Supplementary Material

An Open Reprocessing Pipeline and Detector-Response Characterization
for the ADRON-55 High-Altitude Ionization Calorimeter

Zh. T. Sadykov et al.

Supplementary Material

This document collects the two sensitivity studies that Reviewer 3 asked for in points 5 and 10 of the report, together with the tables they rest on. They are placed here rather than in the article because they measure the stability of results reported there rather than adding results of their own, and because the article is already long. Section numbers and equation numbers refer to the article.

S1. Sensitivity to the quality selection

The composite quality score of Section 3.7 of the article is a heuristic, and a referee asked how much of the angular result depends on it. The score does not enter the reconstruction; it only decides which events are kept, so one pass over all 164 227 events that survive the integrity filter, recording the five sub-scores alongside the reconstructed direction, makes every variation arithmetic on one table. That pass reproduces the published selection exactly: 86 433 events at GOOD or above, of which 990 give a single track with three active planes in each projection. It was run on a different machine from the one that produced the published figures, which sets the scale on which the numbers below should be read: every event of the clean sample is recovered, the per-event azimuth agrees to a median of 0.07° and at worst 2.3°, and the binned flatness statistic moves from the published 15.77 to 16.44, because 147 of the 990 events lie within half a degree of a 10° bin edge and a few of them cross it. The variations below are all measured against that 16.44.

Table S1 collects the variations: the GOOD threshold moved by ± 5 and ± 10 points, the selection opened to POOR and restricted to EXCELLENT, each of the five weights in turn set to zero with the others renormalized, and all five weights made equal. The number of events admitted moves between 63 915 and 86 512, a range of a quarter, and in eleven of the thirteen variations the clean sample is the same 990 events and every angular number is unchanged in every digit printed. Only the EXCELLENT-only selection differs, removing 56 events and moving χ^2/ndf from 16.44 to 16.88.

The reason is visible in the sub-scores themselves. Within the clean sample the penetration term is 75 for every event, the topology term is 100 for every event, the energy-profile term is the constant 70 for every event in the pipeline, and the purity term saturates at 100 for 95% of them; only the fit term varies, and its median is its maximum. Four of the five terms are constants inside this sample, so no reweighting can reorder it. The same dissection explains the grades themselves: no track in the data reaches a penetration depth of four in a projection, so the nominal 0–100 scale never exceeds 88.25, the EXCELLENT threshold of 85 is equivalent to requiring a two-dimensional fit $\chi^2/\mathrm{ndf} \leq 1$, and 71 483 of the 71 562 POOR events are POOR because they yield no track at all. The score is not a finely tuned discriminator; on this data it reduces to two conditions, whether the event has a reasonably penetrating track and whether that track fits.

S2. Sensitivity to the tracking parameters

The parameters of the reconstruction itself are a separate question, because they change what is reconstructed rather than which events are kept, so each value needs its own pass. Table S2 reports one, over every fourth event of the archive. The parameters of Table 2 of the article divide into three groups. Some are fixed by the hardware and are not free: 4000 ADC counts is the top of the digitizer's high-sensitivity range, and the five active-plane threshold is set by the pedestal measurement of Section 3.2, which finds 13 channels of 479 with a pedestal above five counts. Some are inert on this data and we now say so: the 500-count pre-selection removes two events of 243 746, and the three-dimensional fit acceptance $\chi^2_{\mathrm{3D}} < 10$ is never reached, since moving it to 5 or to 20 leaves every number in the table bit-identical. The projection-matching energy ratio is nearly inert as well, 247 and 254 clean tracks at 2 and 6 against 253 at the released value of 3, and forbidding plane skips costs 6% of the tracked events while leaving the clean sample unchanged.

Two parameters do matter, and in both cases the sensitivity is informative rather than reassuring. The width of the search road sets the size of the clean sample almost linearly: 186 tracks at two channels, 253 at three, 417 at five, with the azimuthal statistics moving in step. The clean 3+3 sample is therefore as much a property of the road width as of the instrument, and its size should not be read as a detector property. The cluster threshold matters for a different reason. Moving it from 30 to 20 or 50 counts shifts the mean reconstructed slope by 0.04 and 0.07, which is the scale of the residual misalignment identified in Section 4.4 of the article, and raises the flatness statistic from 4.52 to 7.01 and 6.14. The reason is a coupling that the parameter table does not show: the alignment offsets of Table 3 of the article were derived at a threshold of 30, and the threshold changes the cluster centroids, so the offsets no longer hold when it moves. Changing the clustering threshold requires re-deriving the alignment, and we state that in the released configuration.

The last row of the flatness table is not a variation of the score. It is the illumination cut of Section 4.4 of the article, admitting 85 642 events, within a percent of the published selection, and moving χ^2/ndf from 16.44 to 1.97. The angular results are insensitive to the selection the pipeline tunes and sensitive to a property of the events that the selection does not look at.

Table S1: Sensitivity of the angular results to the heuristic quality selection. Each row is a variation of the composite score of Section 3.7 of the article: the grade threshold moved, each weight in turn set to zero and the rest renormalized, or all five weights made equal. *Kept* is the number of events the variation admits, *clean* the number of them that give a single track with three active planes in each projection, and the remaining columns are that sample's mean reconstructed slopes, its azimuthal flatness and its harmonics. The last row is not a variation of the score at all but the illumination cut of Section 4.4 of the article (†), shown on the same footing. All rows come from one re-run of the pipeline, so they are comparable with each other; the re-run reproduces the published sample event for event and its flatness statistic to 0.7, as the text explains.

Variation	Kept	Clean	χ^2/ndf	$\langle s_x \rangle$	A_1/a_0	A_2/a_0
as published	86 433	990	16.44	$+0.0000 \pm 0.0116$	0.115	0.244
identical in every column below, for 11 further variations:						
threshold 55; threshold 60; threshold 70; threshold 75						
POOR included; weight penetration → 0; weight track_goodness → 0						
weight topology → 0; weight cluster_purity → 0						
weight energy_profile → 0; equal weights						
EXCELLENT only	26 546	934	16.88	$+0.0023 \pm 0.0117$	0.113	0.272
no plane lit edge to edge†	85 642	199	1.97	$+0.1520 \pm 0.0365$	0.532	0.162

Table S2: Sensitivity of the reconstruction to the tracking parameters of Table 2 of the article. Each row is one value of one parameter, everything else held at the released configuration, run over the same fixed set of events: every fourth event of the archive, 41 000 reconstructions spanning both seasons. *Good+* is the number of events graded GOOD or above, *tracked* the number with at least one three-dimensional track, *clean* the number giving a single track with three active planes in each projection. The last four columns are that sample's mean slopes and azimuthal statistics. The rows are comparable with each other but not with the full-sample numbers of Section 4.3, which use four times the statistics.

Parameter	Value	Good+	Tracked	Clean	χ^2/ndf	$\langle s_x \rangle$	A_2/a_0
as released		21 631	16 520	253	4.52	−0.001	0.287
search window	2 channels	21 093	16 105	186	5.14	+0.064	0.466
search window	5 channels	22 228	16 961	417	7.34	+0.029	0.344
3D-fit acceptance	$\chi^2_{3D} < 5$	21 631	16 520	253	4.52	−0.001	0.287
3D-fit acceptance	$\chi^2_{3D} < 20$	21 631	16 520	253	4.52	−0.001	0.287
projection matching	energy ratio 2	21 631	16 512	247	4.57	−0.010	0.282
projection matching	energy ratio 6	21 631	16 523	254	4.68	−0.001	0.294
allowed plane skips	0	21 104	15 479	258	4.37	+0.002	0.266
cluster threshold	20 ADC counts	21 594	16 431	241	7.01	+0.038	0.233
cluster threshold	50 ADC counts	21 704	16 718	245	6.14	+0.068	0.250

S3. The detector-level Monte Carlo

The sub-sample also admits the comparison with a simulation that the sample as a whole does not. We generated straight tracks isotropic in azimuth with a $\cos^n\theta$ parent, digitized the transverse coordinate through the inverse of Eq. (3) of the article at the measured plane positions, applied per-channel efficiencies, and passed the synthetic records through the same four reconstruction calls as the data. The efficiencies are measured on the data by tag and probe: two planes of a projection define a line, the third is interrogated in a window around the predicted position, and a window displaced by twenty channels measures the accidental rate. The corrected per-plane values are 0.847 ± 0.003, 0.921 ± 0.002 and 0.869 ± 0.003 for planes 3, 5 and 7 and 0.597 ± 0.006 and 0.706 ± 0.006 for planes 6 and 8, with the per-channel shape taken from the fraction of events in which each channel carries a large signal. Plane 4 is the exception and we state it plainly: its accidental rate is 0.35 and its residual distribution is flat, so the method does not measure it, and its efficiency is instead set to the single value, 0.147, that reproduces the observed ratio of 3+3 to 3+2 tracks. It is the one tuned parameter in the simulation.

The geometric acceptance of the stack is 124 $\mathrm{m}^2\,\mathrm{sr}$ for $\theta < 80°$. The parent index is fixed by one number, the median reconstructed zenith angle: $n = 3.5$ gives 27.2° against the 26.8° of the properly illuminated sub-sample, a value consistent with a shower-triggered population being more vertical than the $\cos^2\theta$ of atmospheric muons, whereas matching the full sample would require $n = 13.5$, which nothing physical produces and which is a measure of how much the centroid effect steepens the apparent zenith distribution. Everything about the azimuth is then a prediction, since no weight in θ can produce structure in φ. The simulation returns $\chi^2/\mathrm{ndf} = 0.49$ scaled to the size of the data sample, against the 1.18 measured out of sample and the 1.31 measured in sample, and harmonics of $A_1/a_0 = 0.086$ and $A_2/a_0 = 0.043$. The modulation that the geometry and the measured efficiencies predict is thus smaller than the statistical floor of a sample of this size, and the data, once the edge-to-edge events are removed, are compatible with it. The one number that is not compatible is the second harmonic, 0.236 ± 0.128 in the data against 0.043 in the simulation; at this sample size the difference is not significant, and settling whether a second harmonic of that size is real needs more data rather than a better model.

S4. Reproducibility of the released pipeline

The passes above were run on a different machine from the one that produced the published figures. Every event of the clean sample is recovered, the per-event azimuth agrees to a median of 0.07° and at worst 2.3°, the zenith angle to 0.08°, and the binned flatness statistic moves from the published 15.77 to 16.44, because 147 of the 990 events lie within half a degree of a 10° bin edge and a few of them cross it. All rows of Table S1 are measured against that 16.44, so they are comparable with each other.