Data-driven SIREN optical model for ProtoDUNE-HD: Updates

Jay Hyun Jo

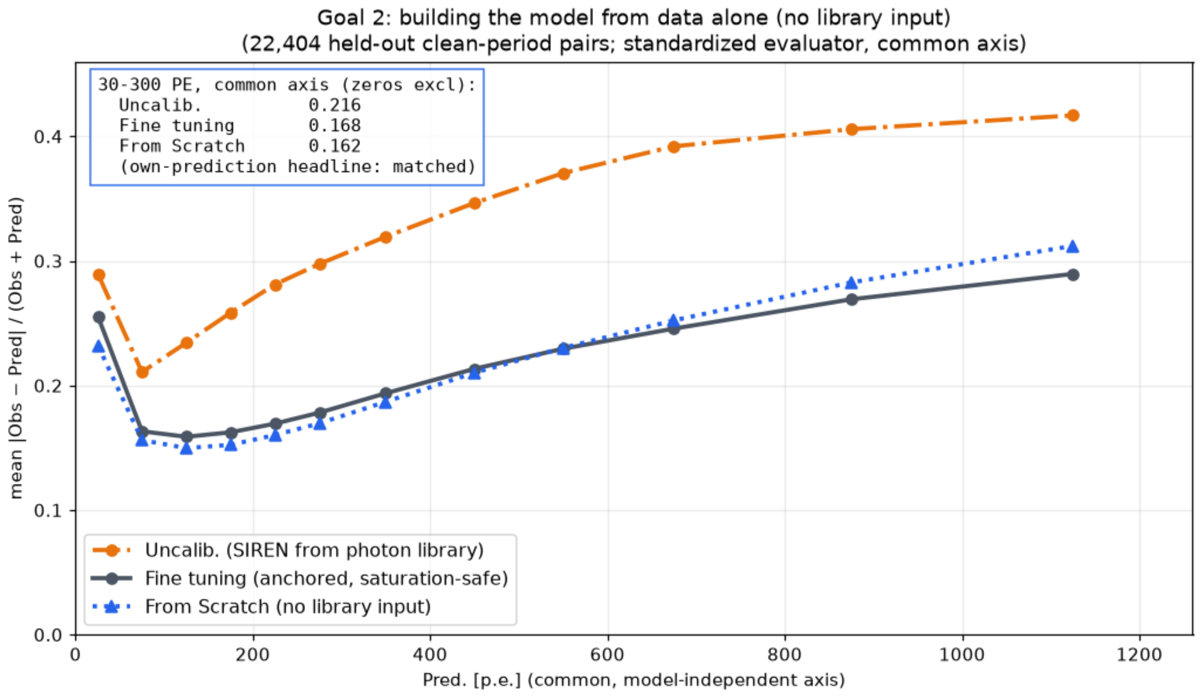
BNL ProtoDUNE Meeting

September 2, 2026

Motivation & Last week (v1)

- Goal 1: calibrate the existing photon library with cosmic data
- Goal 2: build the map from data alone

Since the last talk: a 6% track class was distorting the map, the sample size increased by x1.6



V1 result; large bias with high PE events

At first, I thought this was due to “saturation” of light signals from optical readout

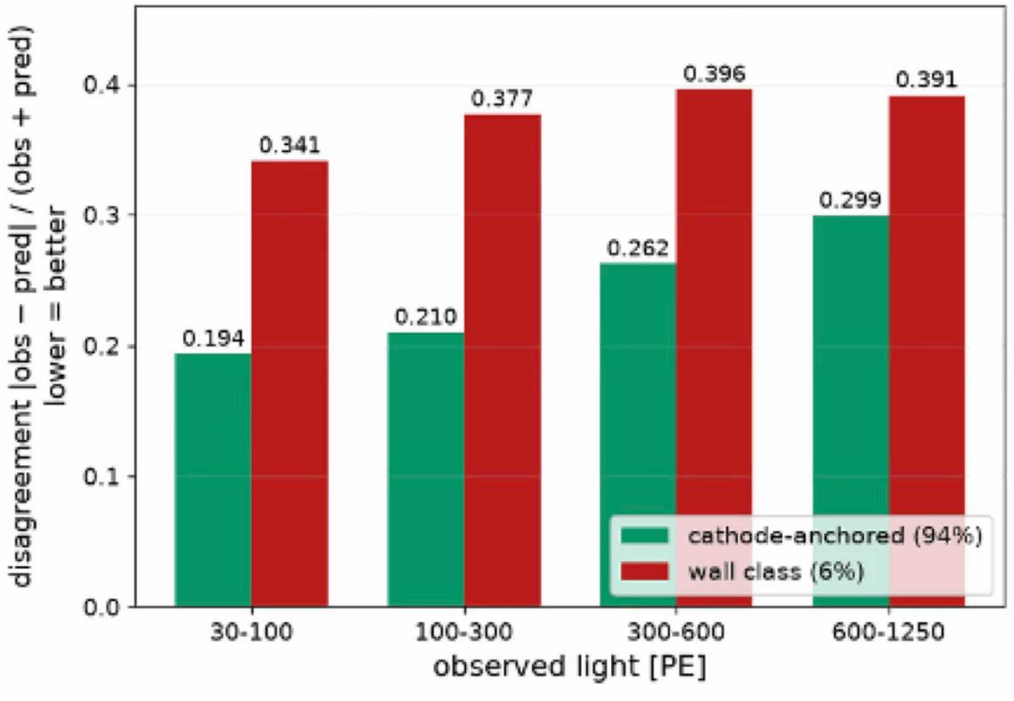
But it turned out I accidentally selected “wall-hugging” events (I wanted cathode-crossing tracks only)

These events very close to APA wall, …

Motivation & Last week (v1)

- Goal 1: calibrate the existing photon library with cosmic data
- Goal 2: build the map from data alone

Since the last talk: a 6% track class was distorting the map, the sample size increased by x1.6



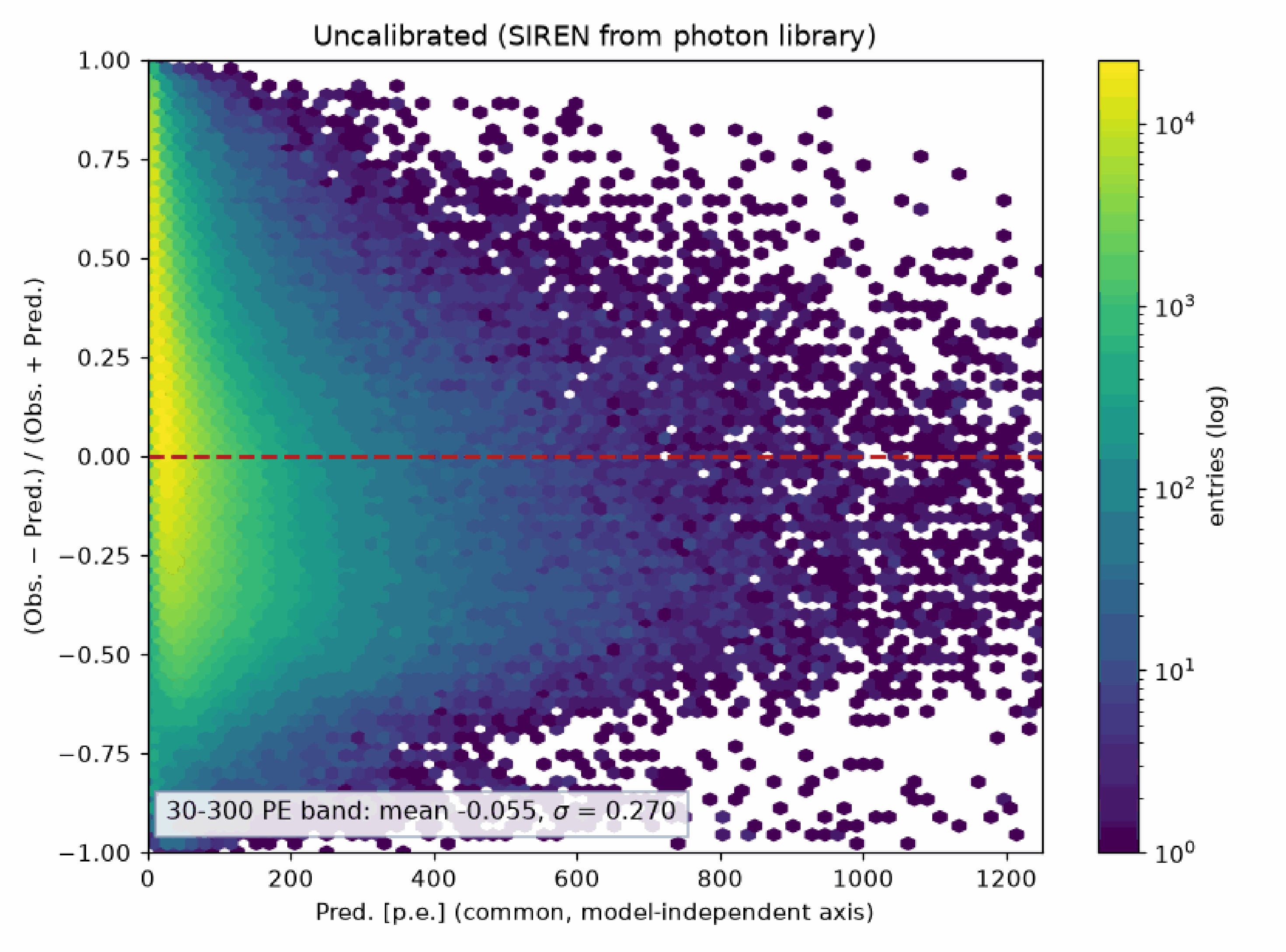
V1 result; large bias with high PE events

At first, I thought this was due to “saturation” of light signals from optical readout

But it turned out I accidentally selected “wall-hugging” events (I wanted cathode-crossing tracks only)

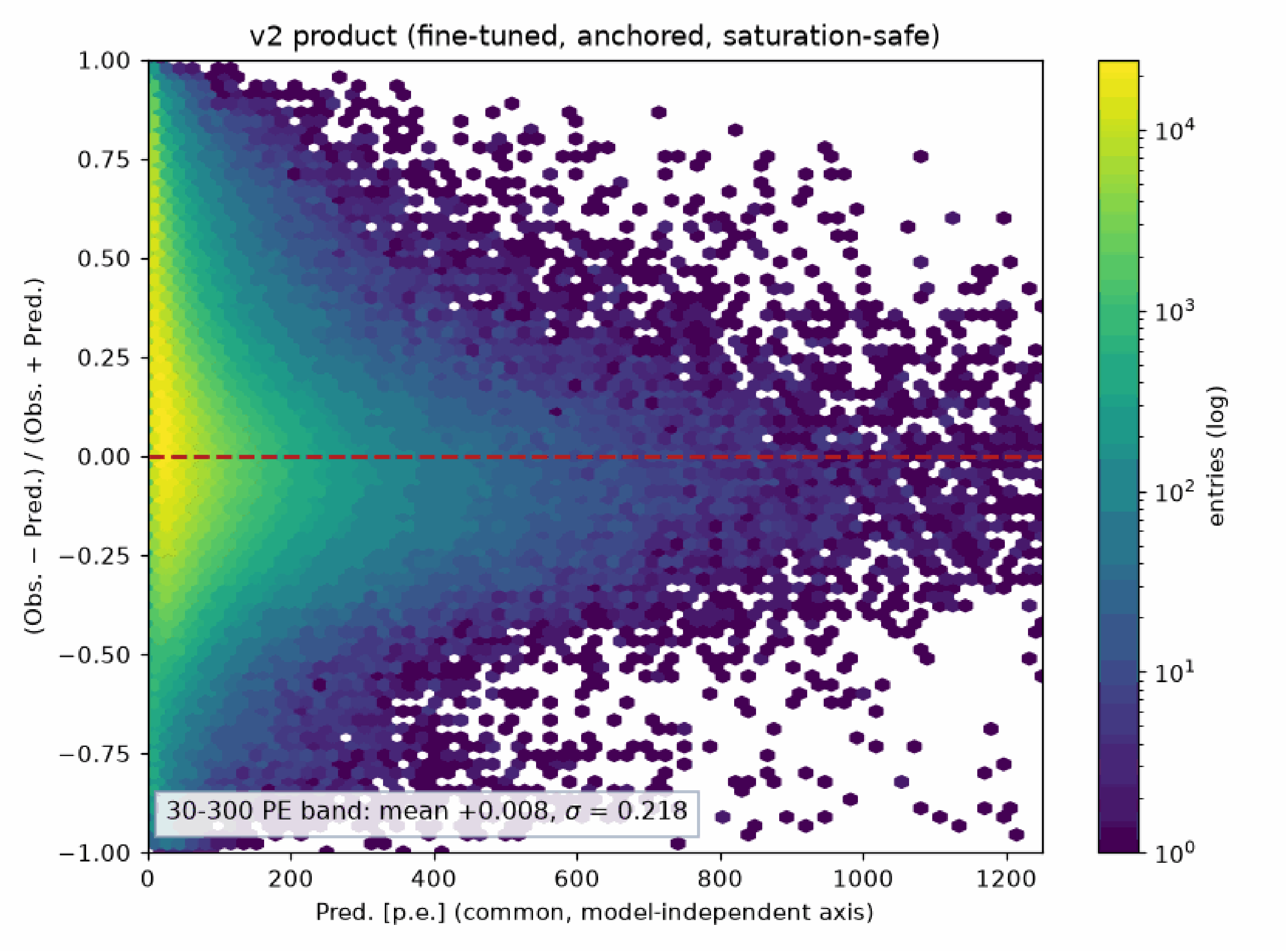
These events very close to APA wall, …

SIREN from simulation (before calibration)



- Uncalibrated: SIREN trained from the photon library (simulation only), evaluated on the held-out pairs
- Systematic under-prediction: band mean -0.055
- Wide residual: σ = 0.270

SIREN after calibration: the v2 product

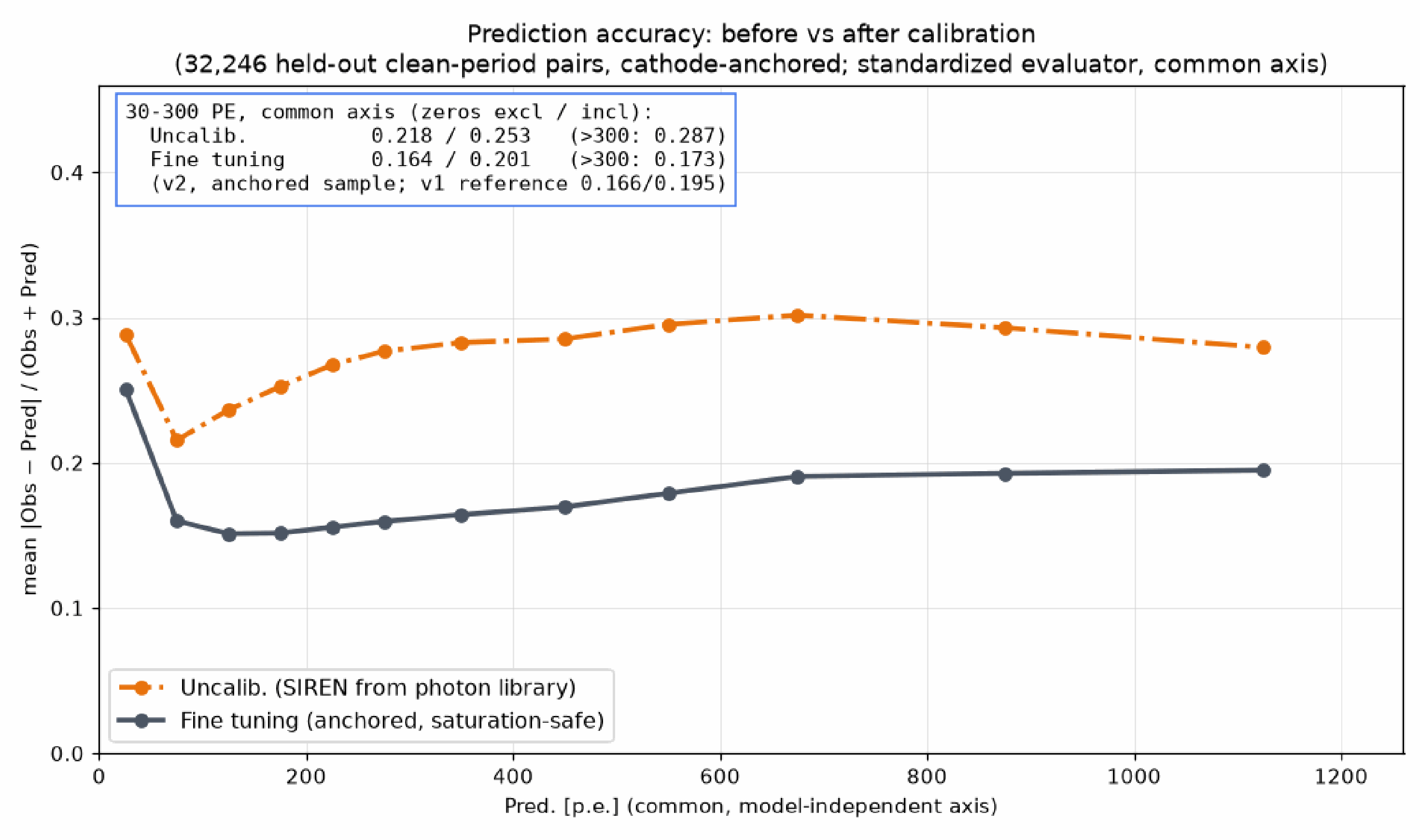


- Same pairs, the v2 product: the library-initialized SIREN re-optimized on the tracks

- Bias removed: band mean +0.008; narrower residual: σ = 0.218

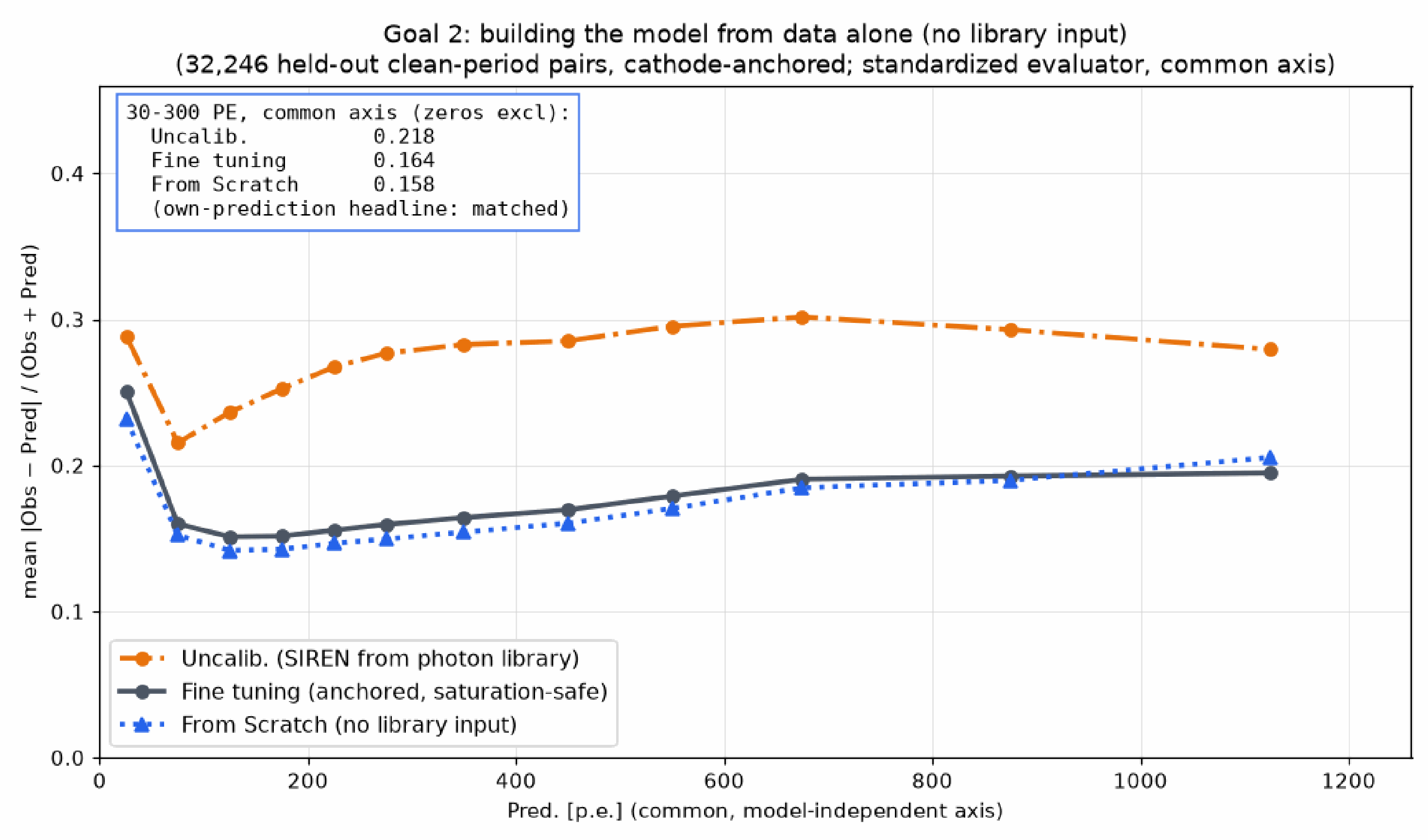
SIREN can be calibrated to remove the data–simulation discrepancy at ProtoDUNE scale.

Goal 1: calibrate the existing photon library with cosmic data



- Per-entry agreement between observed and predicted light
- 0.218 → 0.164 in the plateau and 0.287 → 0.173 in the bright band

Goal 2: build the map from data alone



Uncalibrated

- SIREN trained from the photon library (simulation); suffers the data–MC discrepancy

Fine tuning

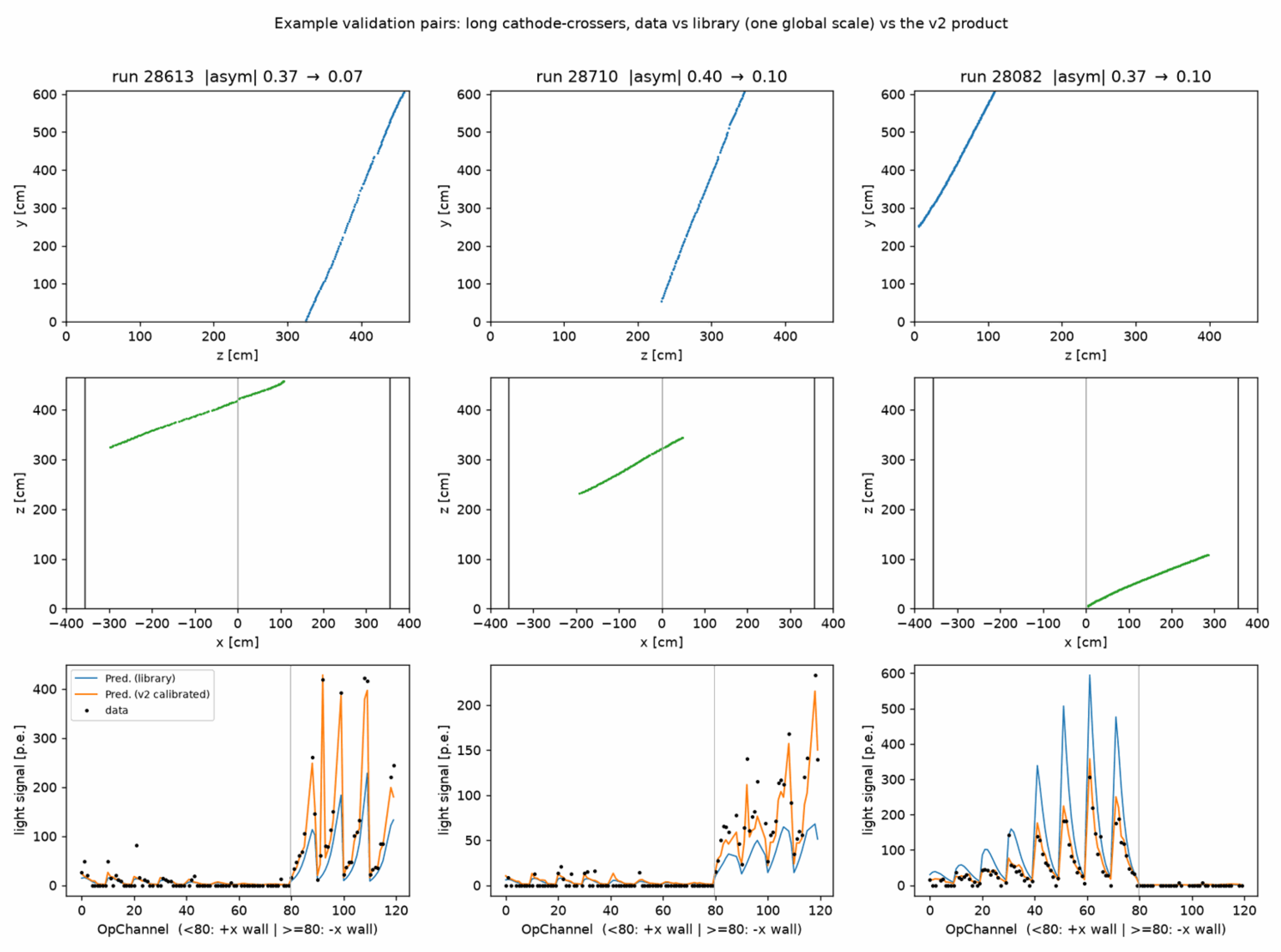
- use the uncalibrated SIREN as initial parameters; re-optimize with tracks; library prior over data-silent regions

From Scratch

- random initialization; optimize with tracks only, no library anywhere

SIREN model can be constructed from data alone, without prior knowledge from simulation, for light prediction.

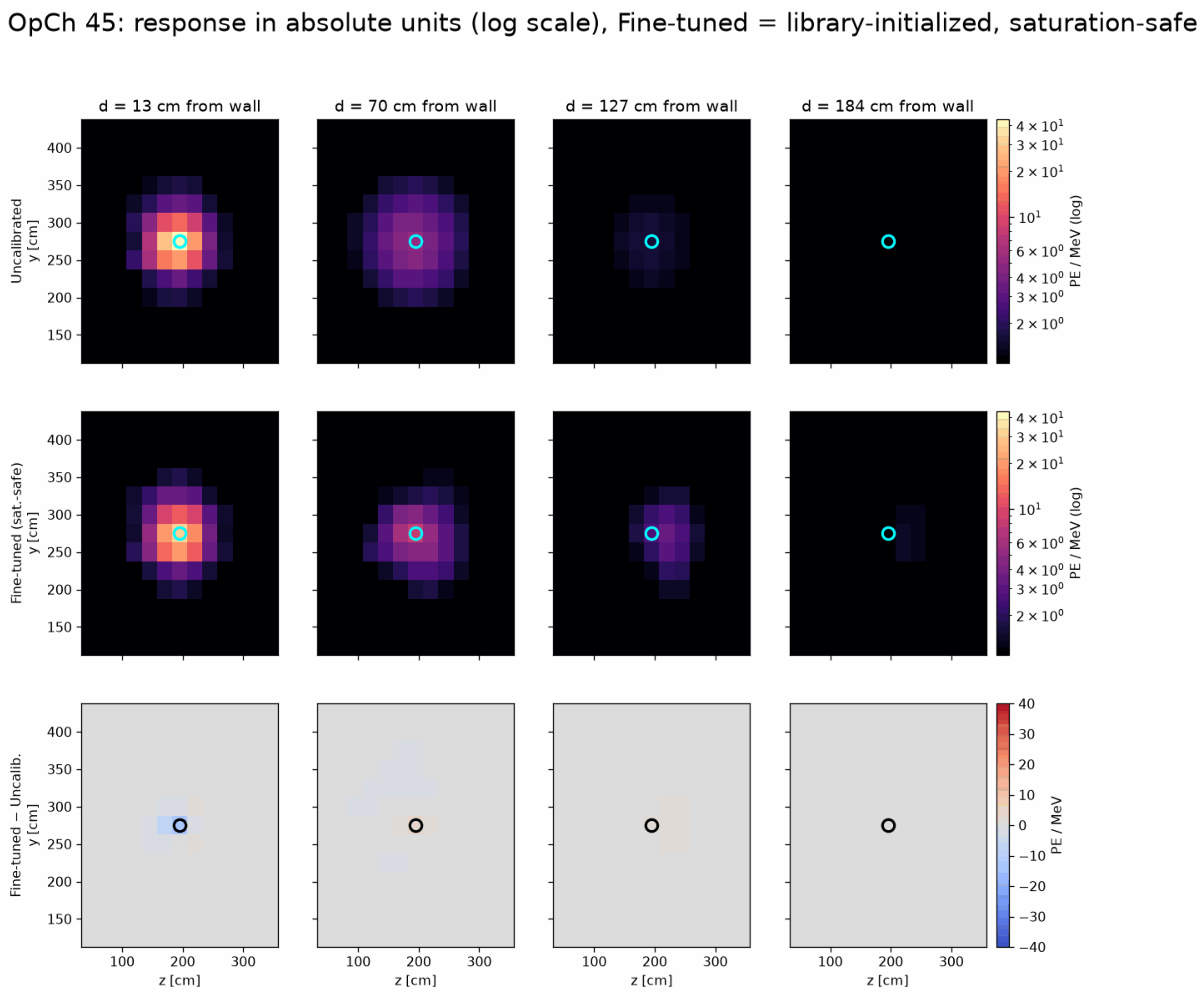
Example validation pairs



Long cathode-crossers with large calibration improvement AND good final agreement.

Better agreement after calibration

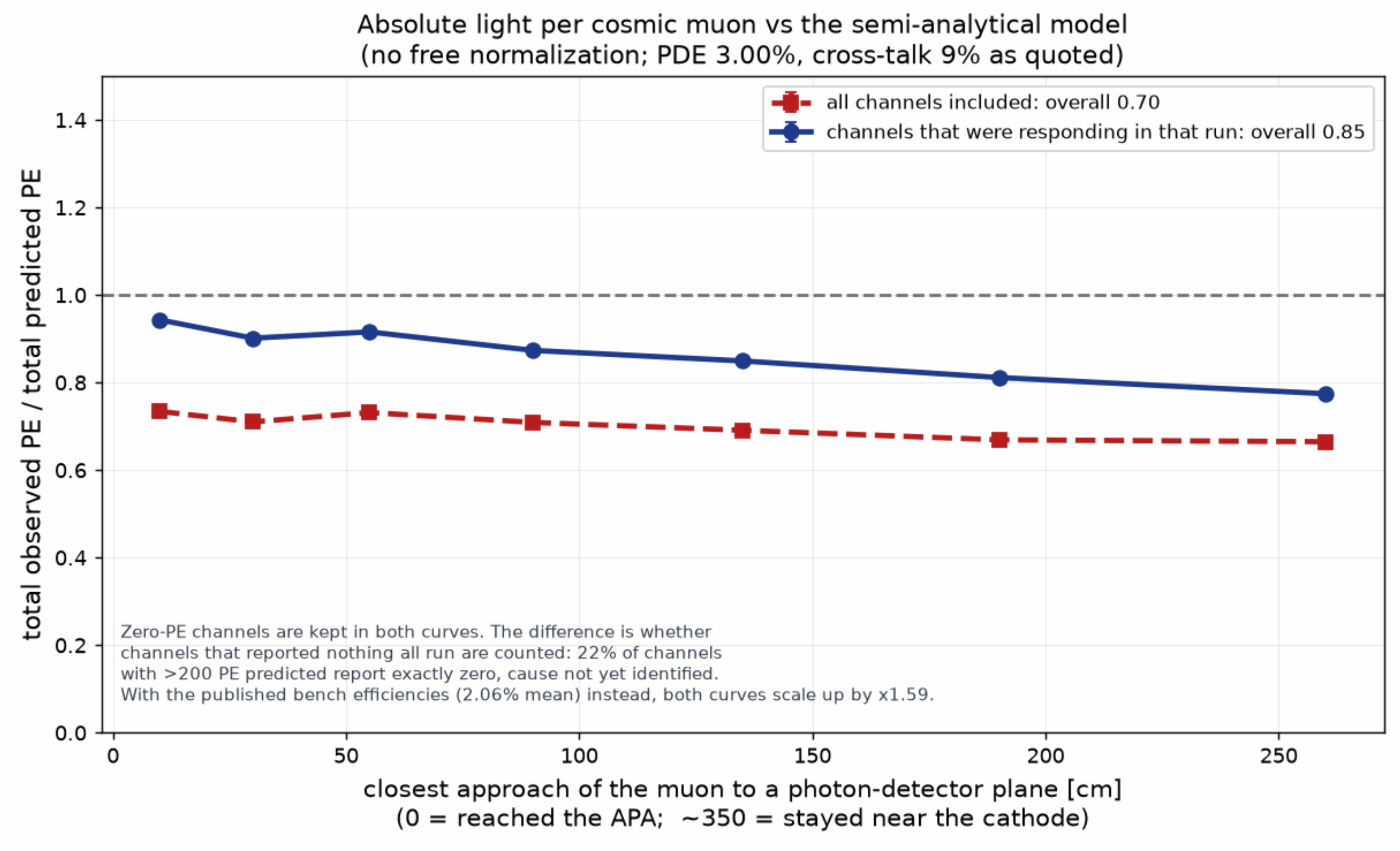
Visibility map: one channel



- One channel, four drift distances; circle = surveyed position

- Model matches the library pretty well

Is there an overestimation from photon library?



Total light seen per muon, divided by the total predicted from the semi-analytical model

- The two curves are the same data counted two ways
  - Upper: only channels that were reading out in that run
  - Lower: every channel, including ones that reported nothing

(About a fifth of channel measurements are exactly zero even when a large signal is predicted)

(An independent analysis (Gabriel B.) reports 0.5 to 0.65 with the same model and the same stated efficiency.)

Conclusions

- **Goal for ProtoDUNE**: calibrate the existing photon library with data, so that it both agrees better with the data and renders light at the accurate detector positions
- We now have the tools and framework to generate and calibrate a visibility map from data:
  Calibrating the existing library improves light prediction; the same recipe with no library input reproduces it.
- What next
  - Understand pred vs. meas PE difference better with experts
  - Expanding the map to cover APA1 optical detectors with full-streaming readout
  - Use Wire-Cell QL matched tracks for training event selection
  - We already have working version for PD-VD; need to finalize
  - Expand to Xe-doped PD-VD data

Backup: v1 → v2, four changes

change	what it bought
1. Removed the wall class 6% of pairs, drift unusable	Map shape returned to the library: drift falloff 0.63 → 0.29 (library 0.34), near-wall amplitude 0.28 → 0.77
2. Saturation mask 400 → 2000 PE the old cut assumed a readout compression we later disproved	The most effective single change: bright band 0.2019 → 0.1957, and the from-scratch map went 35 → 11 cm in peak offset
3. Sample 1.6x 20,000 more files → 280,433 pairs	Plateau ~1.5%. No further bright-end gain: the mask had already taken it
4. Near-wall anchor new soft prior term, bounded by detector geometry	Peak offset 8.1 → 6.4 cm mean, localised channels 61 → 74%, summed-map correlation 0.66 → 0.96

Sample purity 91.8%, measured from off-time windows, the fifth consecutive sample at that value with no retuning.

Backup: the wall class, in the event display

class	share	verdict
cathode crossers	~88%	drift fixed by the crossing, KEEP
cathode touchers	~6%	calo truncated but the trajectory does cross, KEEP
wall class	6%	never reaches the cathode, $\|x\|$ = 280–384 cm, points up to 27 cm outside the volume, EXCLUDE

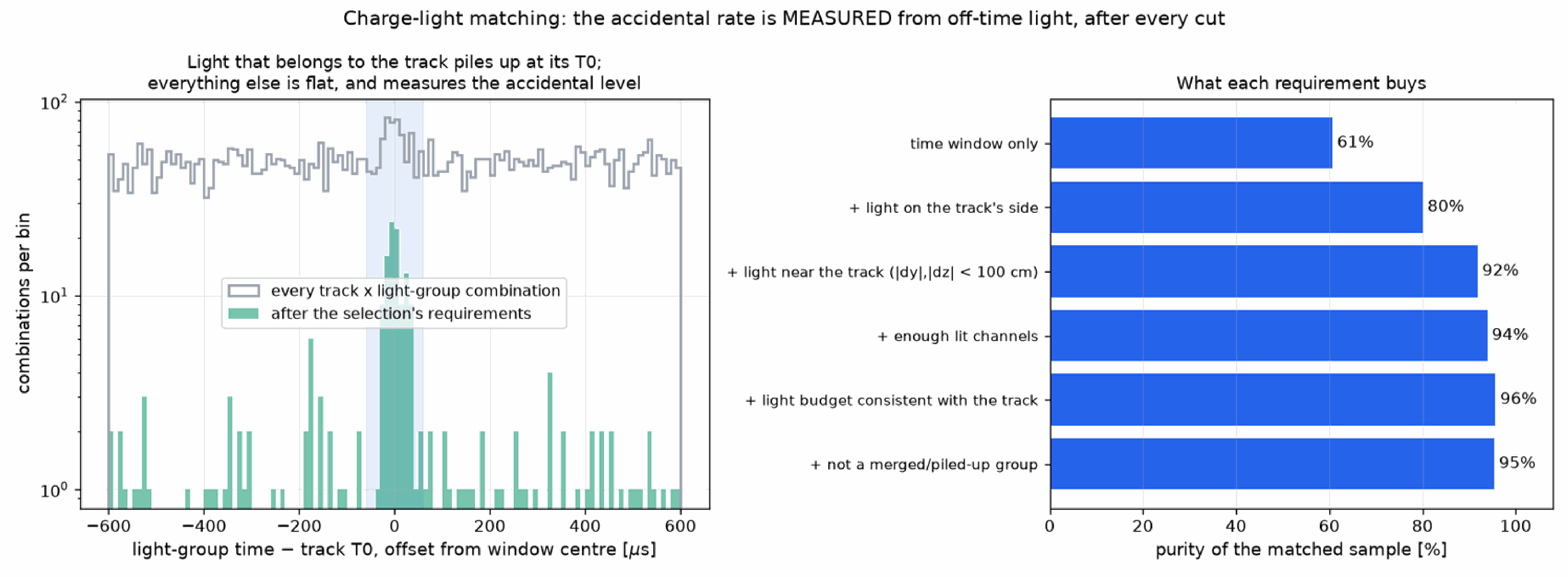
- Found by scanning selected events, not by a metric
- The class carries ordinary Pandora T0 values; provenance is an open question for the reconstruction experts
- Any T0-based selection inherits it, including external flash-matching work

Backup: track-flash matching

How the (simple) pairing works

- Light: merge the fragmented flashes into light groups (a median of 5 flashes per group, so the full scintillation tail is captured)
- Require isolation: no other significant light nearby in time. This alone removes 79% of groups
- Charge: cathode crossers only for t0 (5.6% of “long tracks” selected)
- Pair on time, in a coincidence window measured from the data itself, then loose consistency cuts: light barycentre within 100 cm of the track, and total light within a factor 16 of a bare $1/(4\pi d^2)$ geometric expectation
- Require a unique 1-to-1 pairing. If a track has two viable groups, or a group two viable tracks, both are discarded

Backup: how good is the charge-light matching?



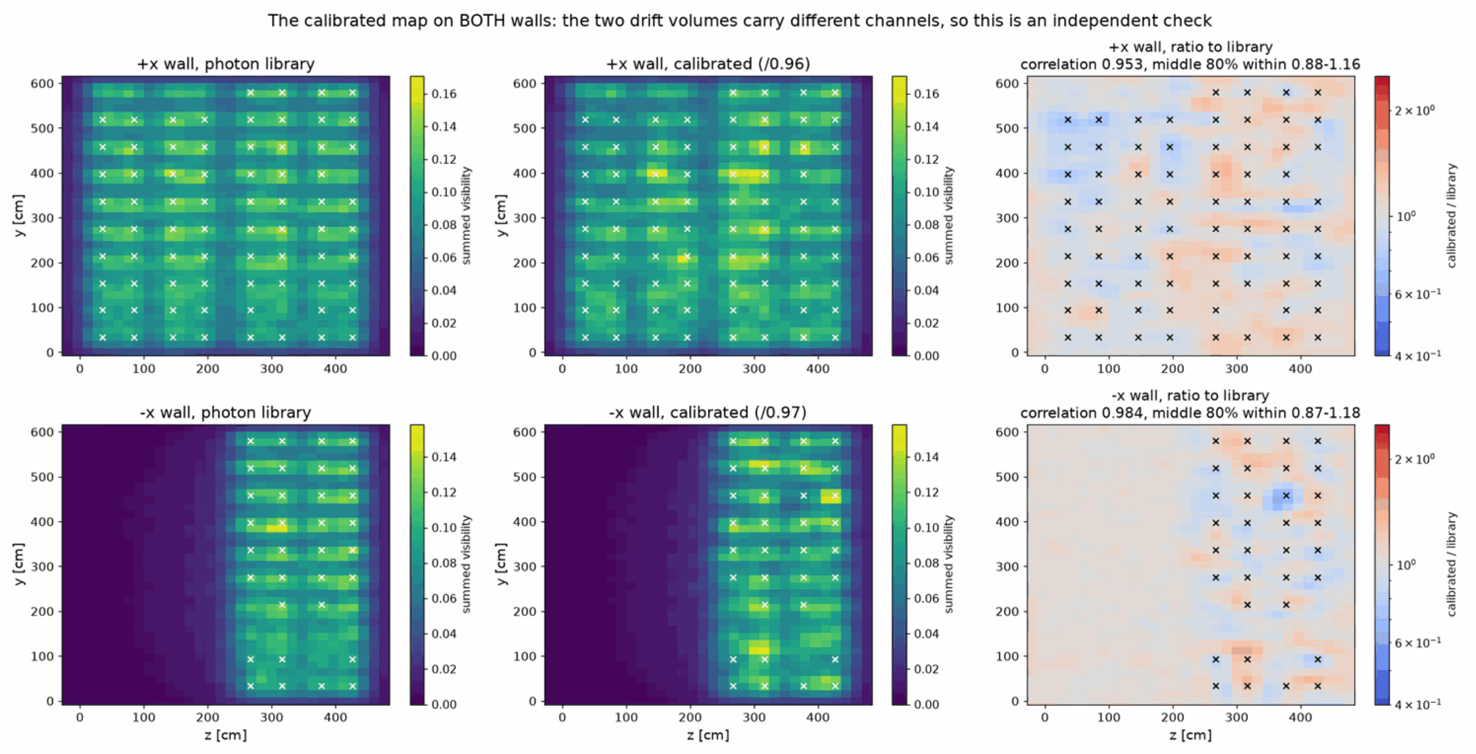
- Light belonging to the track piles up at its T0; unrelated light is flat in time, so the accidental rate is measured, not assumed
- Time window alone: 61% correct. With the selection's requirements: ~95%
- Cross-checked independently with off-time windows at ±500 μs over 20,000 files: 91.8%, the fifth sample at that value

Backup: where the matching-window numbers come from

number	provenance
+35 µs center	fitted peak position. It maps to ~5.5 cm in apparent x; suspected SCE at the cathode (Pandora stitches on uncorrected positions). A calibration of the T0 convention, not physics input
±60 µs window	2.4σ of the fitted σ = 25 µs. The width is Pandora stitching resolution (σ × vdrift ≈ 3 cm), not optical: each track's own light has RMS 4–5 µs. Wide costs purity only; t0 is re-anchored to the flash after pairing
15 µs merge gap	scanned (8/10/15): DAPHNE self-triggers fragment one burst over ~16 µs; at 10 µs the track's own light self-vetoes, −26% signal
±25 µs isolation	design choice pinned by two measured scales: above the 16 µs fragmentation span (no self-veto), matching the 25 µs association resolution (closer light is unattributable anyway)

δt = tflash − t0Pandora for all track–flash pairs, 620 events: a coincidence peak on a flat accidental floor.

Visibility map: summed view, both walls



- All channels summed at a wall slice: library, calibrated, and the ratio. Crosses mark the photon detectors