FTOF and Calorimeters
Data-MC comparison for EB kaons

Cooper Bell (SULI ANL)
Maria Zurek (ANL)

Argonne NATIONAL LABORATORY

Sample and objectives

Samples

- **Data:** CLAS12 RGA pass-2, Fall 2018
- **MC:** CLASDIS simulation
- Forward Detector tracks identified as K+ by the EventBuilder
- MC: require an available truth match, but retain misidentified particles
- Common vz cut: -8 < vz < 2 cm

Two samples compared (both give very similar conclusions)

- **Inclusive:** no additional DIS kinematic cuts
- **SIDIS:** $Q^2 > 2$ GeV², W > 2 GeV, 0 < y < 0.75, $M_X(eKX) > 1.6$ GeV

Objectives: The comparison between data-MC came as a detailed audit of variables that can be reliably used in the ML Kaon ID that is the cleanup of the EB PID=321 sample based on BDT.

Kinematic binning and variables compared

- **Energy:** `REC::Scintillator/Calorimeter.energy`
- **Time:** `REC::Scintillator/Calorimeter.time - REC::Particle.vt`
- **Path:** `REC::Scintillator/Calorimeter.path`

Momentum (GeV)	theta=5-15 deg	theta=15-25 deg	theta=25-35 deg
1–2	Bin 1	Bin 2	Bin 3
2–3	Bin 4	Bin 5	Bin 6
3–5	Bin 7	Bin 8	Bin 9

Quantifying Data-MC agreement

Three primary metrics

- **Normalized Wasserstein distance:** difference in distribution shapes, including shifts and broadening
- **Population Stability Index (PSI):** differences in the relative populations across the distribution
- **Maximum local residual:** largest relative bin-by-bin discrepancy

Additional diagnostics

- Kolmogorov–Smirnov (KS) distance and chi2/nof

Automatic classification

- **KEEP:** all three primary metrics indicate good agreement
- **CANDIDATE:** at least one metric indicates a discrepancy
- **DROP:** at least two metrics exceed their DROP thresholds
- Classifications are diagnostic flags, visual inspection and statistical limitations also considered

General conclusions Data-MC agreement

FTOF response

- **FTOF 1B:** Reasonable agreement in both energy and timing
- **FTOF 1A:** Energy reasonably reproduced; larger timing discrepancies, particularly at low momentum and small polar angles

Calorimeter response

- **Energy:** Data distributions generally shifted toward lower deposited energies relative to MC
- **Timing:** More significant discrepancies across calorimeter layers (shape)
- Largest overall discrepancies appear in **PCAL**
- **ECOUT**/**ECIN** showing comparatively better (but not good) agreement

Path lengths

- Generally good agreement across FTOF and calorimeters.
- Primarily geometrical observables, strongly correlated with track kinematics and potentially redundant

Backup

Data-MC agreement: criteria

Metric	KEEP	CANDIDATE	DROP
Population Stability Index (PSI)	< 0.10	0.10–0.25	≥ 0.25
Wasserstein distance / IQR	< 0.05	0.05–0.20	≥ 0.20
Maximum local residual	< 0.30	0.30–0.80	≥ 0.80

Combined decision (in each momentum-θ bin)

- **KEEP:** All available metrics satisfy KEEP thresholds
- **CANDIDATE:** At least one metric exceeds its KEEP threshold, but fewer than two reach DROP thresholds
- **DROP:** At least two of the three metrics reach DROP thresholds

Additional diagnostics

- KS distance > 0.05 flags shape disagreement
- Detector hit-fraction difference > 5 percentage points is flagged separately