# IceCube Muon Direction
# Reconstruction with GNNs

Applied Machine Learning · University of Copenhagen

**Thor Kjærsgaard Jensen, Leonora Misciattelli Mocenigo Soranzo, Oskar Friis Holløse, Nikolaj Strikart Nielsen & Emma Bruun**

**~2M**
MC events

**32.1°**
MLP baseline

**3.38°**
best model (GNN unc.)

**89.5%**
improvement

INTRODUCTION

# The IceCube Observatory

### South Pole

Buried 1,450–2,450 m deep in Antarctic glacial ice

### 5,160 DOMs

Digital Optical Modules on 86 vertical strings

### 1 $km^3$

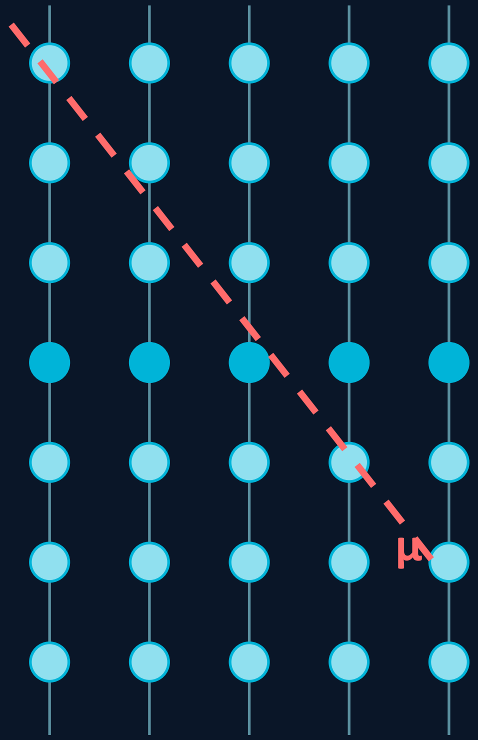
Cubic-kilometre detector volume

### Cherenkov light

Muons produce light cones as they pass through ice

### Goal

Detect neutrino sources, dark matter, cosmic rays

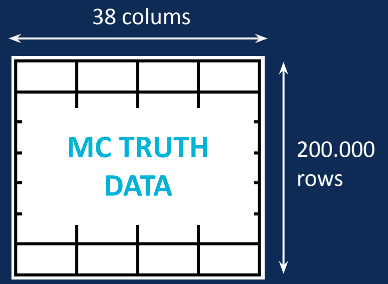


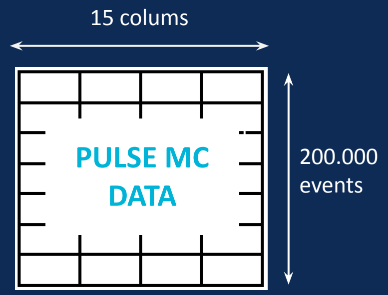

*Glacial ice*

# Monte Carlo Simulated Data & Real Data

## Monte Carlo Simulated Data

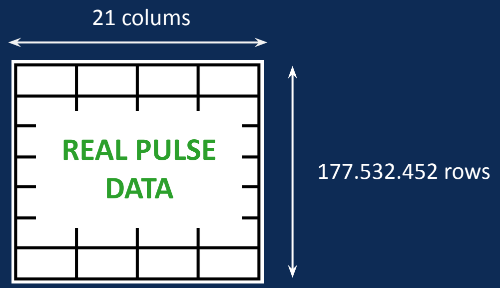
- 1,077,387 stopped **muon events**
- 947,513 through-going muon events
- Truth labels: zenith, azimuth, energy, vertex
- Two file types per class:
  Truth file ⟶ 1 row per event (38 cols)
  Pulse file ⟶ ~72 rows per event on average
  Pulse features: x, y, z, time, charge, HLC
- **Sampled to 200.000 events for this project**
  (stopped=106.290, through=93.710)
- Split 80% train / 10% val / 10% test




## Real IceCube Data (IC86.21)

- Detector recordings with no truth labels
- Same pulse format as Monte Carlo:
  ~72 rows per event ⟶ approx. 1.9M events
- Used for MC vs data comparison:
- Applied best trained GNN to real data
- Compare predicted zenith/azimuth distributions
- Agreement = model generalises to reality
- Key MC/data difference: pulse width discrete in MC, continuous in data⟶ width feature excluded from GNN input




| | | | |
|---|---|---|---|
| 200k randomly sampled events | ~72 pulses / event | 14.5M pulse rows | 80/10/10 train/val/test |

MOTIVATION / GOAL

# Project Objective

## Supervised Regression

Reconstruct direction of event based on additional parameters ⟶ we want to find the **incident angle** of the muon.

**Target values**

- **Truth data: dir_x/y/z**
- **Pulse data: Particle trajectory from dom_x/y/z at dom_time**

$$\vec{d} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$$

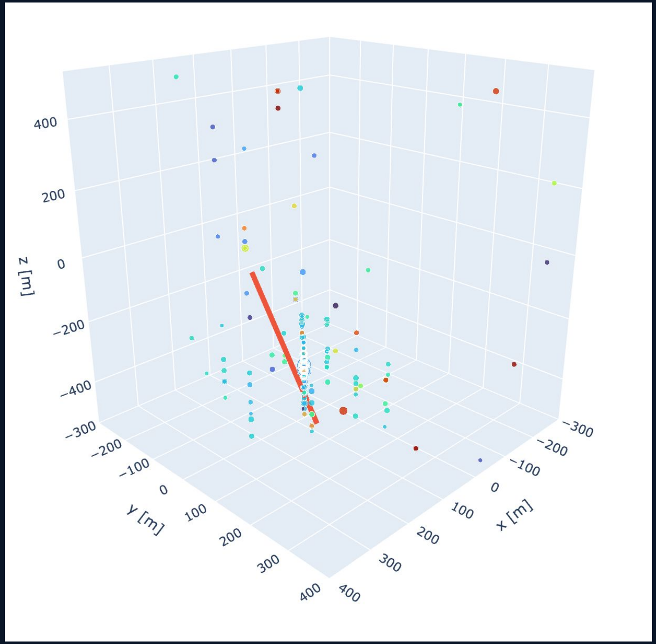
## Figures of Merit (two metrics)

**Median angular error**

$$\Delta\psi = \arccos(\vec{d}_{\text{pred}} \cdot \vec{d}_{\text{true}})$$

**Loss function for model : Cosine loss [0,2]**

$$\mathcal{L} = 1 - \vec{d}_{\text{pred}} \cdot \vec{d}_{\text{true}}$$




Visualization of a MC muon event: Points show detections in DOMs in event no. 1944718 from mc_SplitInIcePulses_merged_stopped.parquet. Color represents time of detection (blue = early, red = late) and size of the points gives the magnitude of charge. The red line is the true muon track given by mc_truth_stopped.parquet. We extended the trajectory as a line for visual purposes to show direction of muon.
From the pulse data, the goal is to reconstruct the red line given by the truth data.

# Models

## MLP

The Idea Behind It:

- Used as a simple, fast benchmark to prove that advanced geometric deep learning methods are necessary

What it Sees:

- 8 summery features

What it's good for:

- Extremely low computational cost

Struggles:
Collapsing data into averages destroys the raw spatial and temporal structure. It cannot exploit brighter, higher-energy events and fails on near-horizontal tracks.

## GNNs

The Idea Behind:

- Matches the physical reality of the detector and preserves the exact geometry of the particle track. .
  - Nodes = Pulses in sensors (DOMs)
  - Node features = recorded charge, time and position of hit, etc.
  - Edges = Adjacent sensors (neighbours)

What it Sees:

- Processes the raw, variable-length pulse matrix directly. Nodes represent individual sensor hits with feature vectors containing (x, y, z, t, q).

What it's good for:

- Naturally handles events with any number of sensor activations and scales dramatically with more data.
- Well-suited for variable event sizes and the irregular geometry of detector location and spacing

MODEL 1 — BASELINE

# MLP Baseline

*8 summary features per event*

| | |
|---|---|
| `cog_x/y/z` | Charge-weighted centre of gravity |
| `t_mean` | Charge-weighted mean hit time |
| `t_std` | Temporal spread of hits |
| `log_q` | log(1 + total charge) |
| `n_hits` | Total pulse count |
| `n_hlc` | Reliable HLC pulse count |


## Result

**32.1°** median error

**41.9°** 68th pct

**62.9°** 90th pct

**6.8%** events < 10°

MODEL 2

# GNN Default — EdgeConv

*Every pulse is a graph node · k=8 nearest neighbours · hidden dim 128*

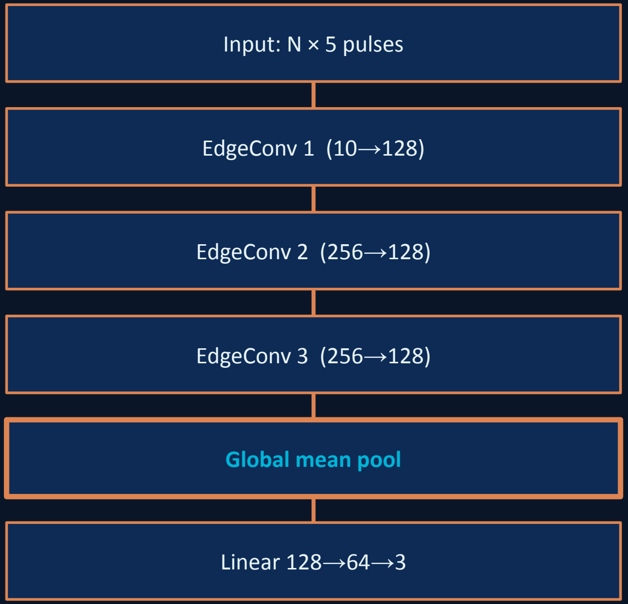
## EdgeConv key idea

For each pulse node i, aggregate from k=8 nearest neighbours:

$$h' = \mathrm{MLP}( [h_i \,||\, h_{\square} - h_i] )$$

## Why $h_{\square} - h_i$ matters

- Encodes relative timing Δt between neighbouring DOM hits
- Encodes relative position Δr between DOMs
- Cherenkov light speed in ice ⟶ Δt + Δr constrain direction
- MLP could never see this ⟶ it only sees per-event averages




## Result

| | | | |
|---|---|---|---|
| **5.46°** | **7.38°** | **13.3°** | **82.2%** |
| median error | 68th pct | 90th pct | events < 10° |

**83% improvement over MLP · best single model so far**

MODEL 3

# GNN Optimised

*Every architectural choice motivated by a physical or training reason*

| Change | Default | Optimised | Why |
|---|---|---|---|
| Pulse selection | First 128 in file | Top 128 by charge | Brightest = most info |
| Neighbours k | 8 | 12 | More spatial context |
| Hidden dimension | 128 | 256 | Double capacity |
| Conv layers | 3 | 4 | Deeper propagation |
| Batch normalisation | No | Yes | Stable training |
| Residual connections | No | Yes | Prevent grad vanishing |
| Global pooling | Mean pool | Add pool | Preserves event size |
| LR schedule | Cosine 20ep | Warmup + cosine 30ep | Careful convergence |

**Result:**

| 3.82° | 5.25° | 9.50° | 91.1% | 65.5% |
|---|---|---|---|---|
| median | 68th pct | 90th pct | < 10° | < 5° |

MODEL 4

# GNN Uncertainty

*The uncertainty extends the Optimised GNN with a second output head*

### Outputs:

Instead of outputting simple 3D point estimates the model produces both the direction vector and a concentration parameter (kappa) quantifying the certainty per event:

- predicted direction: $d \in R^3$
- vMF concentration: $\kappa > 0$

High $\kappa \longrightarrow$ prediction is concentrated around 1 direction
Low $\kappa \longrightarrow$ model assigns larger directional uncertainty

### Information:

Same optimised EdgeConvGNN with:

- 4 EdgeConv layers
- Residual connections
- Batch normalisation
- Global add pooling

### Loss function:

von Mises–Fisher negative log-likelihood

$$\mathcal{L} = -\kappa(\hat{\mathbf{d}} \cdot \mathbf{d}_{true}) + \log(\sinh \kappa) - \log \kappa$$

The loss rewards accurate directions with high confidence, but penalises high confidence when the predicted direction is wrong.

### 2 heads:

- Direction head: $R_{256} \longrightarrow R^3$
- Kappa head: $R_{256} \longrightarrow R_{>0}$

| Result: | 3.38° | 4.75° | 8.92° | 91.9% | 70.5% |
|---|---|---|---|---|---|
| | median | 68th pct | 90th pct | < 10° | < 5° |

MODEL 5

# GNN Multitask

*Predicting primary muon energy and direction simultaneously through a shared geometric backbone*

- Neutrino events in IceCube carry multiple correlated signatures. Rather than training isolated models, a shared GNN optimises direction and energy jointly.
- **Joint Parameter Extraction:** Sparse, irregular event patterns in the ice inherently encode both directional trajectory paths and energy-scale event brightness (PE count).
- **Geometric Regularization:** Forcing EdgeConv message passing layers to map physical energy parameters acts as a robust structural constraint, preventing direction-head overfitting.
- **Zero-Cost Enrichment:** Simultaneously predicts continuous $\log_{10}$(energy) in GeV for free, with virtually no degradation to baseline directional tracking accuracy.

**Loss function:**

$$L_{total} = L_{direction} + 0.1 \times L_{energy}$$

**Trajectory MSE**: measure angular alignment between predicted and true 3D unit vector. Traces coordinate pathway

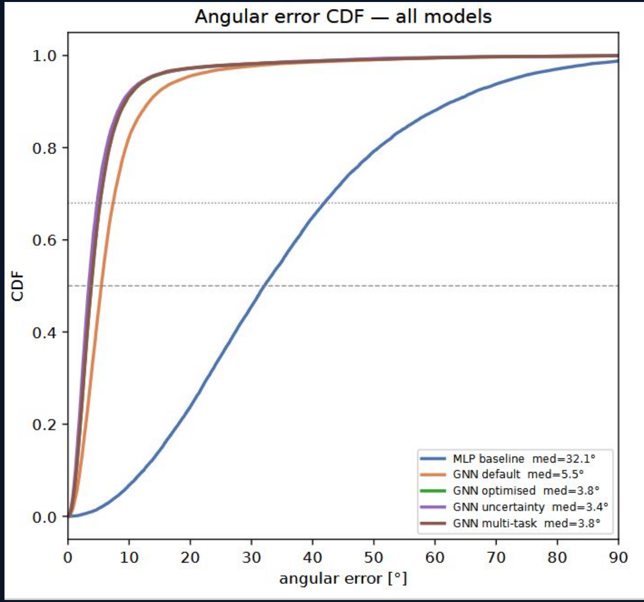
**Energy MSE:** measure discrepancy in $\log_{10}(E)$. Prevents high energy events from breaking backpropagation

**Scaling Weight:** Stabilizes gradient magnitudes, guaranteeing that directional geometry remains primary priority.

| Result: | 3.79° | 5.24° | 9.39° | 91.3% | 65.5% |
|---|---|---|---|---|---|
| | median | 68th pct | 90th pct | < 10° | < 5° |

RESULTS

# Results — Five-Model Comparison




| Metric | MLP | GNN def. | GNN opt. | GNN unc. | GNN MT |
|---|---|---|---|---|---|
| Median [°] | 32.1 | 5.46 | 3.82 | 3.38 | 3.79 |
| 68th pct [°] | 41.9 | 7.38 | 5.25 | 4.75 | 5.24 |
| 90th pct [°] | 62.9 | 13.3 | 9.50 | 8.92 | 9.39 |
| < 5° | 1.7% | 44.7% | 65.5% | 70.5% | 65.5% |
| < 10° | 6.8% | 82.2% | 91.1% | 91.9% | 91.3% |
| < 20° | 23.8% | 95.5% | 97.2% | 97.2% | 97.3% |

- MLP: flat 30–50° across all energies ⟶ cannot use event brightness
- GNN default: 5.46° ⟶ 83% better, preserves spatial/temporal pulse structure
- Optimised GNN: 3.82° ⟶ charge selection + deeper layers key
- GNN Uncertainty: 3.38° best ⟶ vMF kappa head improves direction too
- GNN Multi-Task: 3.79° best ⟶ additional energy prediction does not cost anything, slight improvement

EXTRAS

# Additional Analyses

## Uncertainty Estimation

- Model predicts kappa ⟶ confidence per event (vMF distribution)
- Kappa:
  High kappa ⟶ model is sure
  Low kappa ⟶ uncertain
- Best model: 3.38° median (GNN unc.)
- Key plot: kappa vs angular error ⟶ model "knows" when it's wrong
- High-kappa events cluster near 0° error

## Ablation Study

- Retrain GNN 6× ⟶ one input removed at a time
- No position: 49.2° ⟶ position is critical
- No time: 8.43° ⟶ timing carries direction signal
- No charge: 6.04° ⟶ charge helps but less critical
- k=16 vs k=8: minimal difference (5.51 vs 5.58) ⟶ timing + position are the essential features

## MC vs Real Data

- Apply trained GNN to real detector data
- Compare predicted zenith/azimuth distributions
- Checks all events
- Tests if simulated data is compatible with real data
- MC and data should agree if model generalises

## Multi-Task GNN (direction + energy)

- Second output head: predicts log_10(energy) simultaneously
- Combined loss: $L = L_{direction} + 0.1 \times L_{energy}$
- Result: 3.79° ⟶ matches optimised GNN
- Shared representations benefit both tasks

## Ablation — real results

| Experiment | Val median | vs default |
|---|---|---|
| k=4 | 6.35° | +0.77° |
| k=8 (default) | 5.58° | — |
| k=16 | 5.51° | −0.07° |
| No charge | 6.04° | +0.46° |
| No time | 8.43° | +2.85° |
| No position | 49.2° | +43.6° |

CONCLUSION

# Conclusion & Outlook

## Key Findings

- MLP: 32.1° ⟶ spatial/temporal structure cannot be captured by averages
- GNN default: 5.46° ⟶ 83% improvement by preserving full pulse geometry
- GNN optimised: 3.82° ⟶ charge selection + BN + residuals + deeper layers
- GNN uncertainty: 3.38° (best) ⟶ vMF kappa head improves direction too
- GNN multi-task: 3.79° ⟶ energy prediction shares useful representations

## Future Work

- Scale to full 2M events on GPU cluster for MC training
- Test an Energy + direction + uncertainty GNN
- More data!

**32.1°**
MLP baseline

**5.46°**
GNN default

**3.38°**

GNN uncertainty

**89.5%**
improvement over MLP

*Spatial and temporal pulse patterns contain the directional signal, the GNN recovers it, the MLP cannot.*

THE END

# THANK YOU FOR YOUR ATTENTION!

APPENDIX

# Group statement

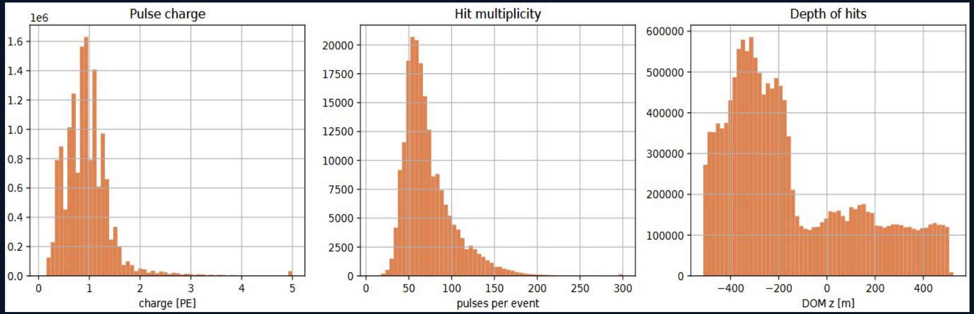
"Each member of the group—Thor, Leonora, Oskar, Nikolaj, and Emma—made an equal contribution to this project."

APPENDIX

# Raw Data visualisation

## Visualisation of Monte Carlo simulated data

- The simulated data have the following parameters:
  Energy, Zenith, Azimuth, Pulse charge, Hit multiplicity, Depth of hits.
- The distribution of these can help us visualise the data.
- Plots / Variable overview are from IceCubeComplete.ipynb

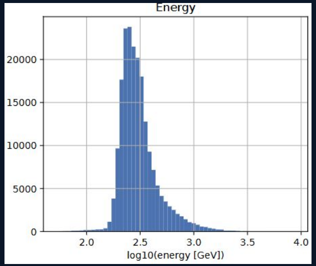



*Charge of pulses*

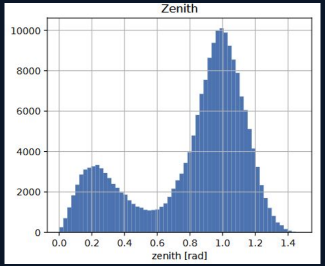
*Amount of pulses hit for each muon*

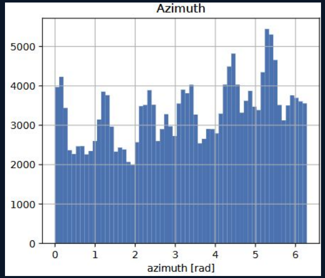
*How far into the ice the last detection happens*

*The energy distribution using log scale energy*




*Zenith in radians and counts*




*Azimuth in radians and counts*

APPENDIX

# MC vs Real Data:

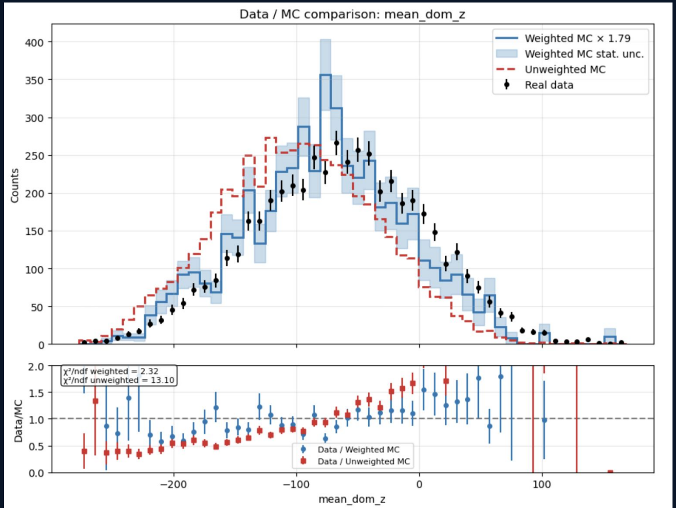
*Comparison of variables*

**In the beginning, when we got an overview over the data, we compared different variables of MC and real data. Among other things, we measured the importance of weight variable "norm_class_this_db_osc_weight" given in the MC data.**

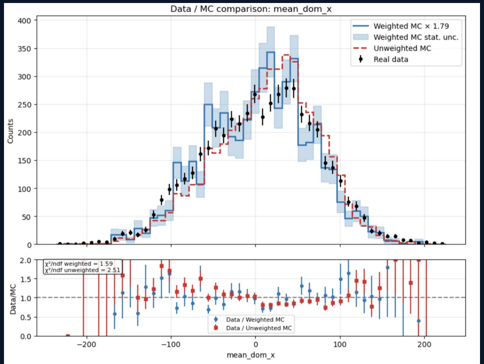
We normalized the weighted data, so the total weight was equal to the total counts of real data and unweighted data points.

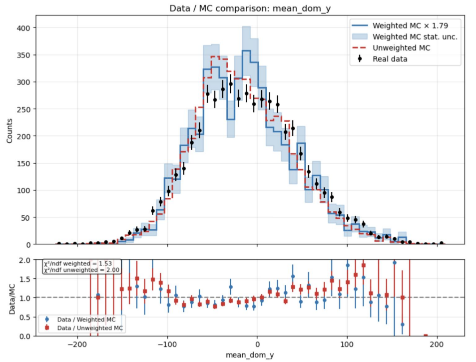
In the first plot we show the counts as a function of mean_dom_z which is given by the average of dom_z values in each event. The second plot shows the residuals given by the ratio between real data and MC data for each interval of the histogram. The errors are calculated using Poissonian statistics.

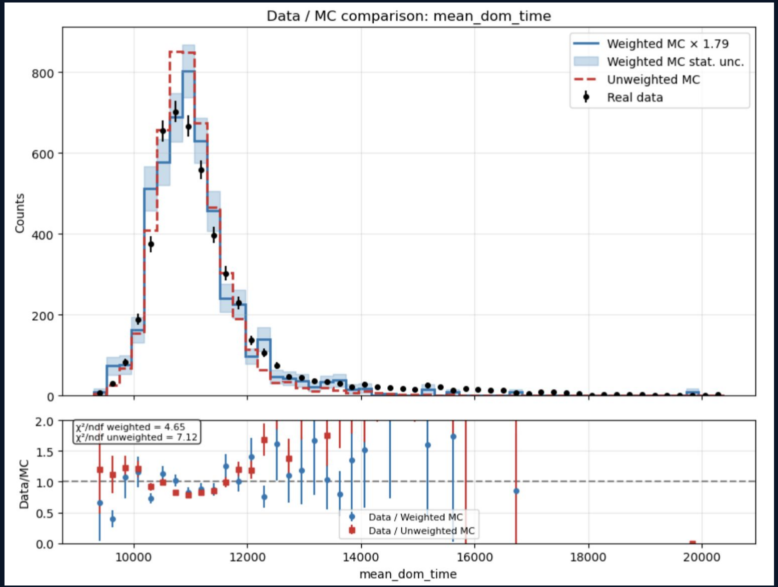
Comparing the unweighted and weighted distribution, the weighted data aligns better with the real data.

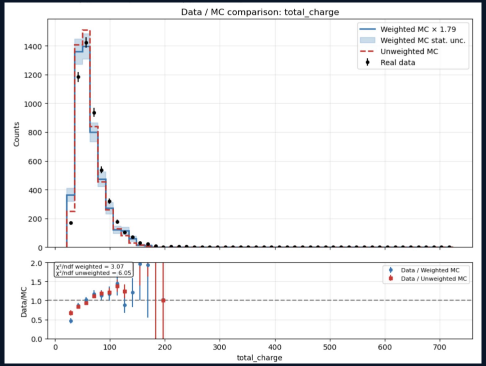


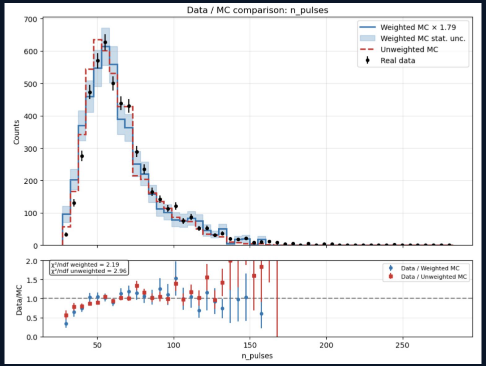
# More plots: Comparing MC and real data



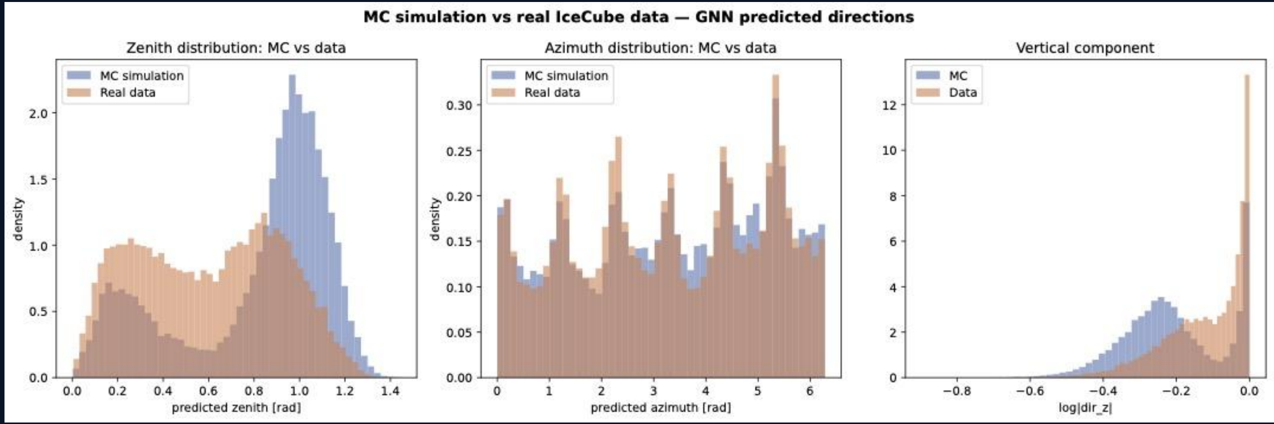



APPENDIX

# Comparing MC with predictions on real data

## Comparing azimuth and zenith

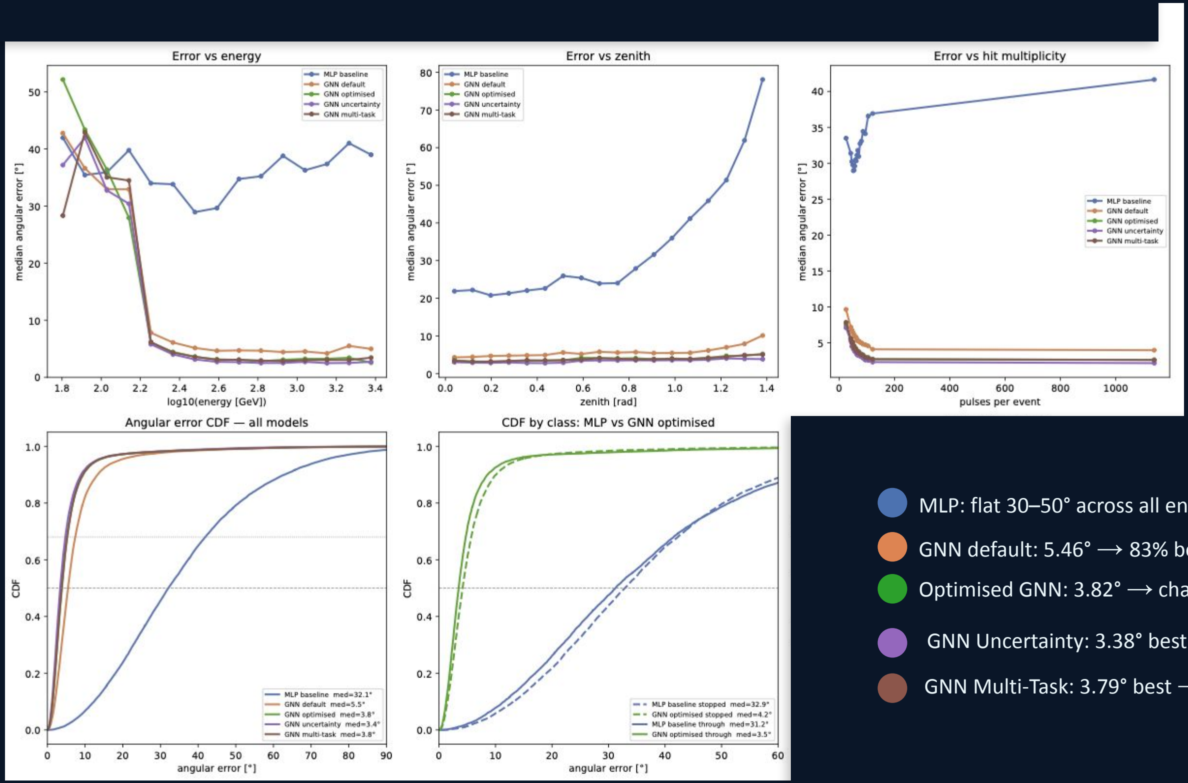
- Zenith shows discrepancies in the distribution of the predictions. The GNN predicts a more uniform distribution
- It does not predict the same large peak at approx. 1 rad
- The Azimuth Shows much better agreement and matches the distribution
- The individual peaks at each radian is also predicted by the GNN




MC vs data from IceCubeComplete.ipynb

APPENDIX

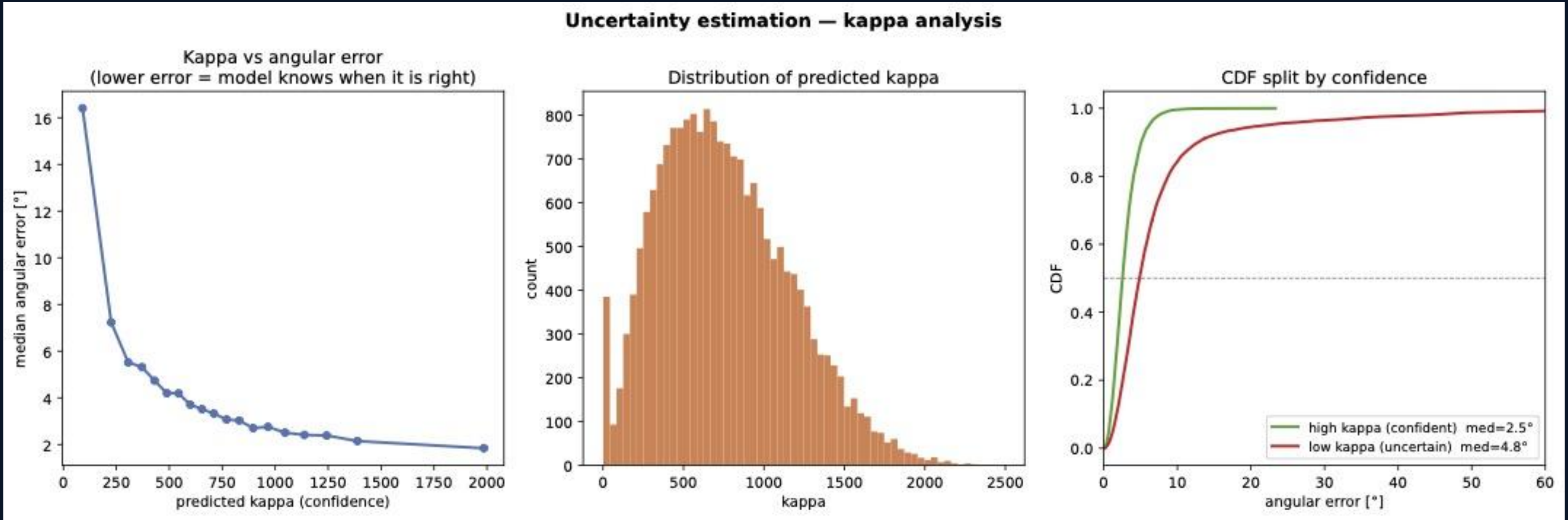
# Full Model Comparison: All 5 models




- MLP: flat 30–50° across all energies ⟶ cannot use event brightness
- GNN default: 5.46° ⟶ 83% better, preserves spatial/temporal pulse structure
- Optimised GNN: 3.82° ⟶ charge selection + deeper layers key
- GNN Uncertainty: 3.38° best ⟶ vMF kappa head improves direction too
- GNN Multi-Task: 3.79° best ⟶

APPENDIX

# Uncertainty estimation: Kappa analysis



# Some event images: