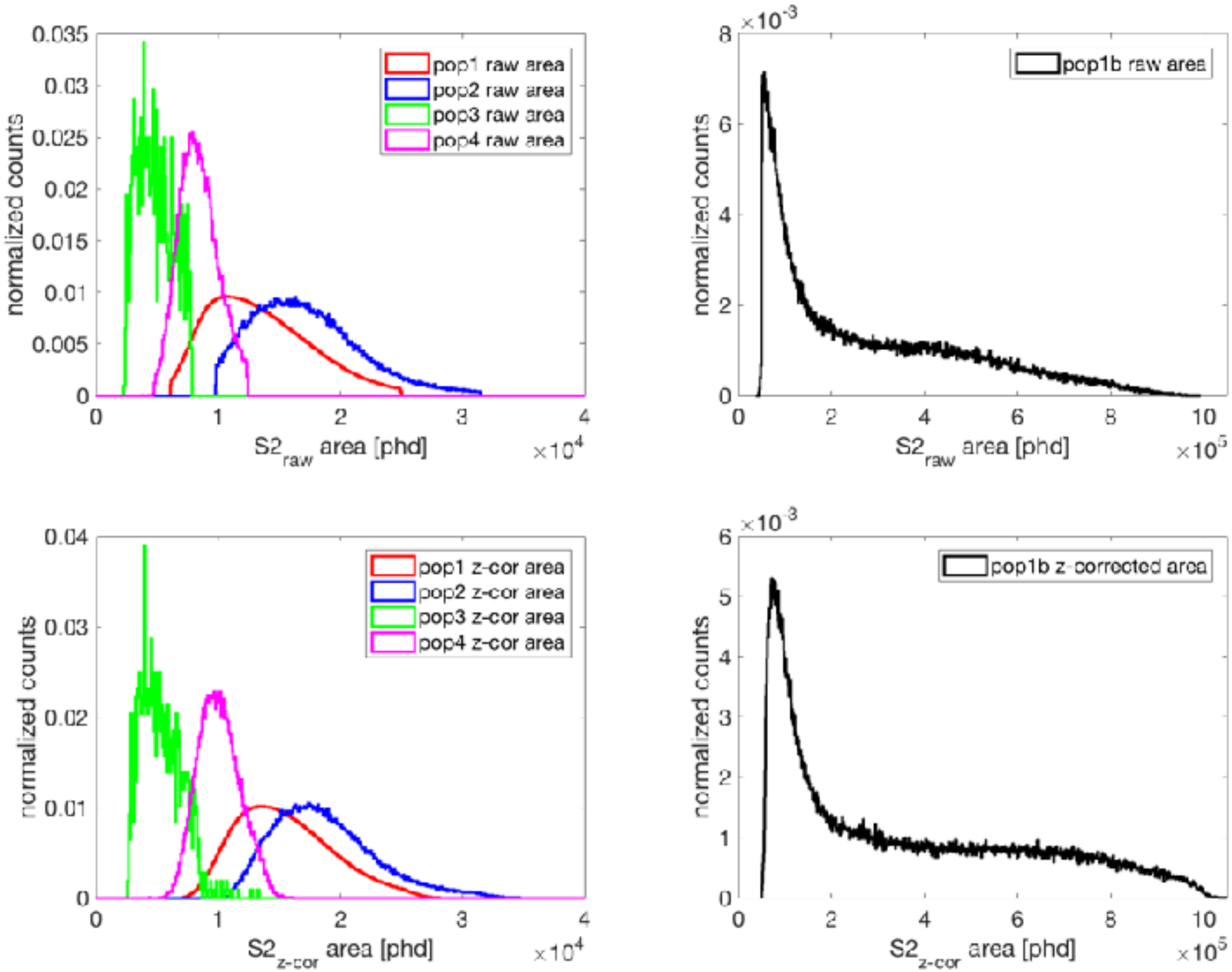
- Population 1 has a hotspot evident in drift vs. R^2.
- Cut to select this hotspot: Population 1 & (drift time < 80 μs) & (500 < R^2 < 600 cm^2)
- *(left)* Plot the x-y positions of events in this hotspot. These largely are concentrated at -y near the 6:00 panel.
- *(right)* PMT map. PMT 26 had problems during Run04; maybe this is the culprit behind high counts? Otherwise, PMTs 23, 24, 31 or some problem near them are suspect.

S1 AREAS



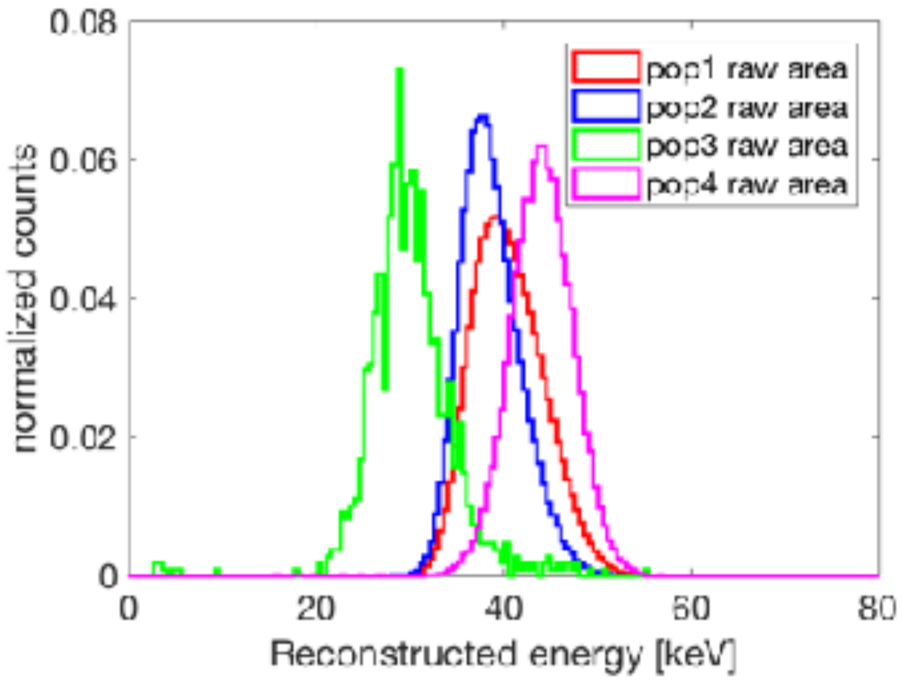
Population 1b has S1 pulse areas much larger than seen in populations 1, 2, 3, or 4.

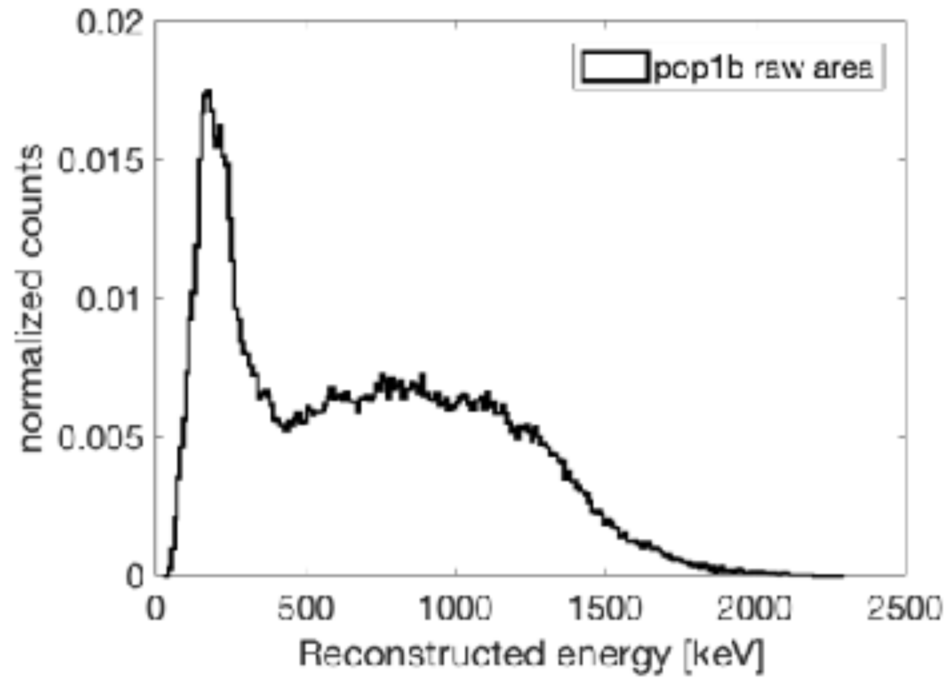
S2 AREAS



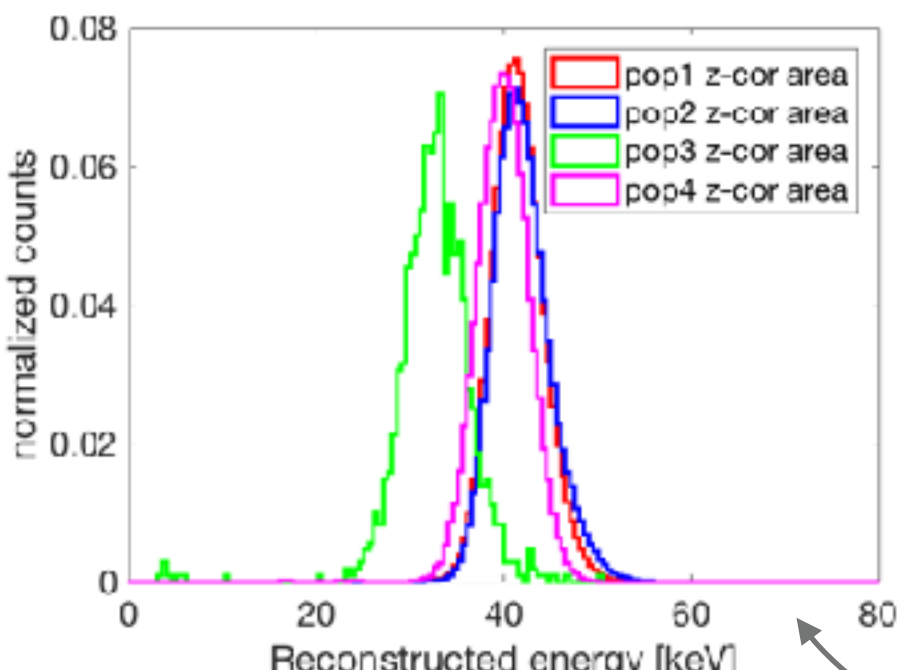
Population 1b has S2 pulse areas much larger than seen in populations 1, 2, 3, or 4.

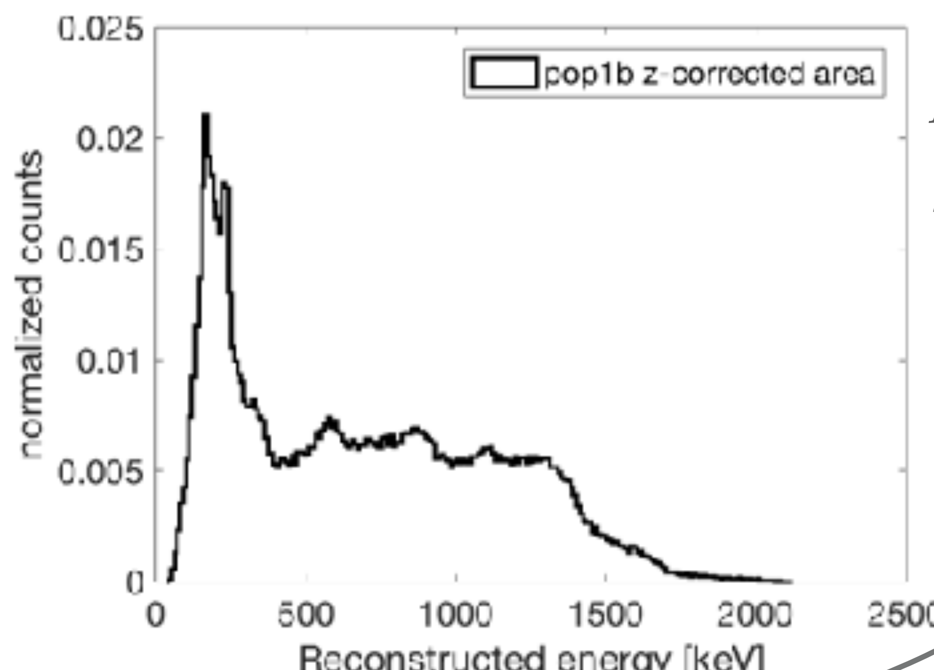
RECONSTRUCTED ENERGIES





Population 1b has reconstructed energies much larger than seen in populations 1, 2, 3, or 4.





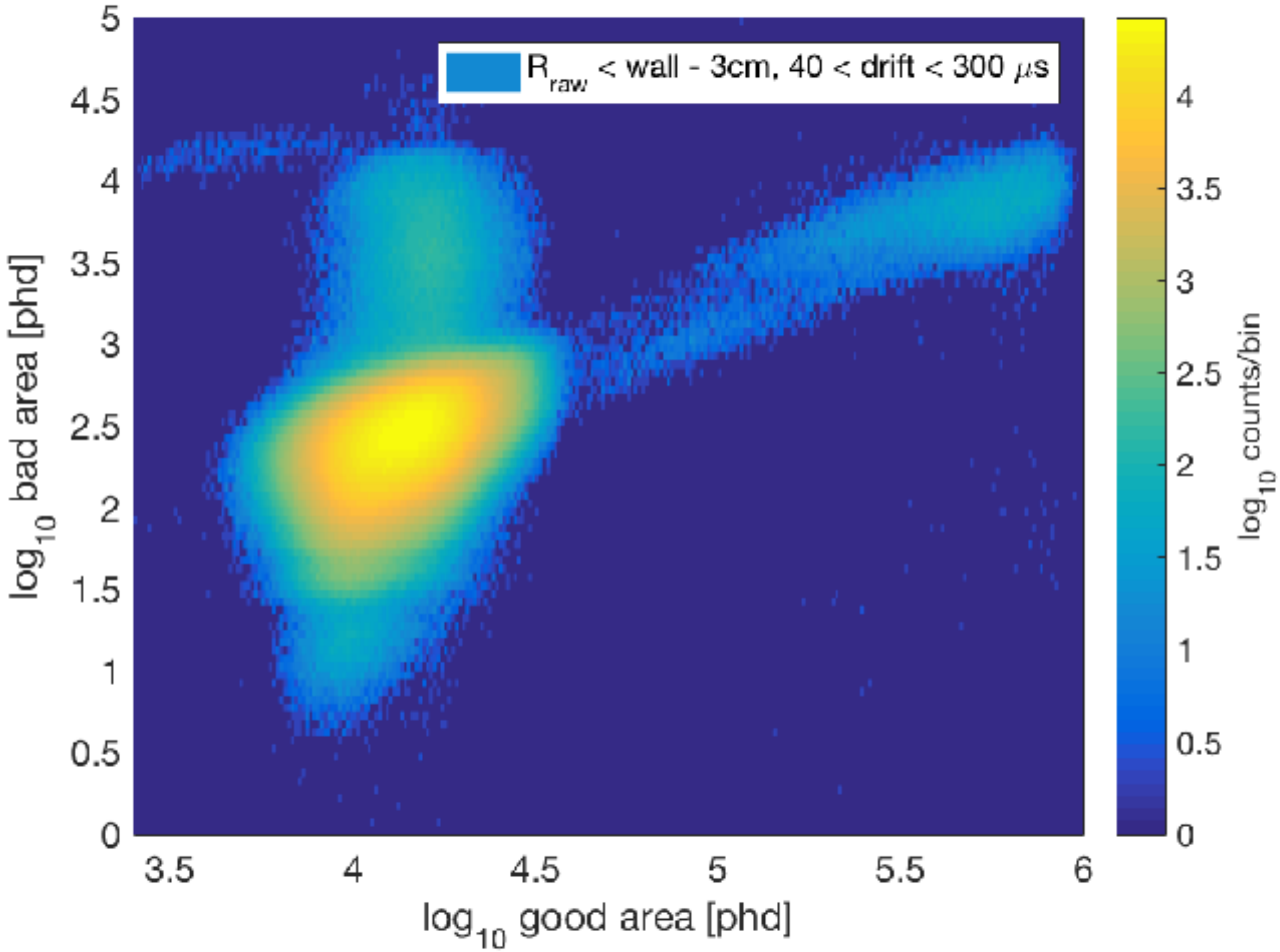
Populations 1, 2, and 4 have energies ~41 keV consistent with ^{83m}Kr.

SUMMARY OF KR-83M POPULATIONS

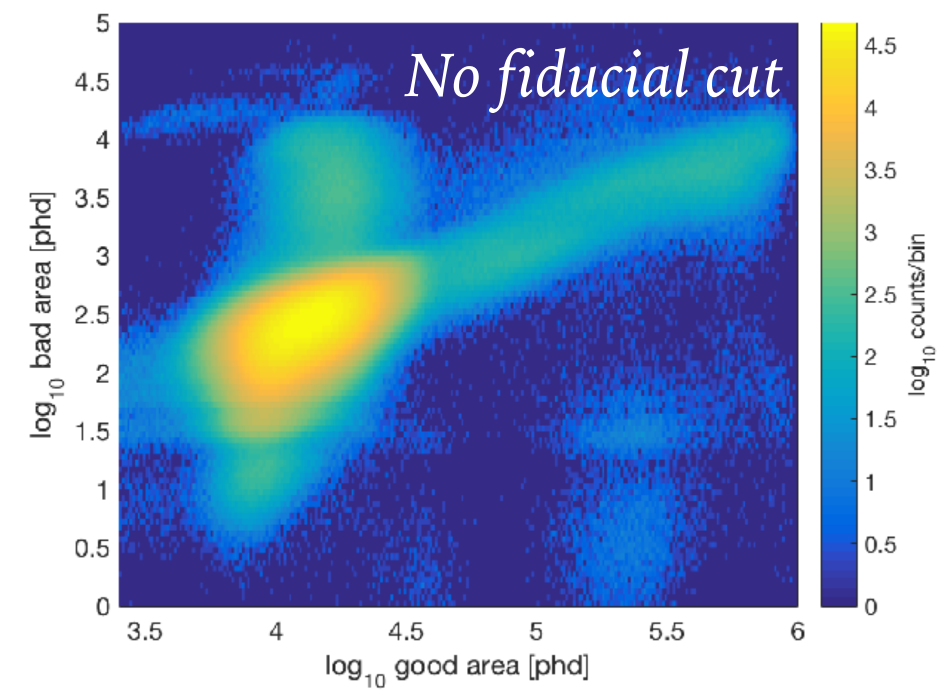
Population	Energy [keV]	XY	Drift	Event anomalies
1	~41	Uniform	Uniform	—
1b	>100	Walls	Bottom, Walls	E-trains
2	~41	-X	Top	Baseline issues
3	~33	Uniform	Top	Misclassified double-scatter
4	~41	~Uniform	Bottom	—

Population 1 has the expected energy, xy, and drift distributions of ^{83m}Kr events, and it is free of anomalies that contribute to excessive bad area. This population extending to log10(good area) = 4.6 will be used to set the bad area cut.

APPLY RUN04 FIDUCIAL CUT



➤ *(above)* Apply the Run04 fiducial cut which uses the raw radius to cut 3 cm in from the wall as calculated in the script setRMax using look-up tables. The fiducial cut also cuts on drift time: 40 < drift time < 300 μs.



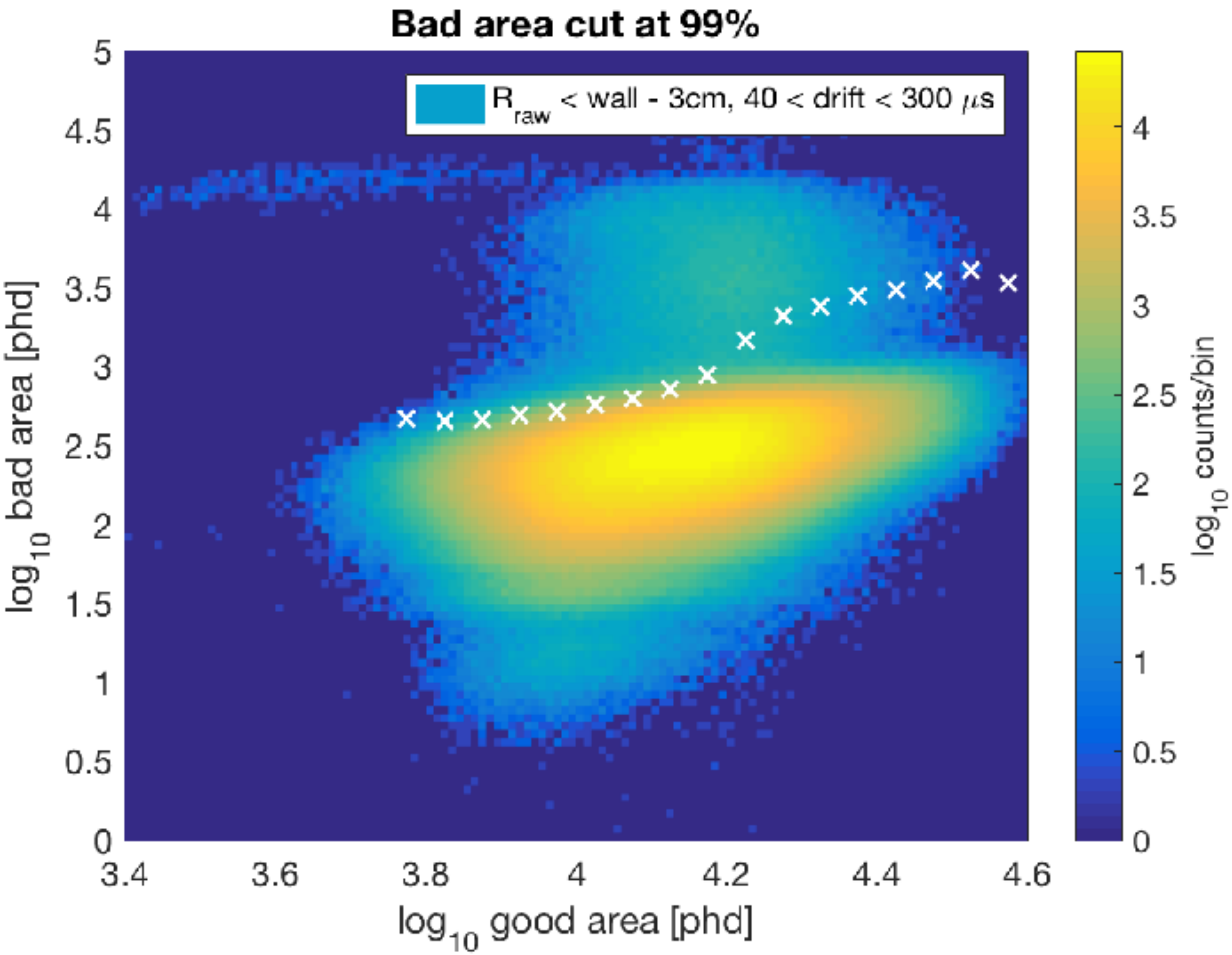
- ➤ *(above)* For comparison, this is the plot of log10(bad area) vs. log10(good area) ***without*** a fiducial cut applied.
  - ➤ Many "extra" populations that were not classified as 1, 1b, 2, 3, or 4 will disappear with a fiducial cut.
  - ➤ Many events in population 1b disappear with the fiducial cut.
  - ➤ The colorbar re-scaled, so the counts are relatively smaller with the fiducial cut.

CONSTRUCTING A NEW BAD AREA CUT WITH ^{83m}KR, FIDUCIAL CUT

1. Bin log10(good area) in the vicinity of ^{83m}Kr data distribution with fiducial cut applied *(see table)*.
2. Calculate the log10(bad area) value at which X% of the data in the bin of log10(good area) is below.
   - Initial Run04 bad area cut determined from tritium data kept **99%** (X=99) of the data within the log10(good area) bin. This only cut 1% of the events as having too much bad area.
3. Determine the best value of X (ie., what percentile to keep).
4. Fit the log10(bad area) values at X% to calculate a cut line as a function of log10(good area) and log10(bad area).

Bin	Min log10(good area)	Max log10(good area)	Counts/bin
1	3.6	3.65	205
2	3.65	3.7	761
3	3.7	3.75	3,431
4	3.75	3.8	14,027
5	3.8	3.85	49,636
6	3.85	3.9	145,712
7	3.9	3.95	346,520
8	3.95	4	659,300
9	4	4.05	976,625
10	4.05	4.1	1,206,623
11	4.1	4.15	1,324,571
12	4.15	4.2	1,311,733
13	4.2	4.25	1,096,535
14	4.25	4.3	724,735
15	4.3	4.35	375,968
16	4.35	4.4	166,485
17	4.4	4.45	69,411
18	4.45	4.5	24,112
19	4.5	4.55	5,621
20	4.55	4.6	741

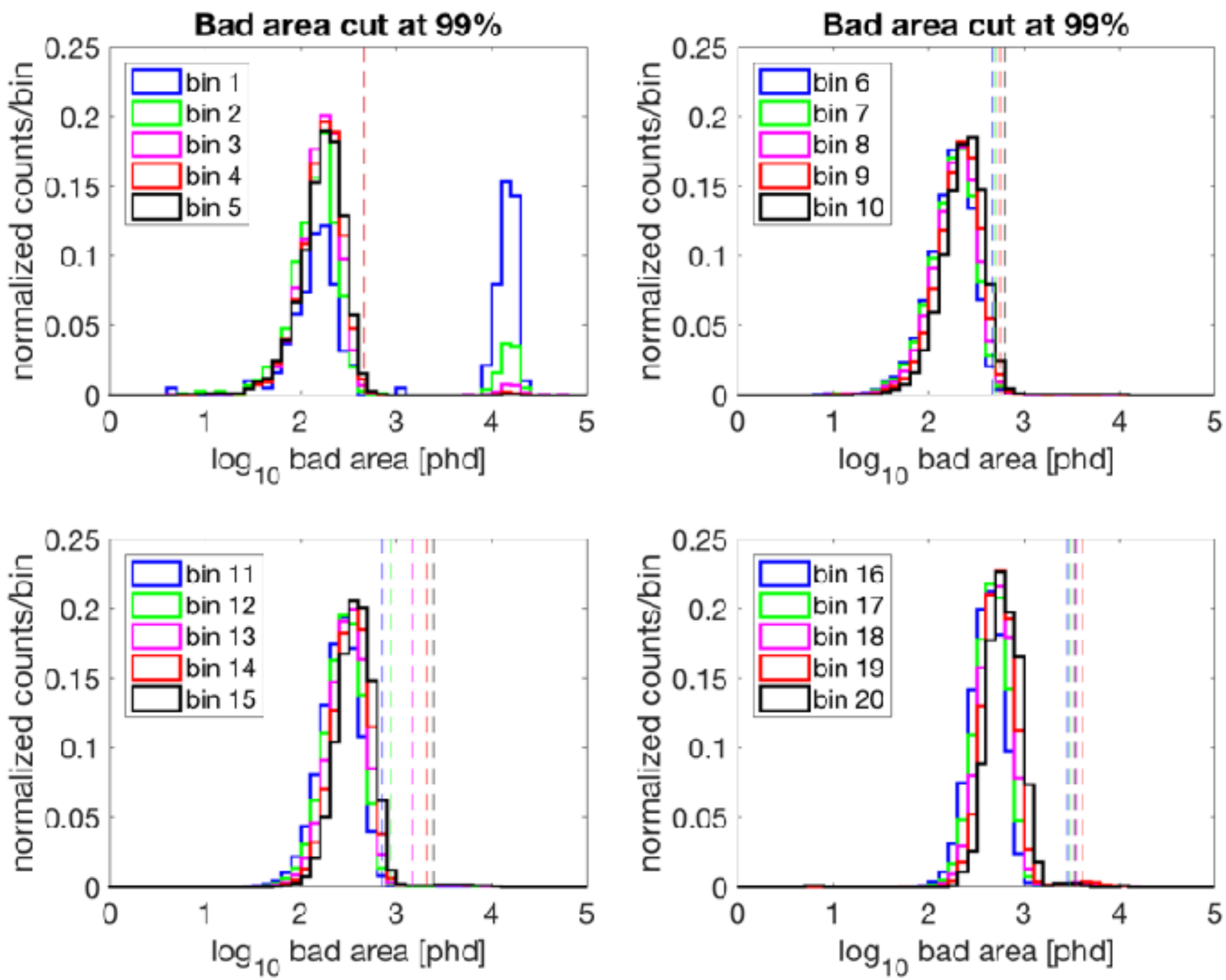
FIND THE 99% VALUE IN EACH BIN OF GOOD AREA



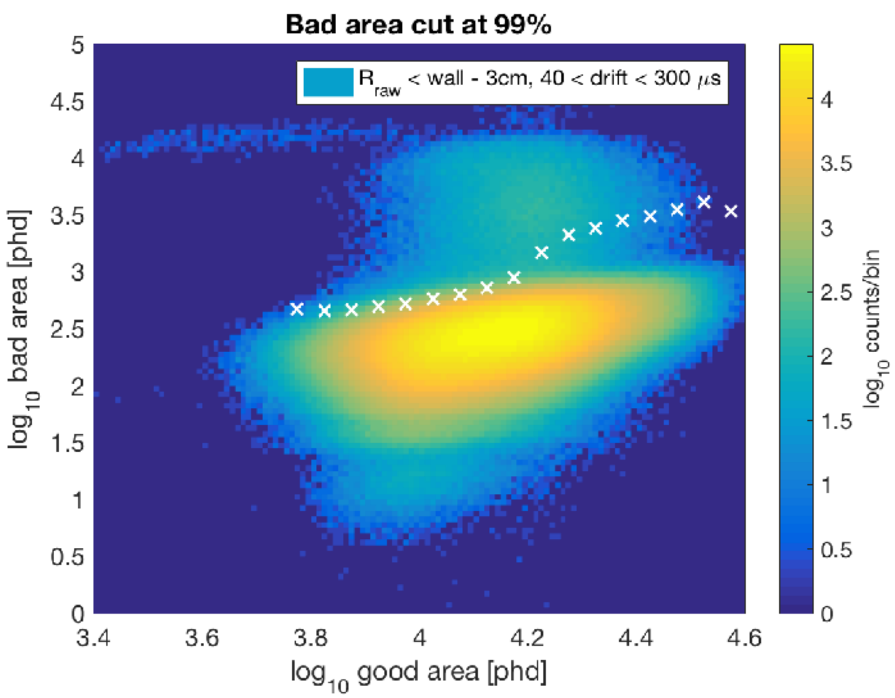
- *(left)* Find the value of bad area at which **99%** of the events in the bin of $\log_{10}$(good area) are below. This 99% value in each bin is drawn as a white X.
- Bins 1-3 do not have a 99% value shown as outliers skewed this point.
- Bins 13+ zag upward to let in bad area from population 2.

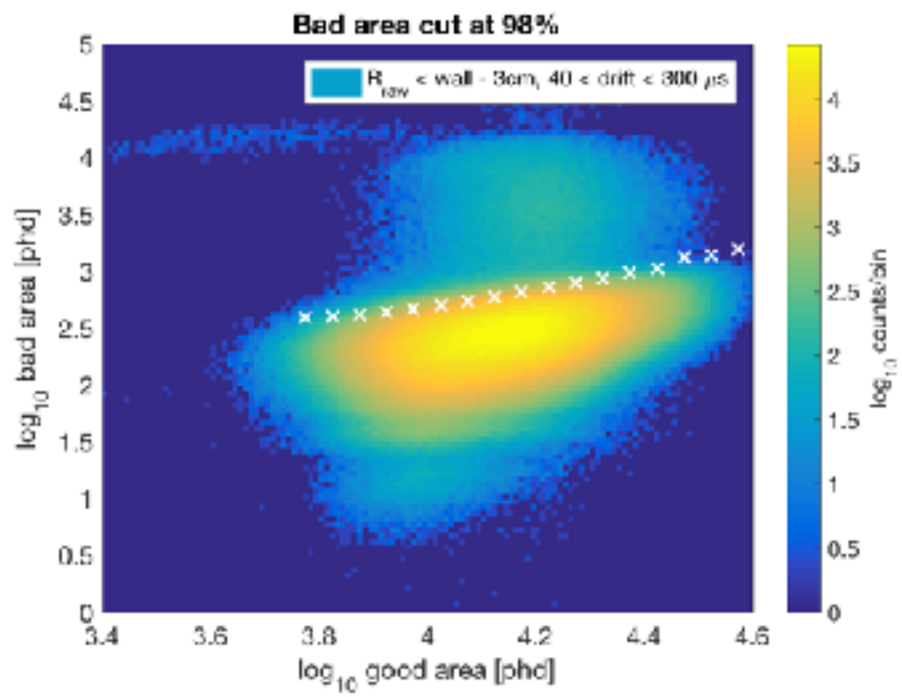
HISTOGRAM THE BAD AREA WITH THE 99% VALUES

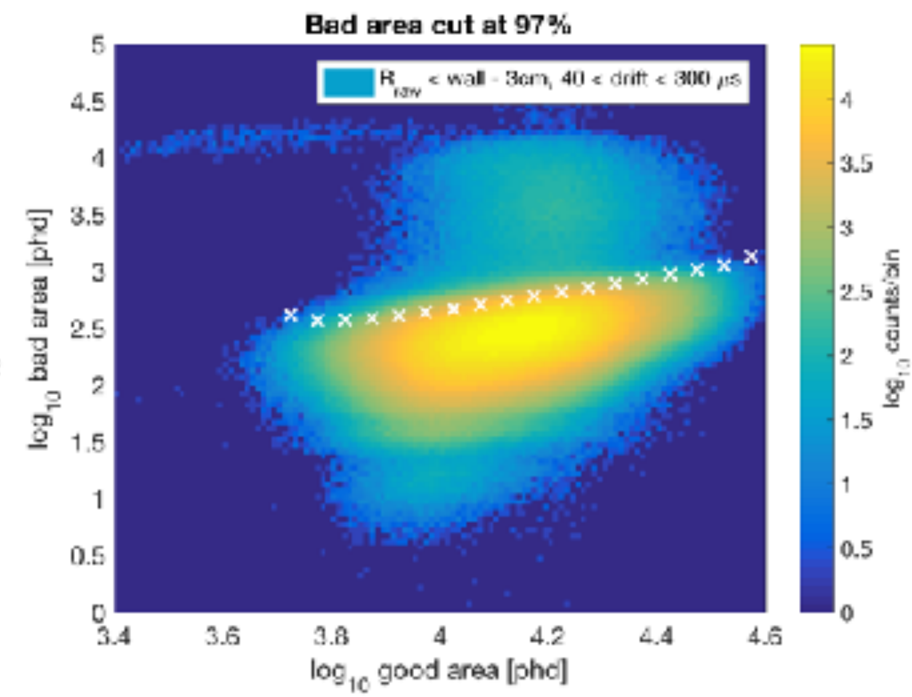
- Plot the histogram of $\log_{10}$(bad area) for each of the 20 bins in good area. The 99% $\log_{10}$(bad area) value for each bin is shown as a dashed line.
- *(top left)* The 99% values for bins 4 and 5 are shown as red and black dashed lines, respectively. Bins 1-3 obviously have large populations near bad areas of ~10,000 phd and have no dashed lines drawn.
- *(bottom right)* The 99% values for bins 16-20 are far from the main distribution due to a tail at high bad area.

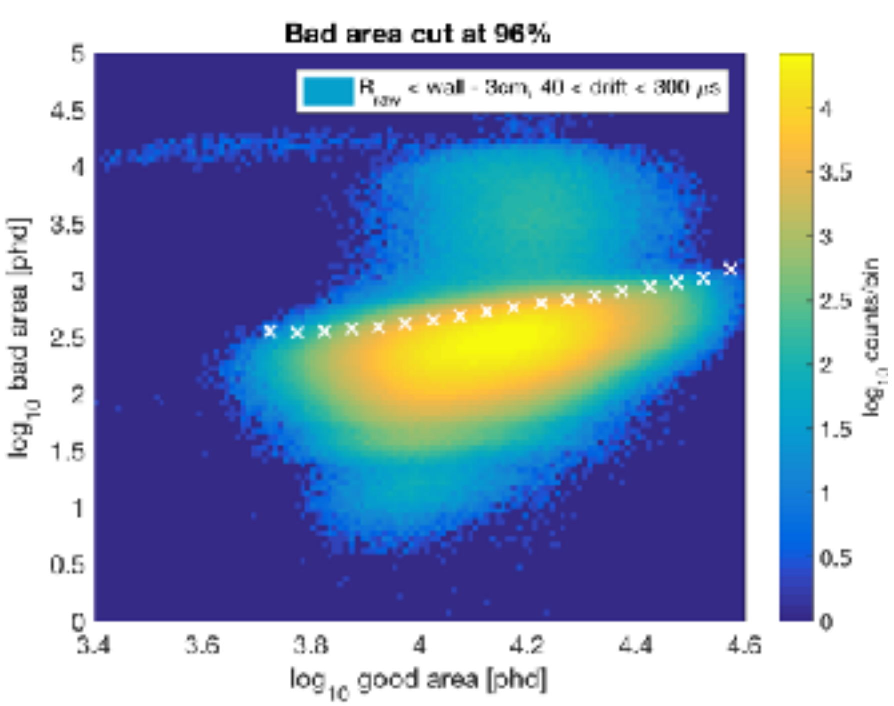


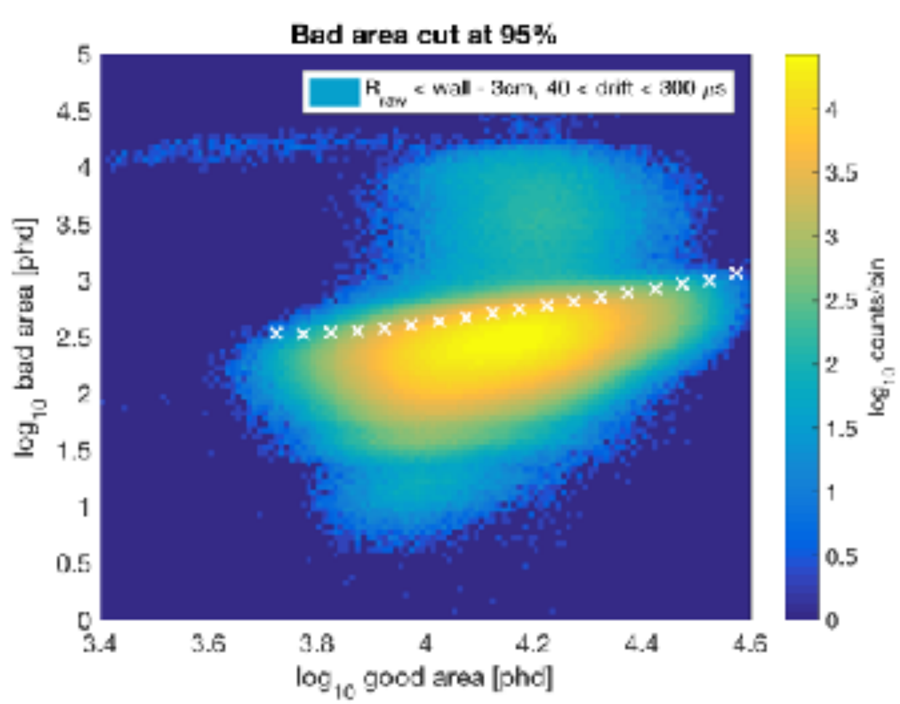
TESTING DIFFERENT X% VALUES, FIDUCIAL CUT APPLIED







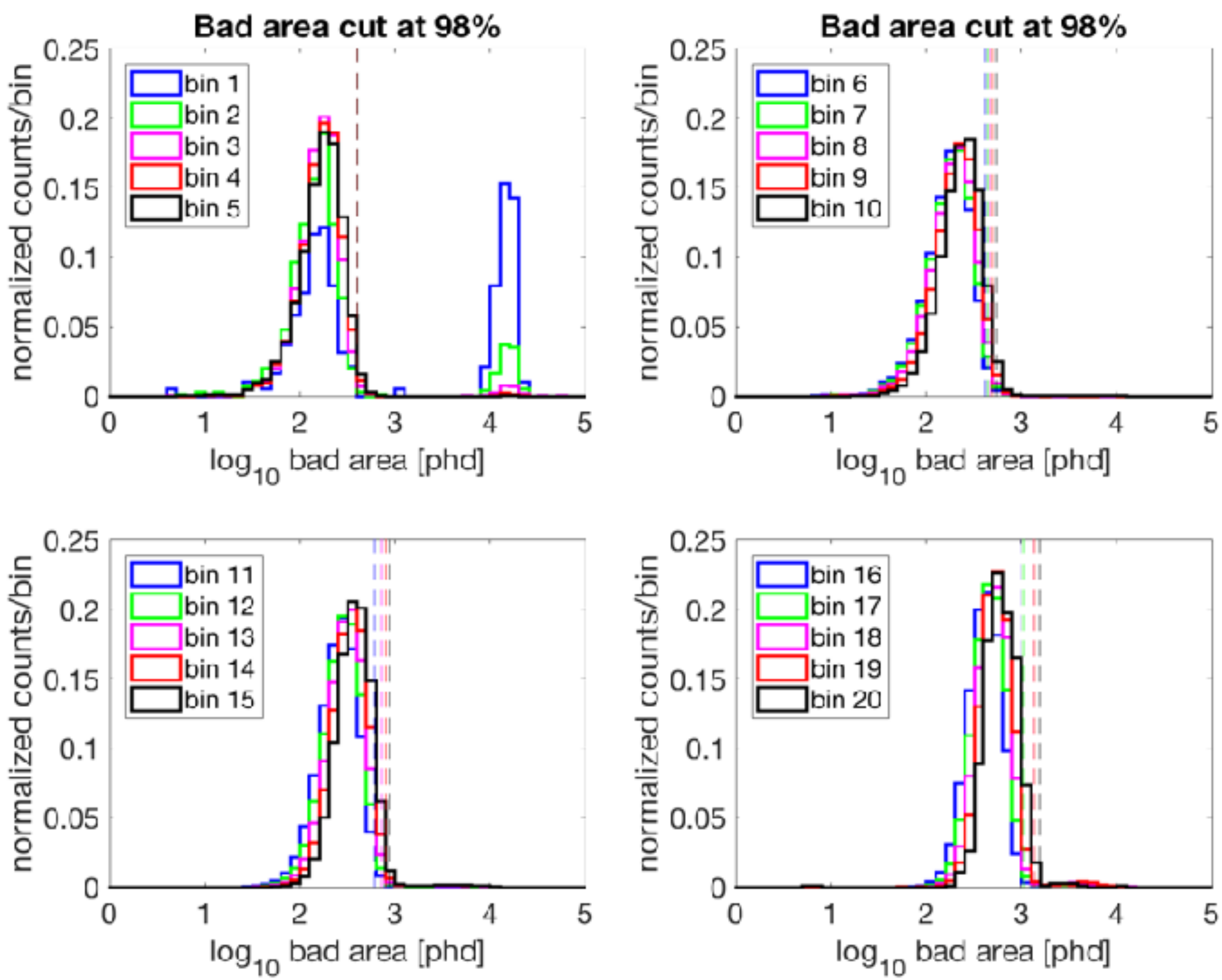




- Mark the value of $\log_{10}$(bad area) at which X% of the events in each bin are below with a white “x”.
- *(top left)* Setting the bad area cut at 99% is too stringent as it lets in events from population 2 which was plagued with baseline issues and other high bad area anomalies.
- *(top center)* Setting the bad area cut at 98% seems optimal as it does not keep events from population 2 but doesn’t cut too deeply into the main Kr-83m blob.

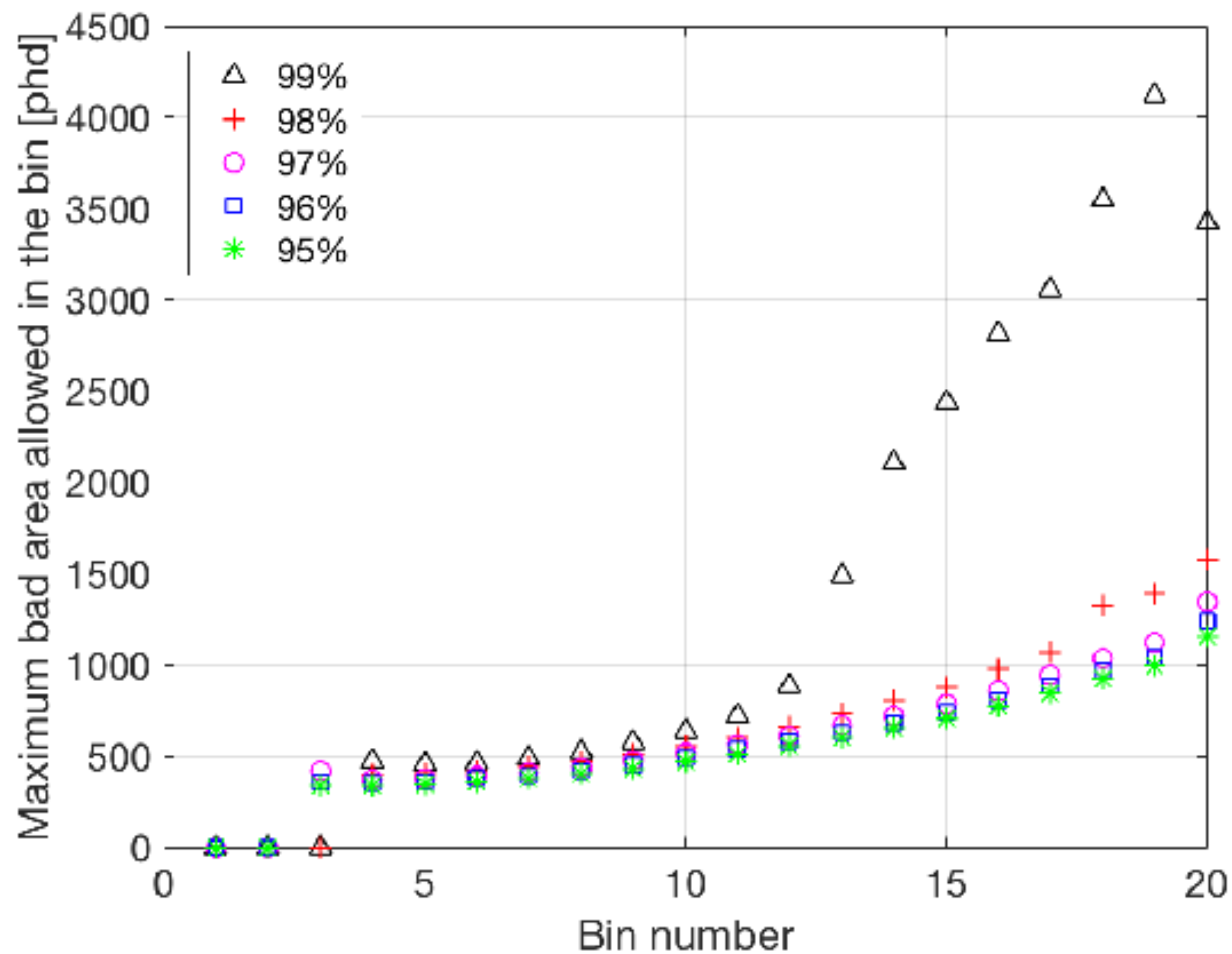
HISTOGRAM THE BAD AREA WITH THE 98% VALUES

- Plot the histogram of $\log_{10}$(bad area) for each of the 20 bins in good area. The 98% $\log_{10}$(bad area) value for each bin is shown as a dashed line.
- *(top left)* The 98% values for bins 4 and 5 are shown as red and black dashed lines, respectively. Bins 1-3 obviously have large populations near bad areas of ~10,000 phd and have no dashed lines drawn.
- *(bottom right)* The 98% values for bins 16-20 are closer to the main distribution than they were at 99%.

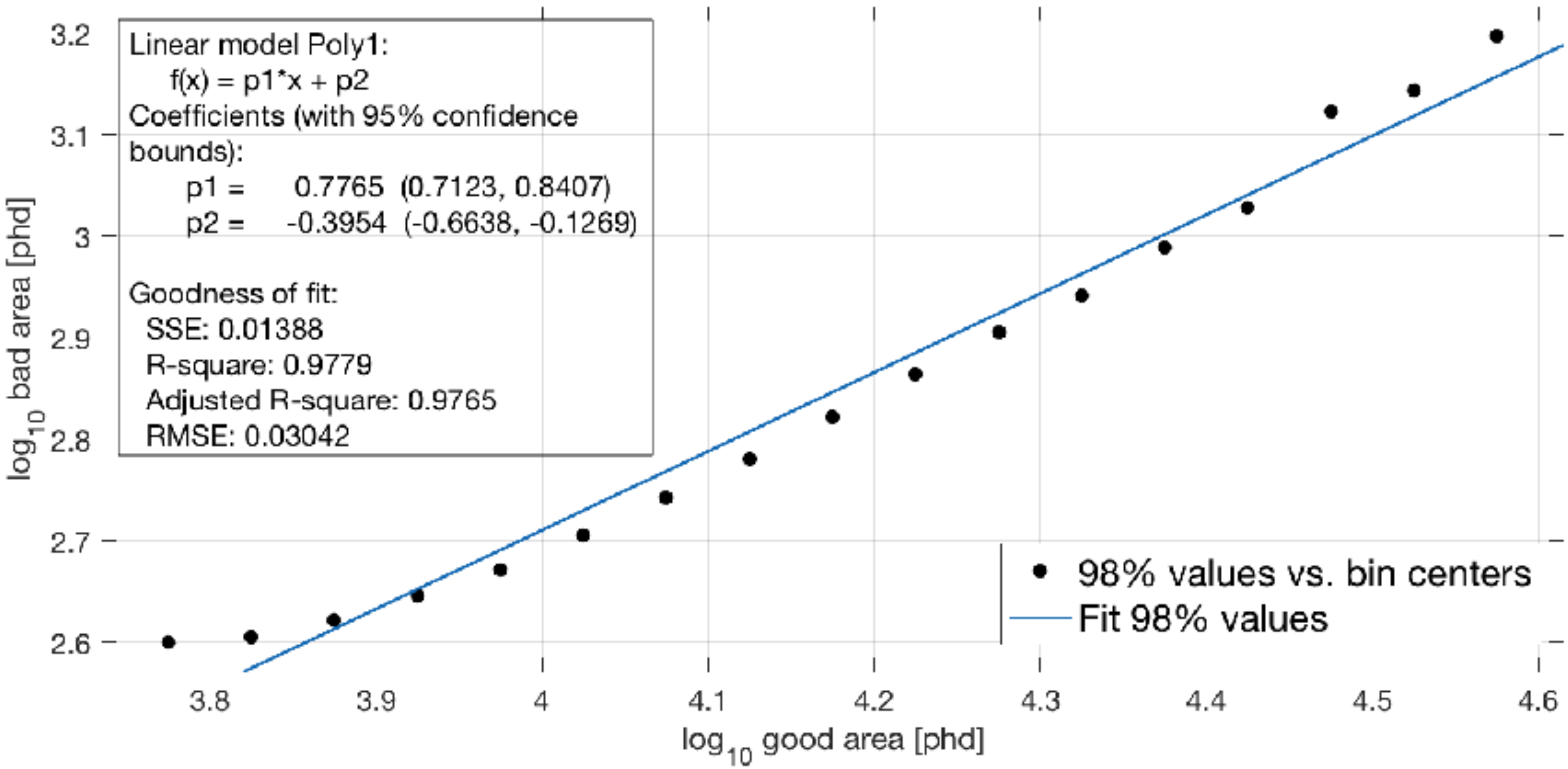


COMPARE VARIOUS X% VALUES ACROSS THE GOOD AREA BINS

- For each of the 20 bins in $\log_{10}$(good area), plot the bad area at which X% of the event in the bin are below.
  - Bins 1 and 2 have smaller stats and too many events with high bad area and are off-scale. These aren't used in the fit.
  - Bin 3 only has reasonable values for 95-97%. As setting the cut at 98% looks reasonable, bin 3 will be excluded from the fit.
- Using 99% is unfortunately too poorly behaved due to the presence of population 2 with its high bad area.



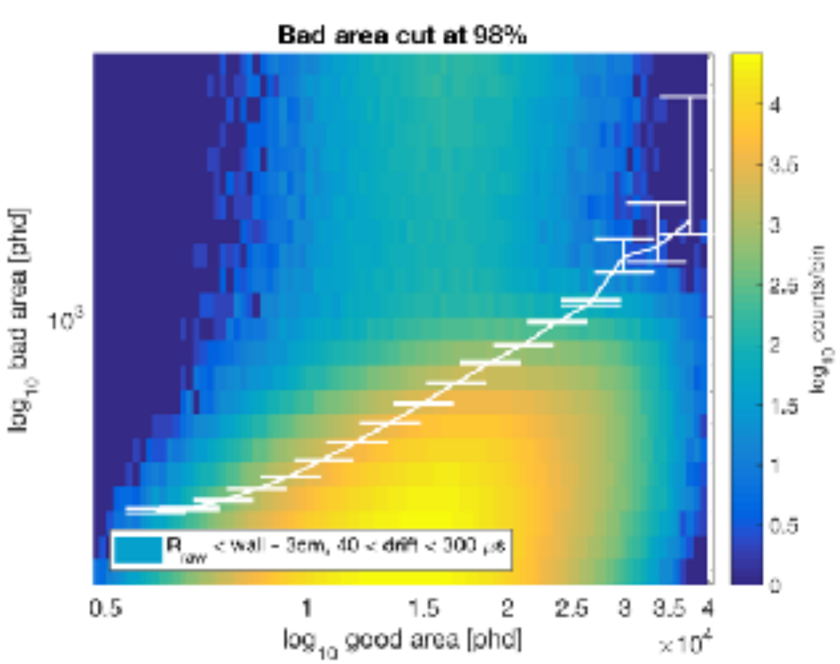
FIT 98% VALUES

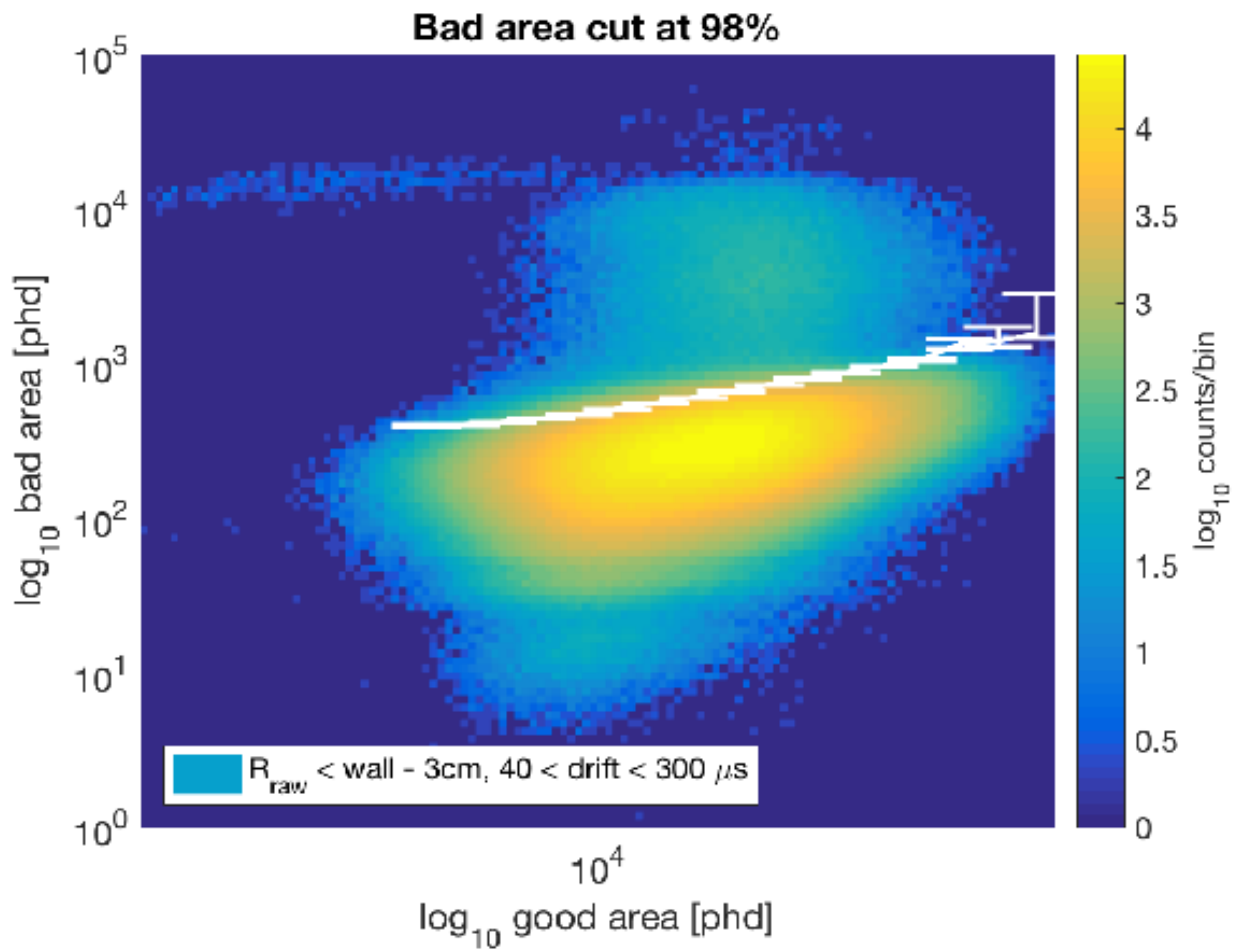


- *(above)* Apply a linear fit to the 98% values of $\log_{10}$(bad area) at the centers of the $\log_{10}$(good area) bins. The first 3 bins are excluded from the fit as there was excessive bad area.
- log10(bad area) = 0.7765*log10(good area) - 0.3954;

APPLY ERROR BARS

- Calculate error bars as follows:
  1. Calculate number of events "N" in the top 2% of each bin.
  2. Sort the bad areas in each bin.
  3. Find the indices sqrt(N) above and below the 98% value's index.
  4. Calculate lower(upper) error bar as the difference between the bad area at the 98% value *and* the bad area at an index sqrt(N) below(above) the index of the 98% value.





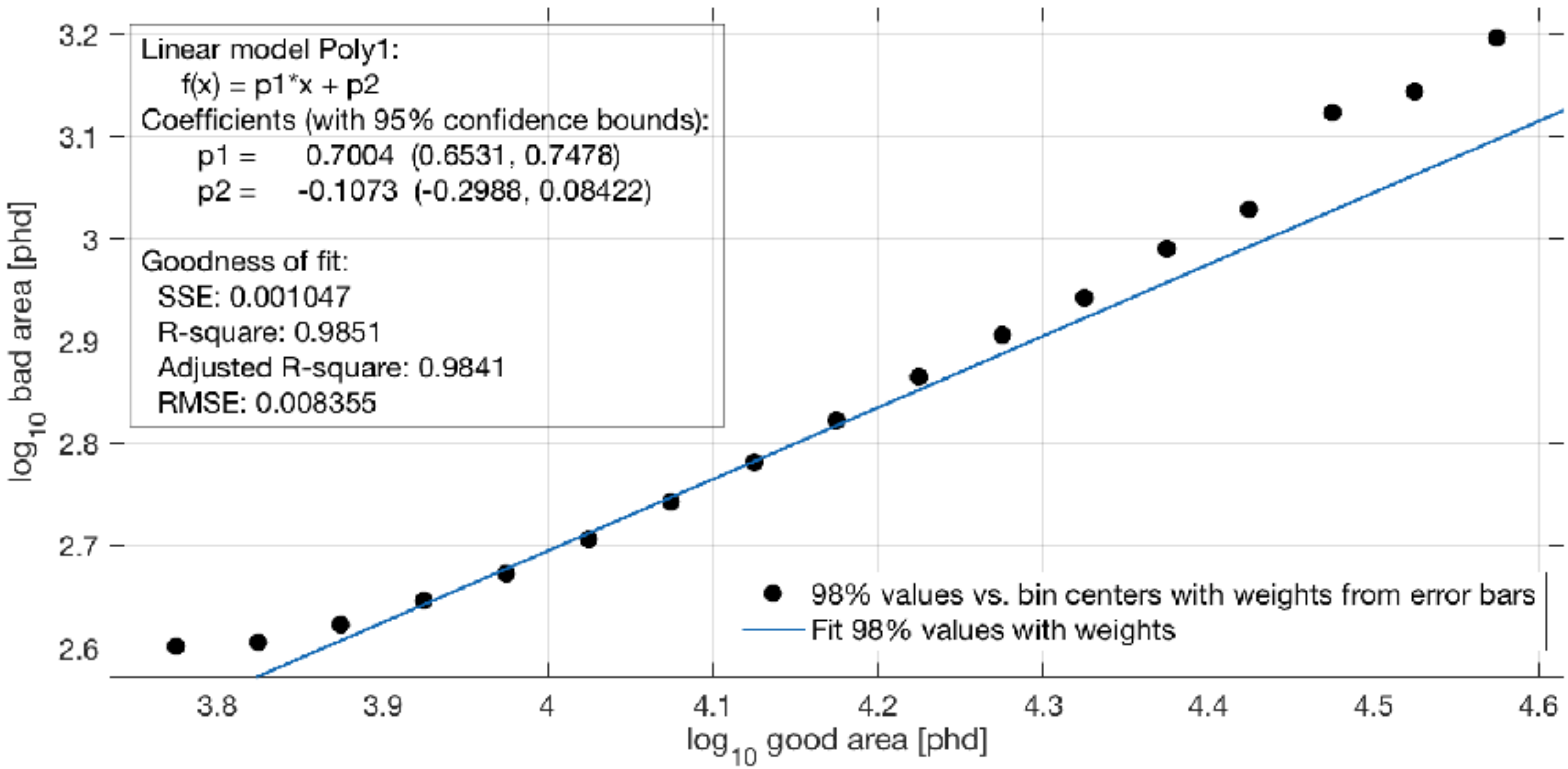
(above) Plot the Kr-83m data with the upper and lower error bars. At large values of $\log_{10}$(good area), the statistics get poorer so outliers become more significant.

(left) Same as above, but zoomed in to see error bars more clearly.

UPPER AND LOWER ERROR BARS FOR 98%

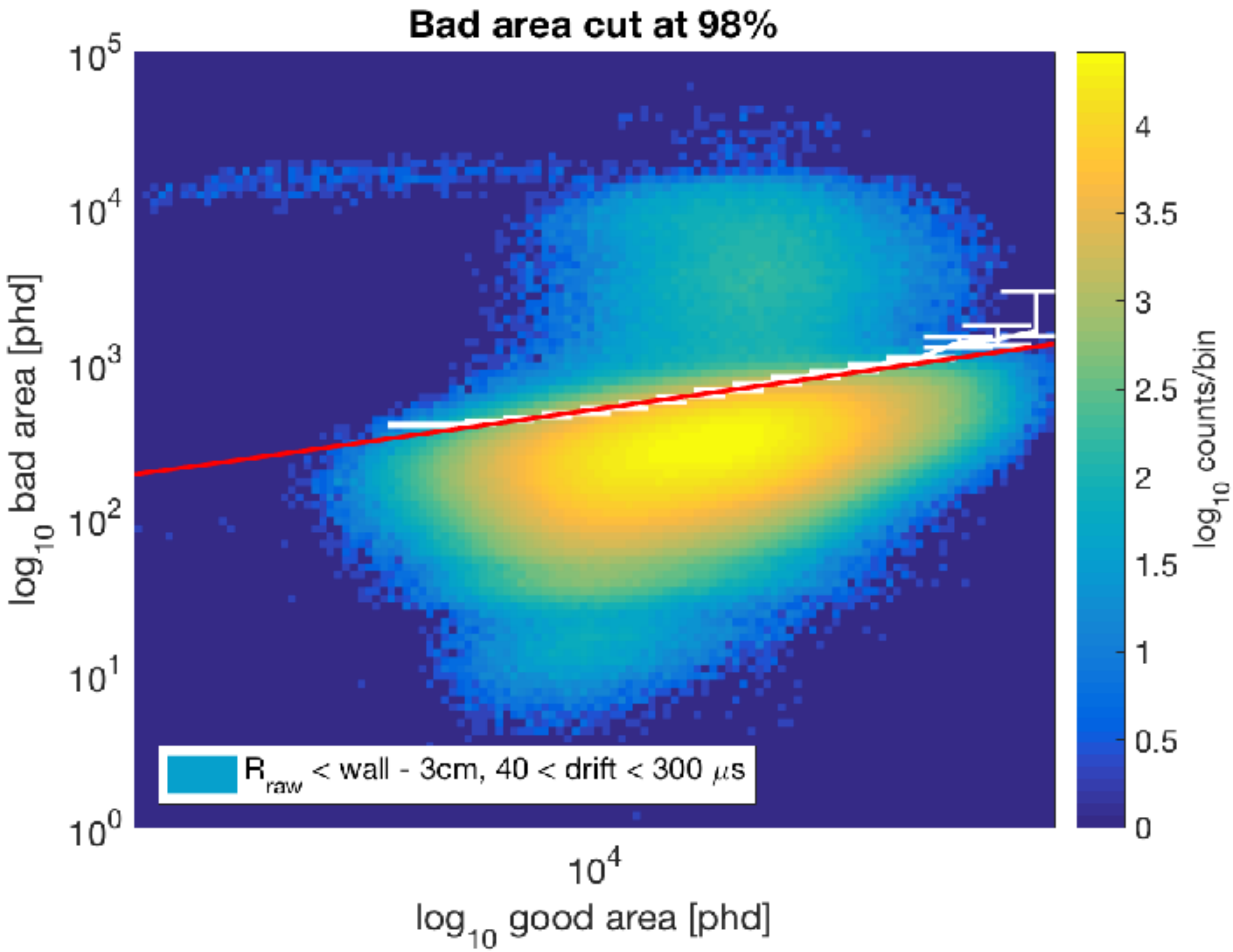
Bin	Min log10(good area)	Max log10(good area)	Counts/bin	Lower error bar [phd]	Upper error bar [phd]
1	3.6	3.65	205	-	-
2	3.65	3.7	761	-	-
3	3.7	3.75	3,431	-	-
4	3.75	3.8	14,027	4.76	4.58
5	3.8	3.85	49,636	2.12	2.47
6	3.85	3.9	145,712	1.14	1.34
7	3.9	3.95	346,520	0.76	0.90
8	3.95	4	659,300	0.61	0.56
9	4	4.05	976,625	0.68	0.55
10	4.05	4.1	1,206,623	0.58	0.56
11	4.1	4.15	1,324,571	0.77	0.69
12	4.15	4.2	1,311,733	0.85	1.03
13	4.2	4.25	1,096,535	1.07	1.38
14	4.25	4.3	724,735	1.91	2.55
15	4.3	4.35	375,968	3.57	4.35
16	4.35	4.4	166,485	7.42	7.28
17	4.4	4.45	69,411	10.49	15.77
18	4.45	4.5	24,112	85.37	119.39
19	4.5	4.55	5,621	94.75	326.28
20	4.55	4.6	741	93.42	1267.64

FIT 98% VALUES, WEIGHTS APPLIED



- *(above)* Apply a linear fit to the 98% values of $\log_{10}$(bad area) at the centers of the $\log_{10}$(good area) bins weighted using error bars. The first 3 bins are excluded from the fit as there was excessive bad area.
  - Weights = $([\text{upper + lower error bar}]/2)^{-2}$ = $(\text{average error bar length})^{-2}$
- Bad area cut at high areas: $\log_{10}(\text{bad area}) = 0.7004*\log_{10}(\text{good area}) - 0.1073$;

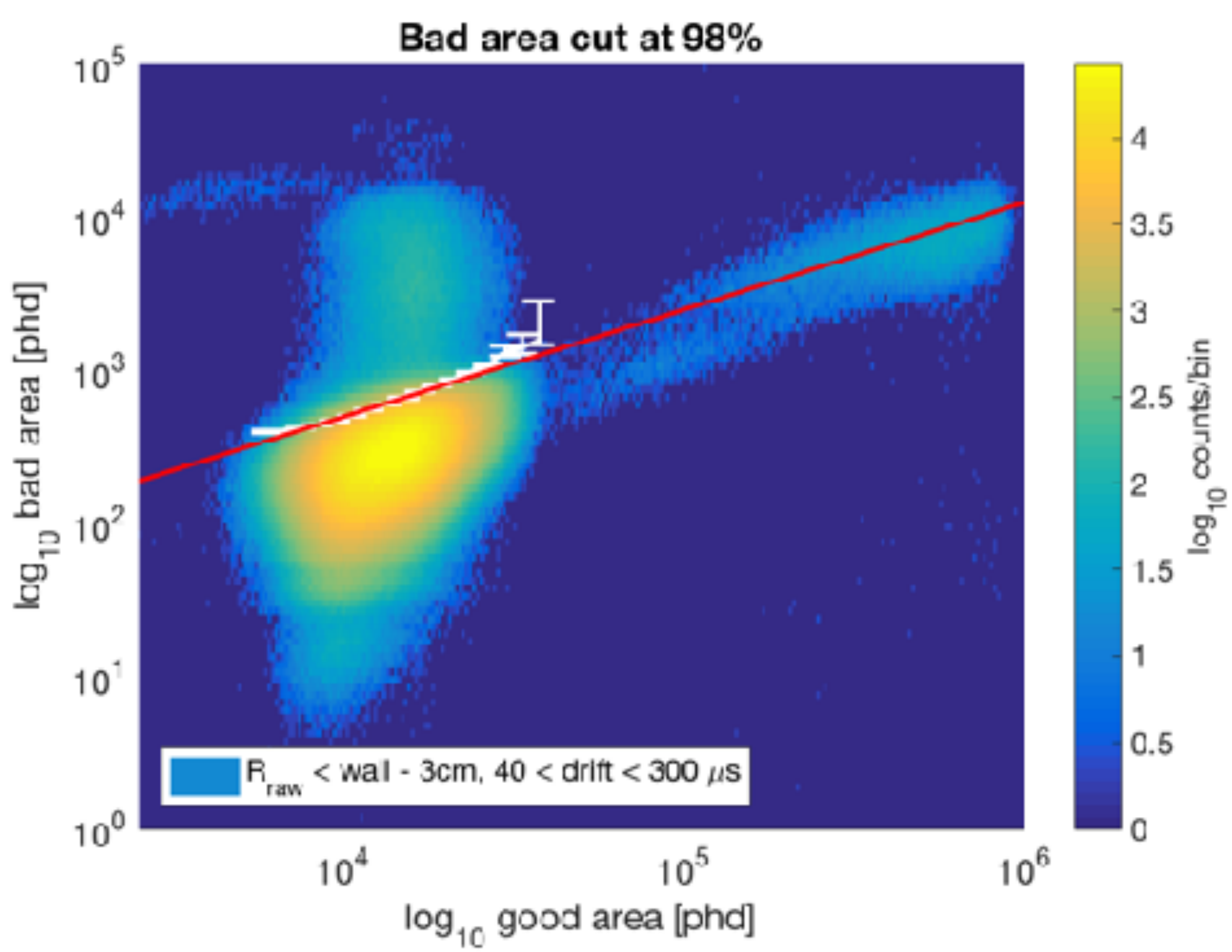
APPLY BAD AREA CUT TO THE KR-83M DATA



$\log_{10}$(bad area) = 0.7004*$\log_{10}$(good area) - 0.1073

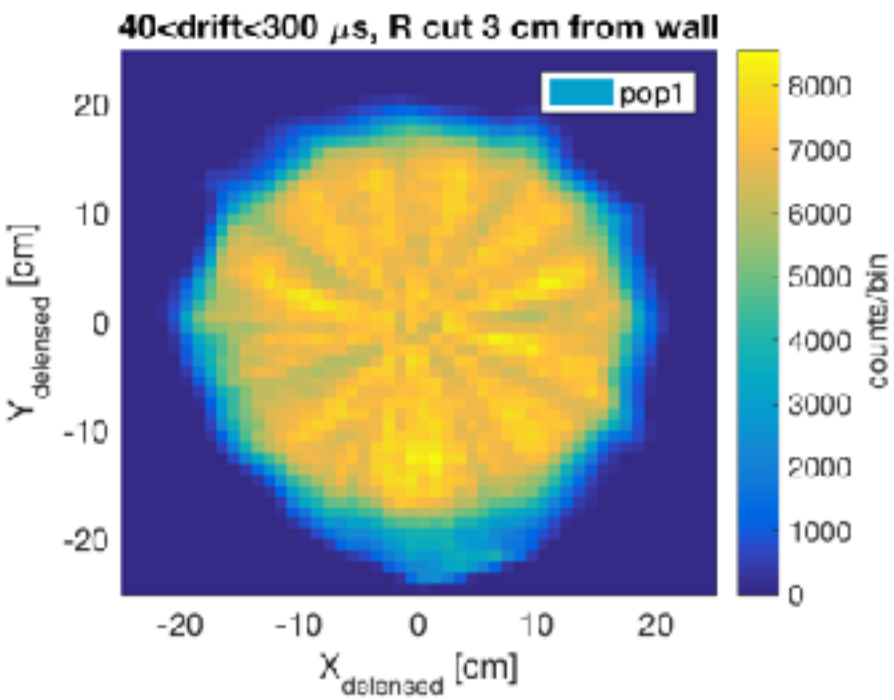
NEXT STEPS

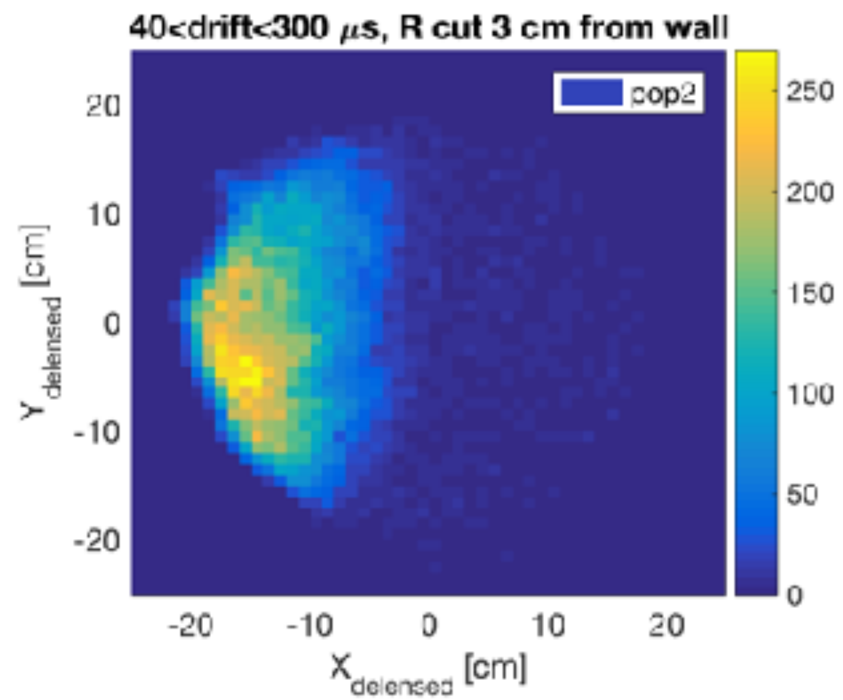
- Apply the bad area cut to the Run04 background data.
- Handle very high areas beyond main 83mKr population.
  - *(below)* Extend x-axis to 10^6 phd where the bad area cut bisects the population 1b letting in too much bad area.
  - Either decide to restrict this bad area cut to only be relevant up to $10^{4.6}$ phd or change the slope again.

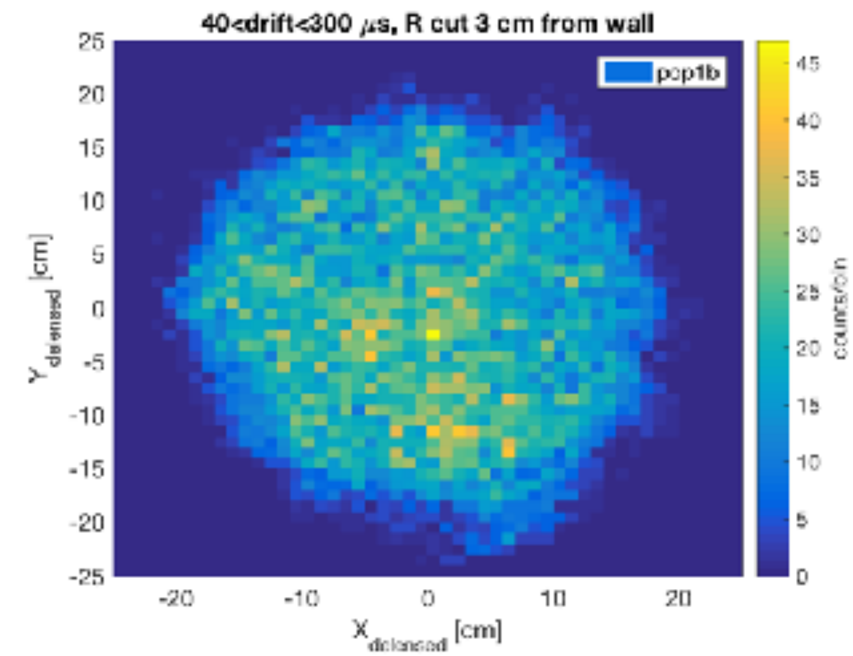


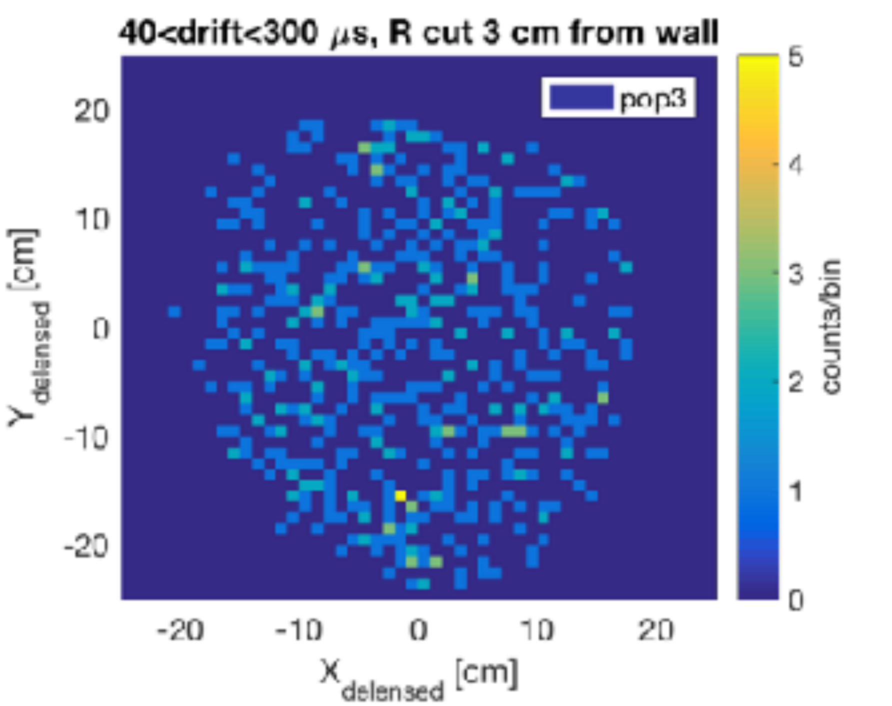
EXTRA SLIDES

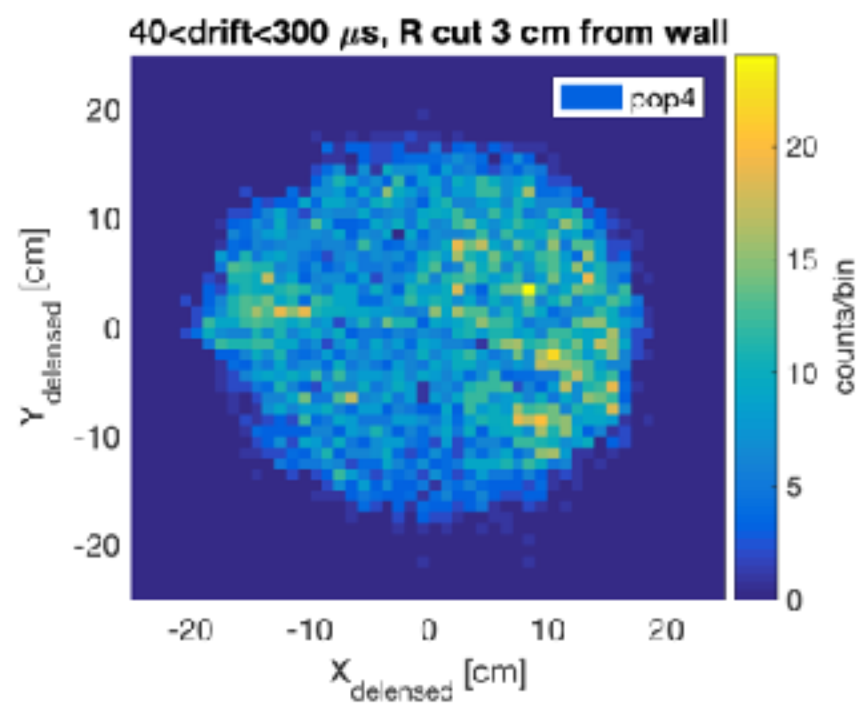
XY POSITIONS OF POPULATIONS: FIDUCIAL CUT APPLIED





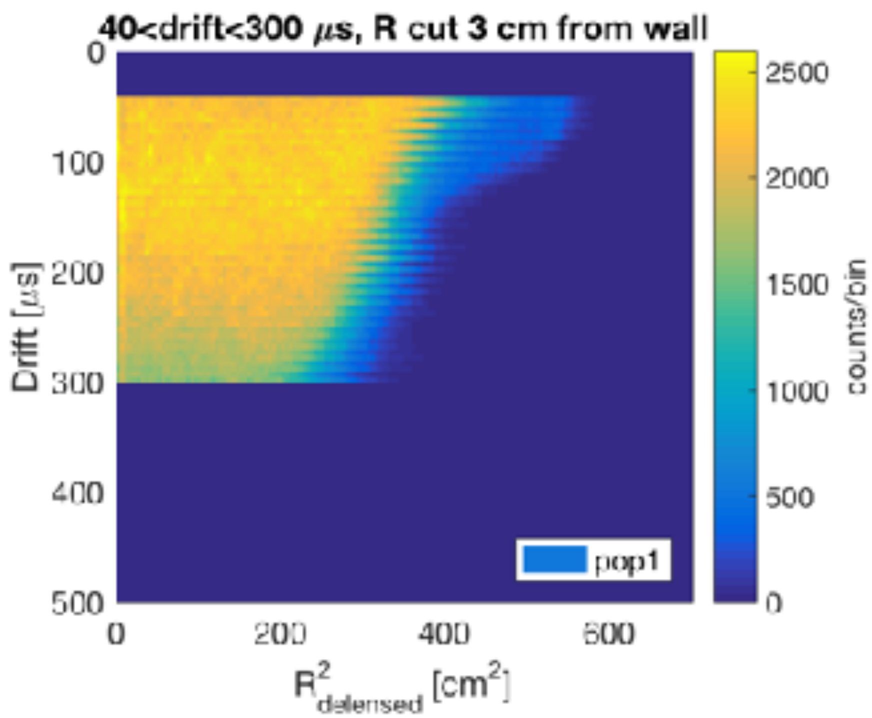


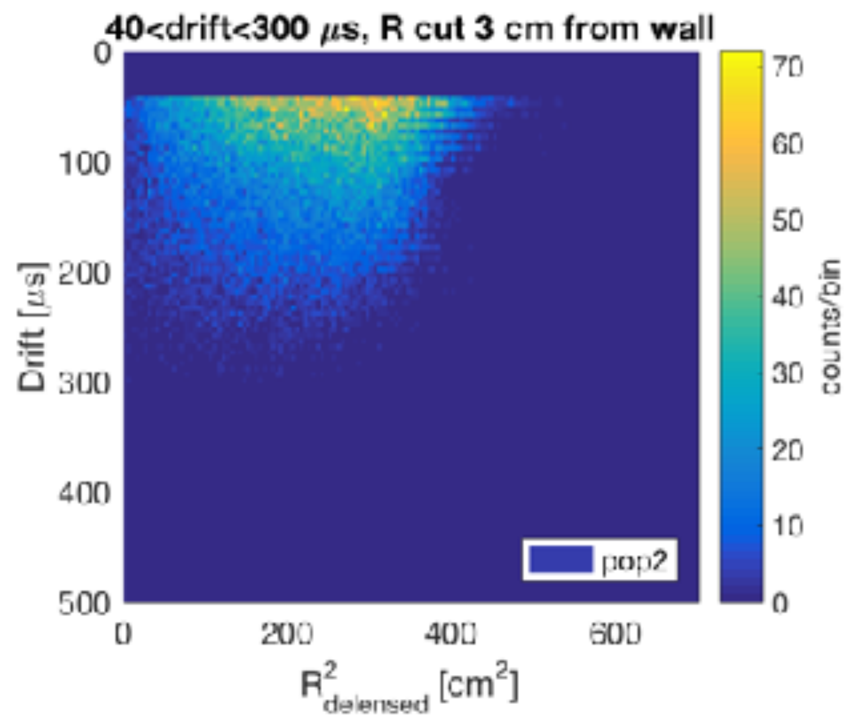


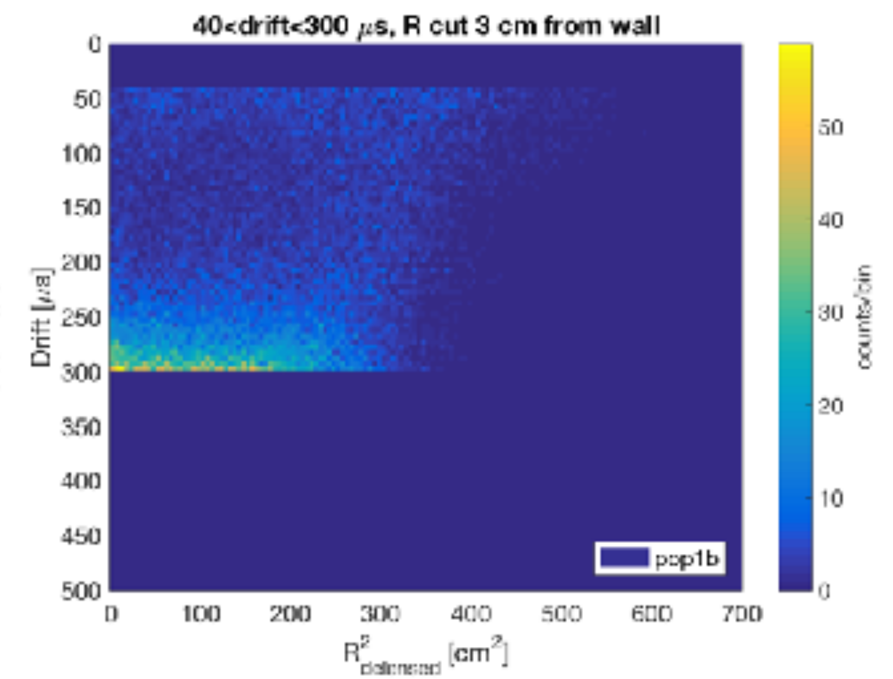


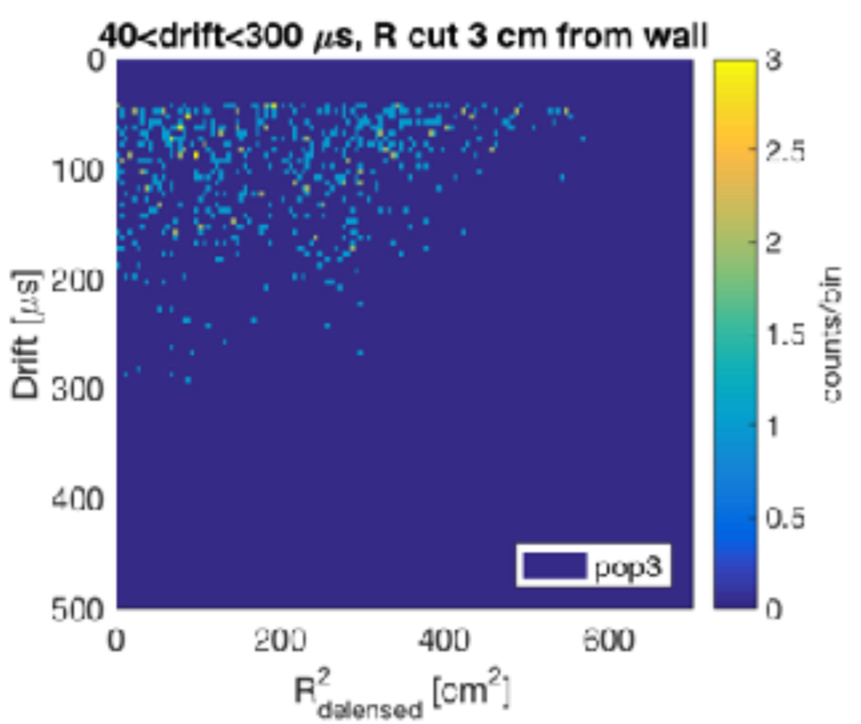
- Applying a fiducial cut results in fairly uniform distributions for all but population 2 which still has a concentration at -x.

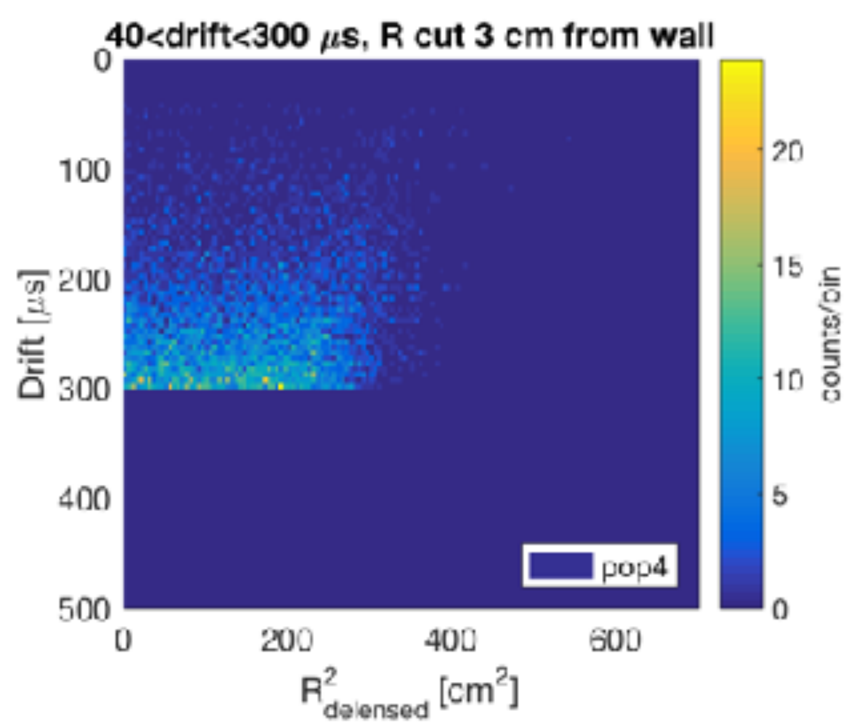
DRIFT POSITIONS OF POPULATIONS: FIDUCIAL CUT APPLIED











- Population 1: Applying a fiducial cut removed the hotspot and left a fairly uniform distribution.
- Population 1b has many wall events removed.
- Populations 2 and 3 are still concentrated near the top.
- Population 4 is still concentrated near the bottom.