Direction Reconstruction using Simulation-Based Inference

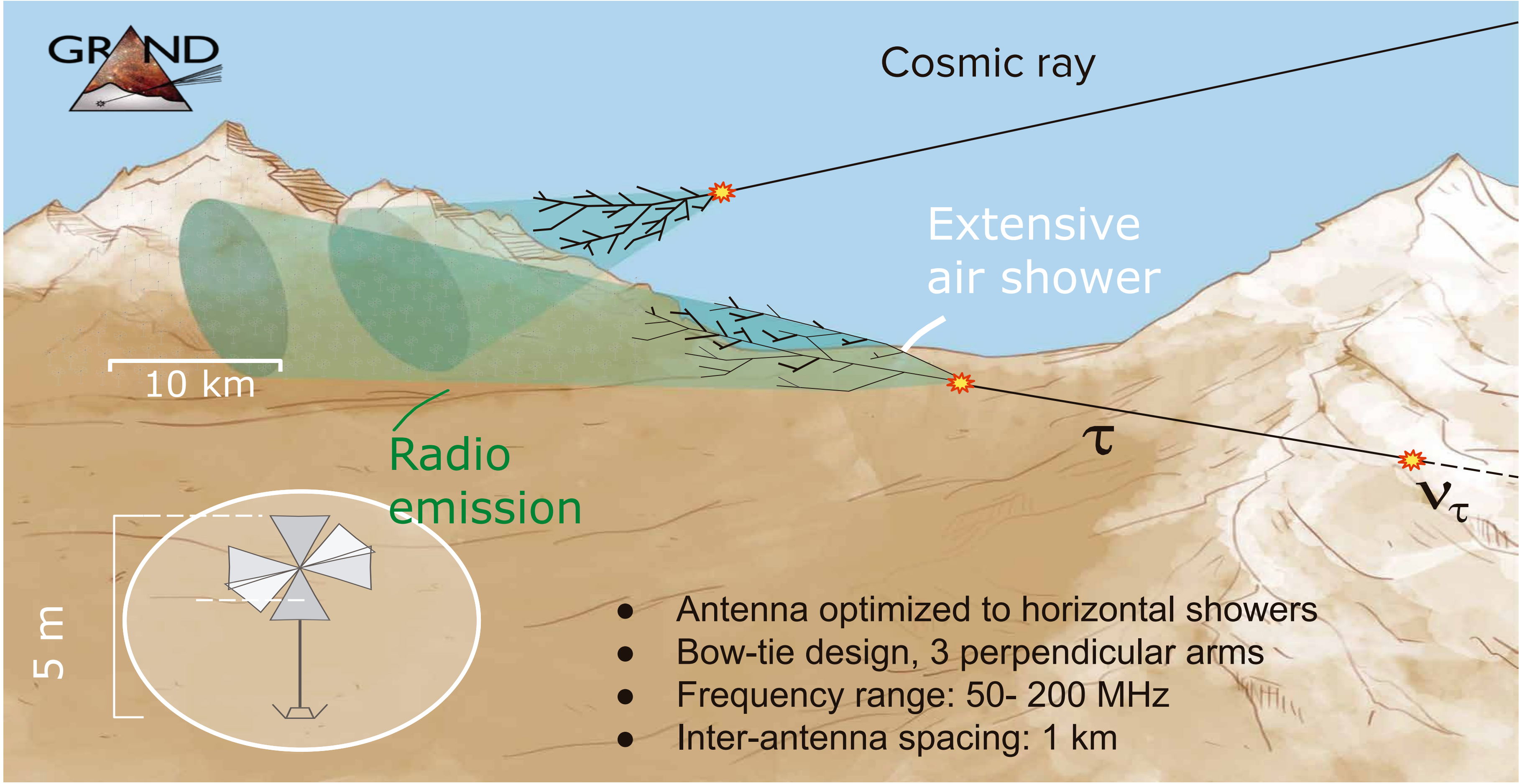
Zachary Mason
(with Oscar Macias & the GRAND ML working group)

Workshop on Machine Learning for Analysis of High-Energy Cosmic Particles

Jan. 27th, 2025

Background

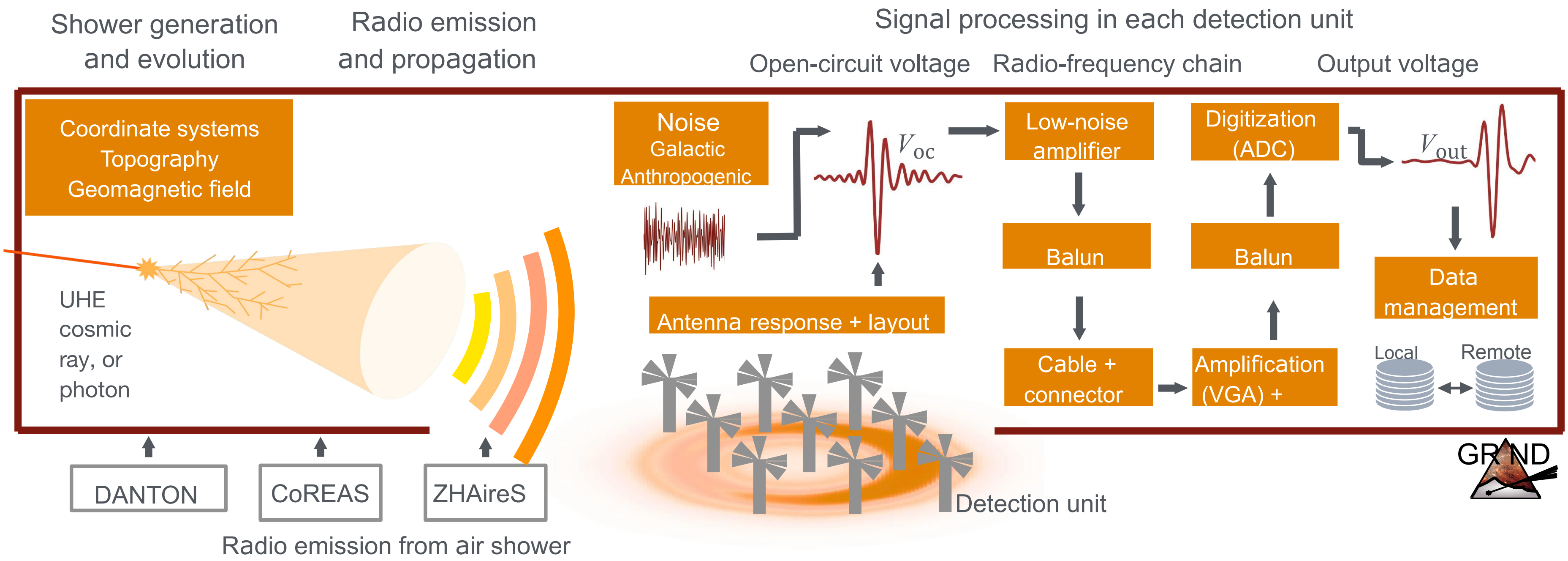
GRAND (Giant Radio Array for Neutrino Detection)



- Antenna optimized to horizontal showers
- Bow-tie design, 3 perpendicular arms
- Frequency range: 50- 200 MHz
- Inter-antenna spacing: 1 km

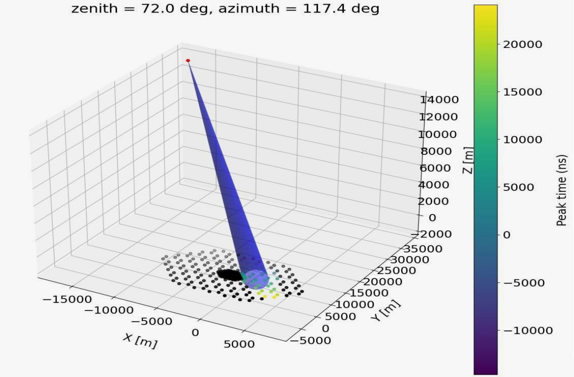
GRANDPipeline

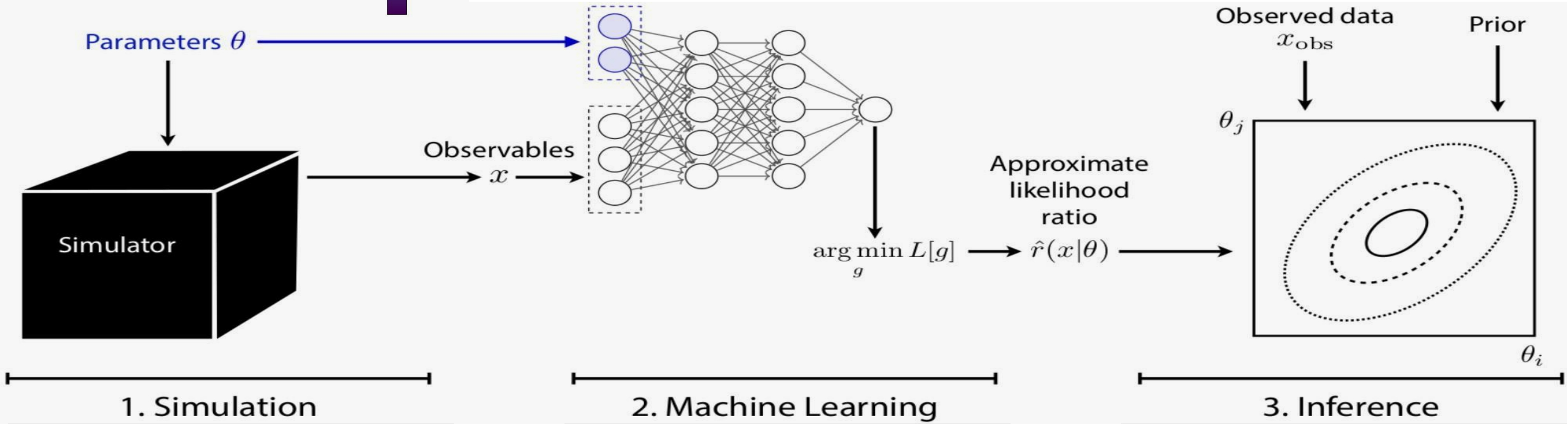
GRAND Collaboration. arXiv:2408.10926



[htt ps://github.com/grand-mother/grand](https://github.com/grand-mother/grand)

Direction Reconstruction with simulation-based inference (LtU-ILI)





~8,000 DC2 Simulations (NJ)
(Filters: events > 5 antennas, >60 micro-Volt/m)

Graph Convolution Network

Neural Autoregressive Flow
+
Masked Autoregressive Flow

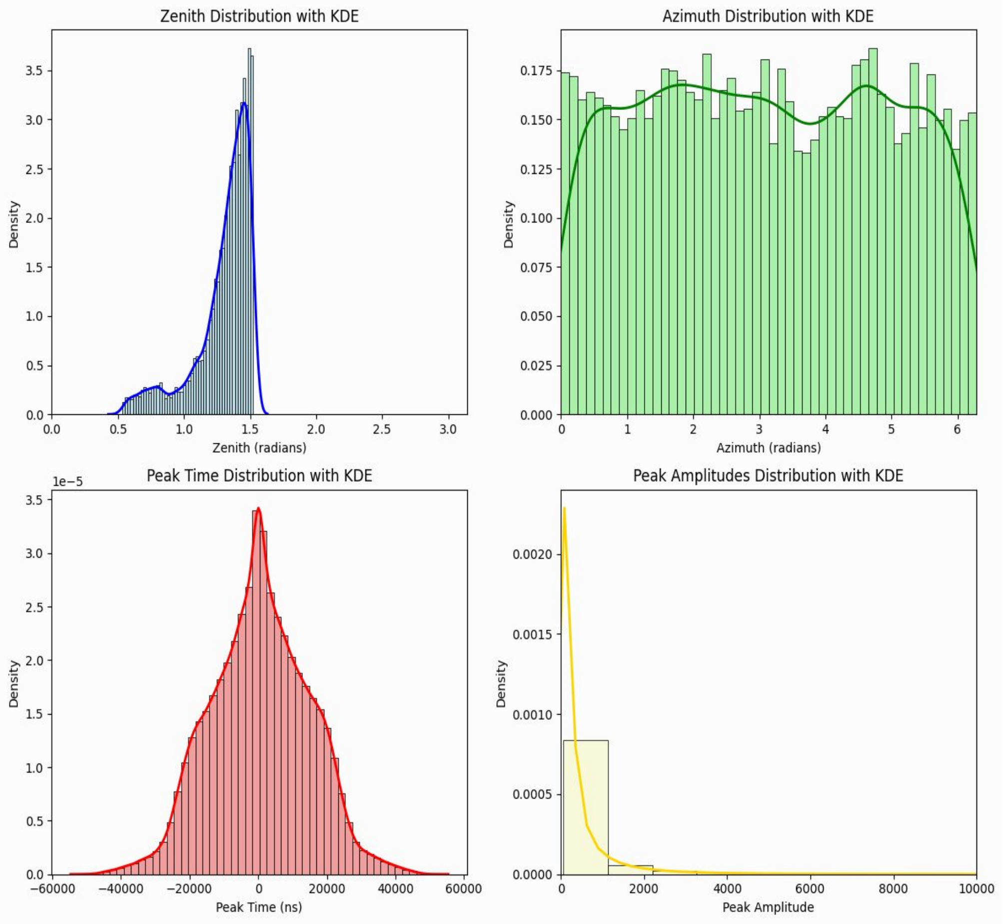
Credit: G. Louppe

ArXiv:2402.05137

LtU-ILI **Ho+ (2024)**

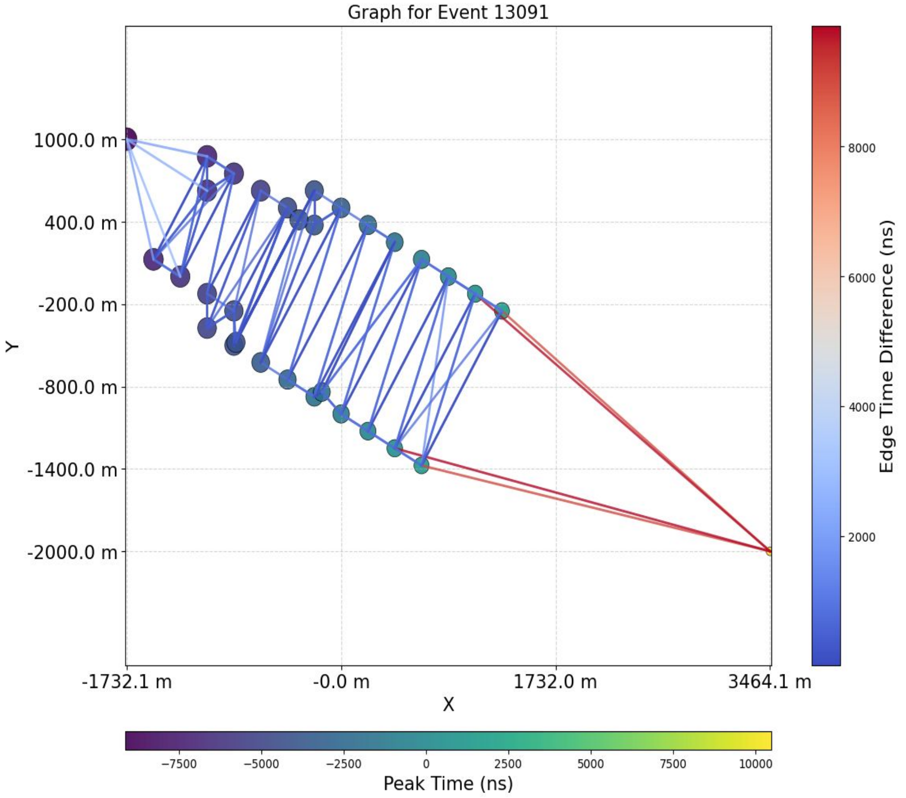
Implementation (DC2 Training Data)

- GRAND Collaboration database of **10,000 ZHAireS simulations**
- Filter simulations based on voltage response level (**>60uV**) and total number of triggered antennas (**>5 antennas**)
- Left with ~**8,200 events** for training our GCN



Implementation (Inference)

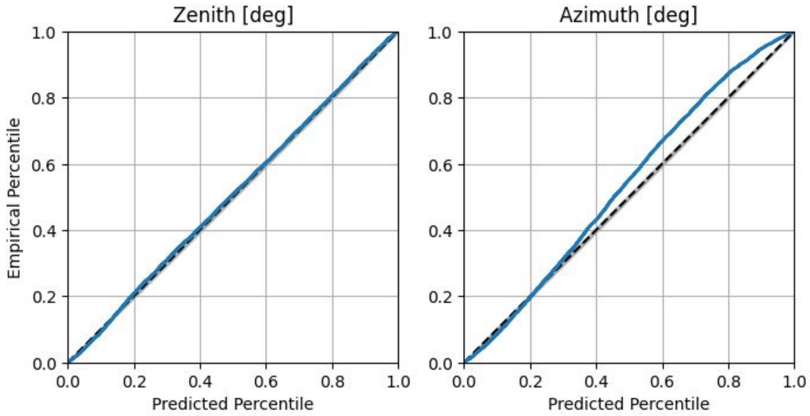
- ➔ Pass in a simulated event from DC2 training data to the GCN
  - ◆ **Nodes** created from antenna location & trigger time
  - ◆ **Edges** based on temporal distance between neighboring antennas utilizing k-nearest neighbors (kNN)
- ➔ Direction of air-shower is implied from resulting graph architecture
- ➔ Outputs posterior distributions of input data-parameter pairs



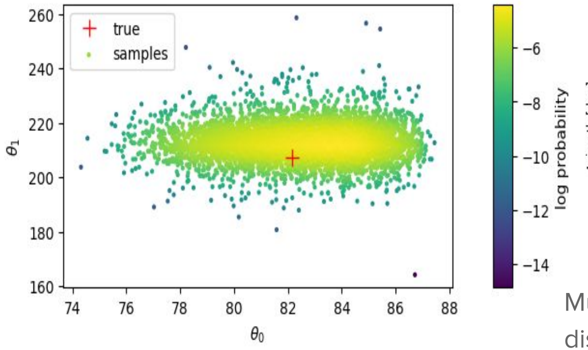
k = 5 nearest neighbors

Implementation (Model Validation)

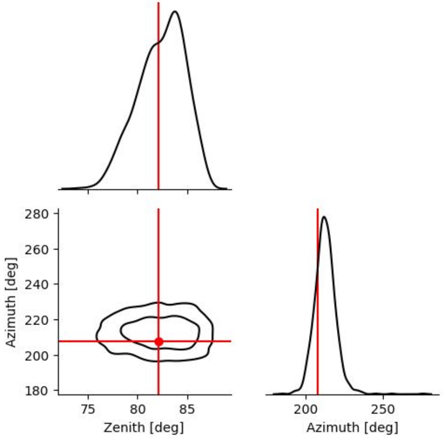
- ~8,200 simulated events (80/20 split for training/validation)
- Takes in posterior distributions from inference step
- Direct comparison to the data-parameter pair distribution from 20% validation dataset



Percentile Coverage Test/P-P plots (error comparison)



Marginal True vs Predicted Plots

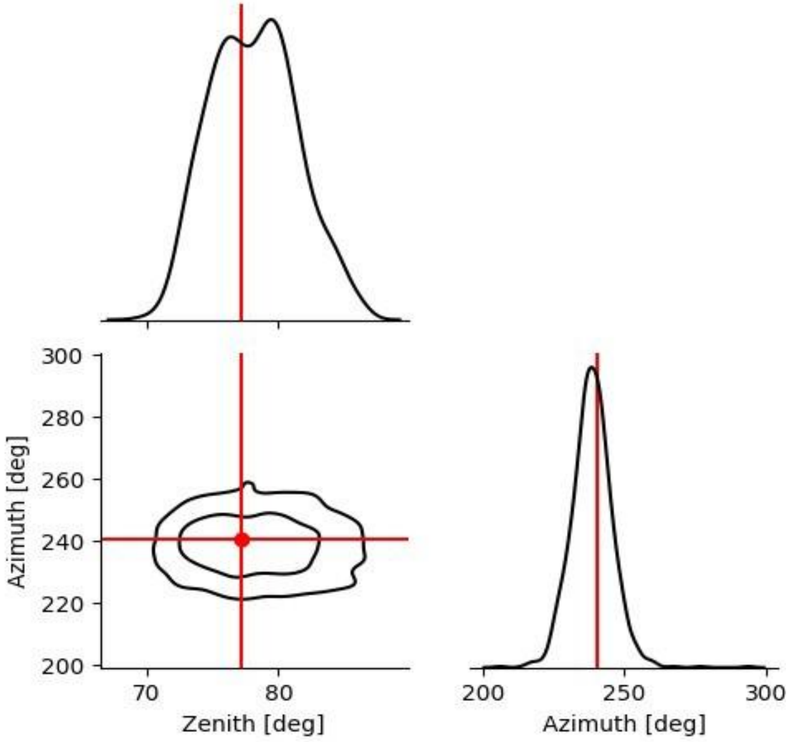


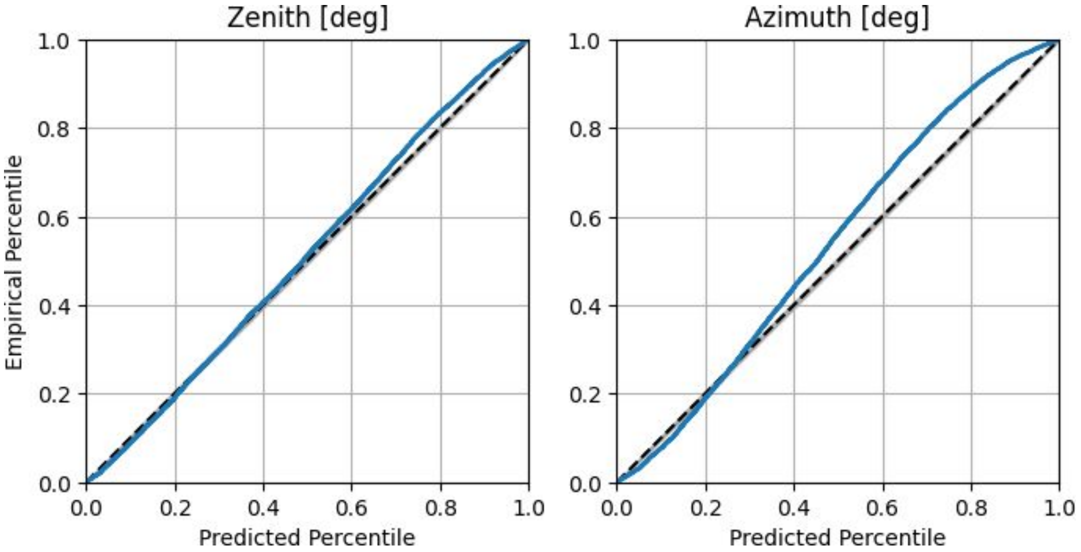
Multivariate corner plots of posterior distributions (model precision)

ArXiv:2402.05137 **LtU-ILI** **Ho+ (2024)**

Preliminary Results

Preliminary Results (data driven)





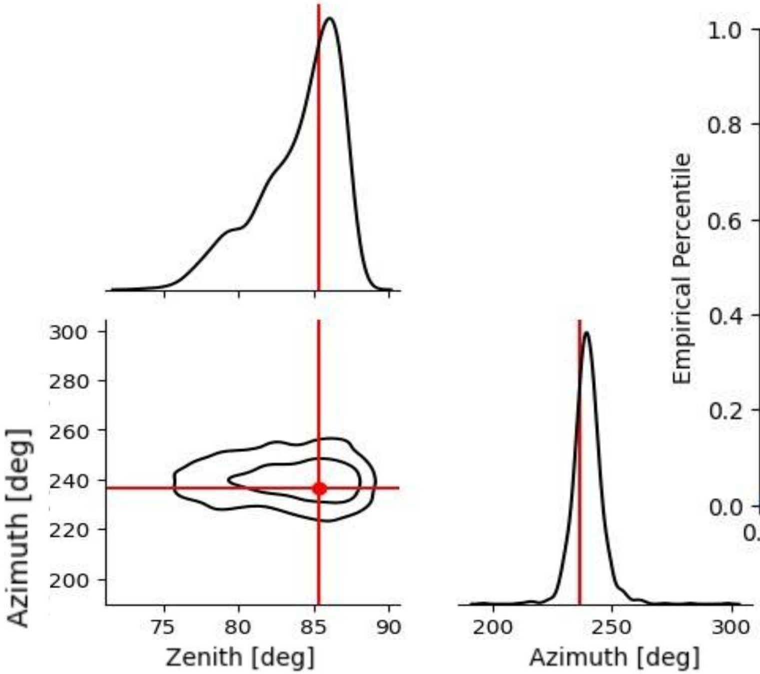
Direction resolution ~10 deg (within 1-sigma of true value indicated by red point)

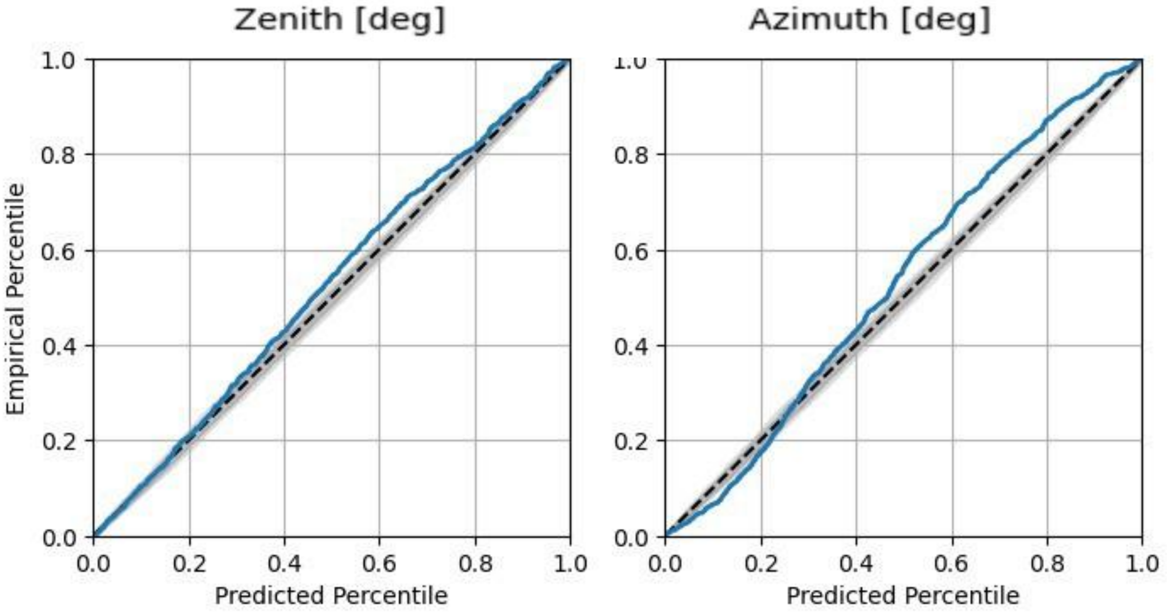
➡ Tuned **k** for graph construction, # of channels on GCN, output channels & drop rate

➡ Best results using:

[k = 5, in_channels = 4, gcn_channels = [16,32], out_channels = 8, drop_p = 0.05]

Preliminary Results (data driven)





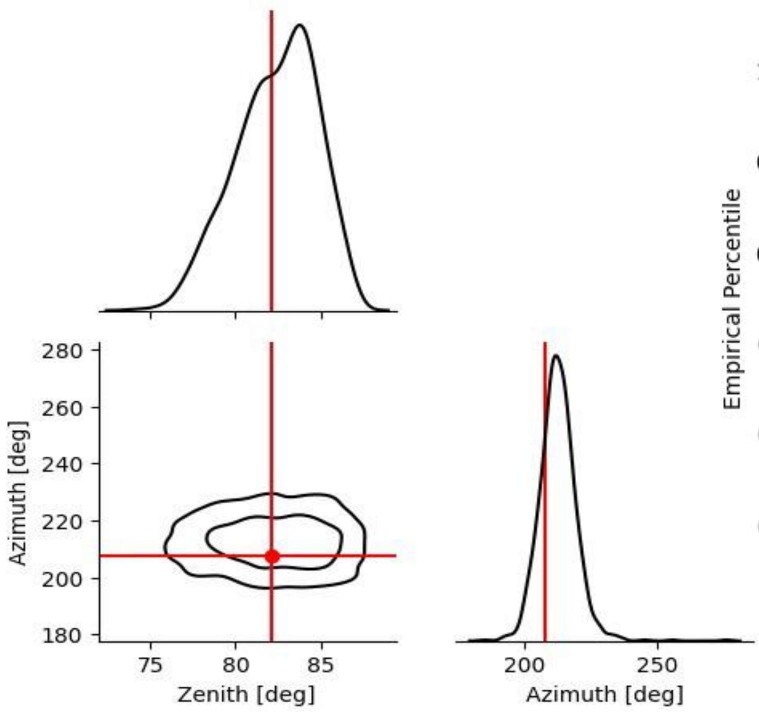
Direction resolution ~5-10 deg (within 1-sigma of true value indicated by red point)

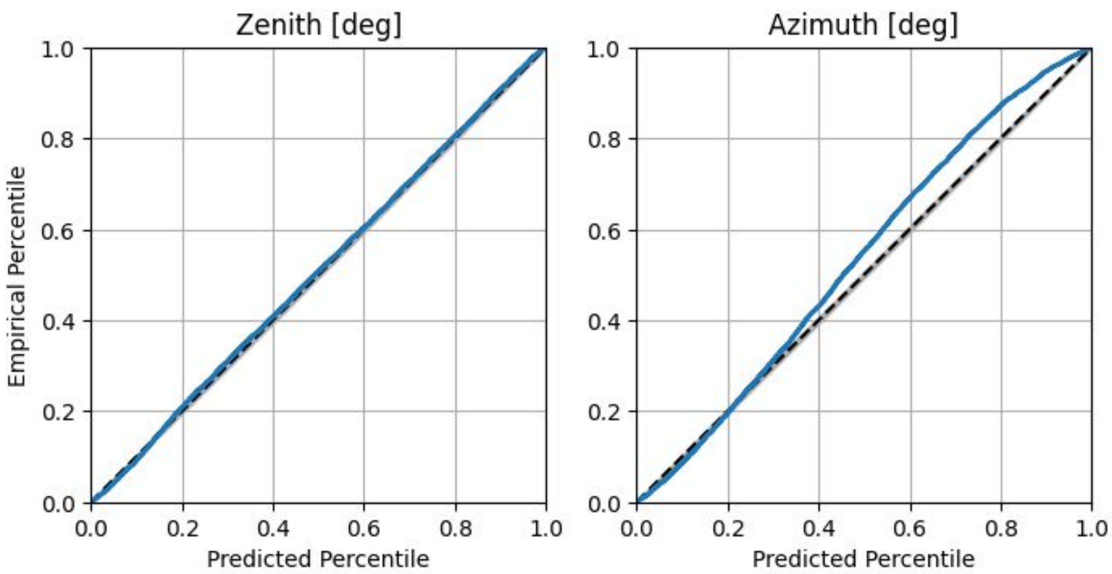
➔ Tuned **k** for graph construction, # of channels on GCN, output channels & drop rate

➔ Best results using:

[k = 5, in_channels = 4, gcn_channels = [16,32], out_channels = 8, drop_p = 0.05]

Preliminary Results (data driven)





Direction resolution ~5-10 deg (within 1-sigma of true value indicated by red point)

➔ Tuned **k** for graph construction, # of channels on GCN, output channels & drop rate

➔ <u>Best results using:</u>

[**k = 5, in_channels = 4, gcn_channels = [16,32], out_channels = 8, drop_p = 0.05**]

Summary

- ➔ Successful implementation of SBI methodology to reconstruct posterior distributions & estimate parameter errors
- ➔ Purely data-driven approach correctly reproduces UHE cosmic ray direction within 5-10 degrees resolution

Next Steps

- ➔ Implement a physics-informed approach to the ML model, hoping to achieve sub-degree resolution (*cf. Arsene Ferriere talk*).

Thank you!

Backup

Extra

- Using **“lampe”** implicit inference backend offered by LtU-ILI for it’s variety of NDEs & greater flexibility in embedded network choice
- Wanted to use a graph type embedded network given the complexity of the training data
- Best results using GCN:
    - [**k** = **5, in_channels** = **4, gcn_channels** = [**16,32**], **out_channels** = **8, drop_p** = **0.05**]
- Tried Graph Attention Network (GAT) as embedded network at first, but it struggled to learn the posteriors, mostly recreated entire training data distribution