ACAT 2017
Seattle

High Statistics and GPU Accelerated Data Analysis in IceCube

Philipp Eller (pde3@psu.edu)
for the IceCube collaboration

PennState

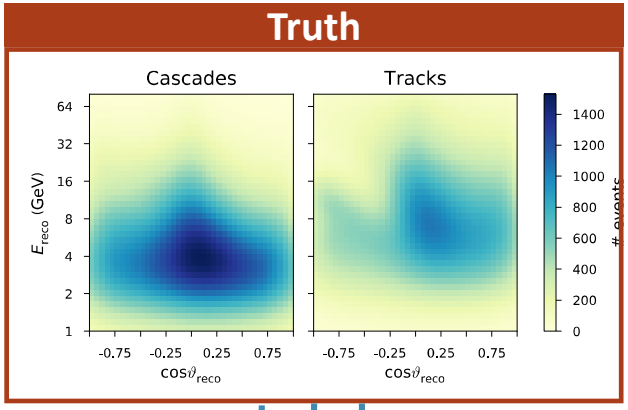
Limited Monte Carlo

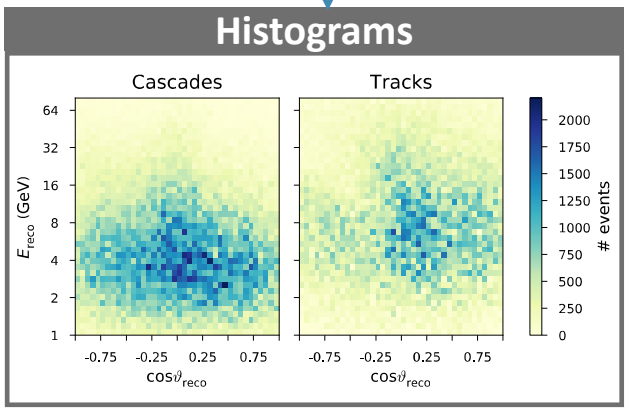
- Experiments usually need loads of MC, especially for high statistics data analyses
- MC production (simulation, reconstruction, …) can be very **time and resource consuming**
  - Yesterday we heard: HL-LHC MC production and storage alone may cost ~$1B (one billion!)
- Furthermore, handling huge MC sets is often **slow**

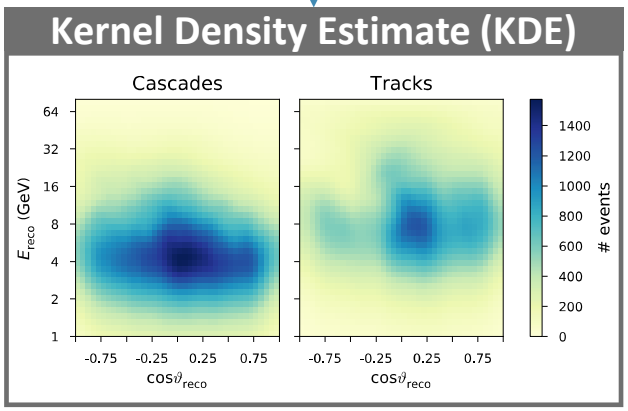
- **-> Can we just live with less MC?**

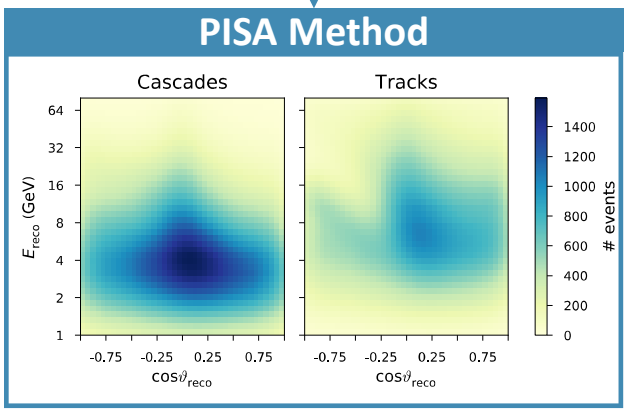
Smoothing techniques

- To mitigate statistical noise
- For illustration, using a toy model with known true pdfs
- identical 10k event samples to estimate original density



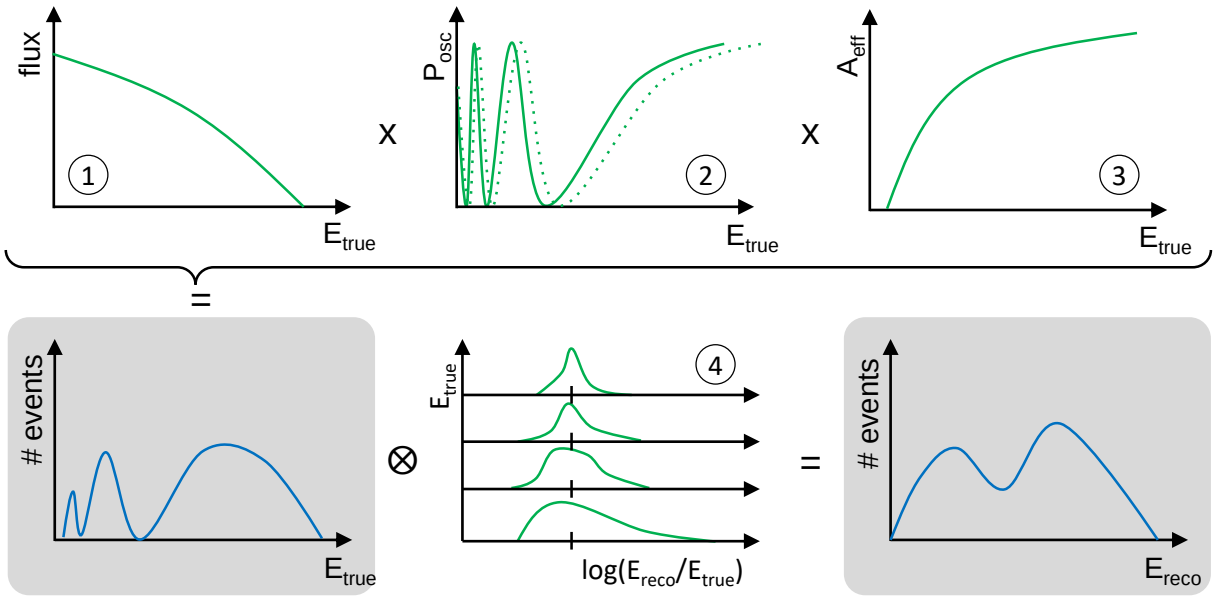






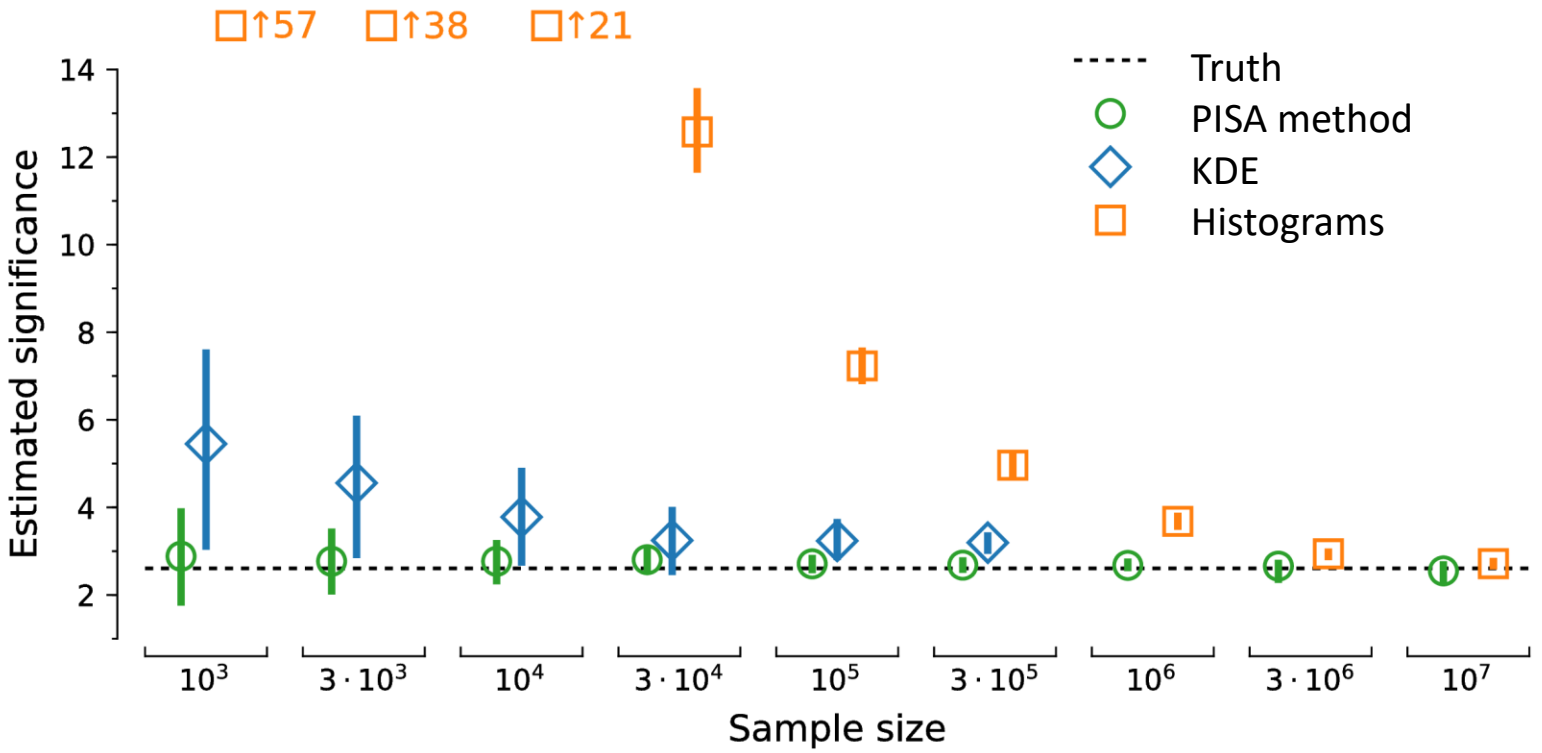
PISA Method

- Instead of applying smoothing to final events
  - Use MC to characterize detector resolutions (4) and selection efficiencies (3)
  - Apply smoothing to these well behaved distributions
- Here: Convolute with atmospheric flux (1) and oscillation probabilities (2) to arrive at final spectra



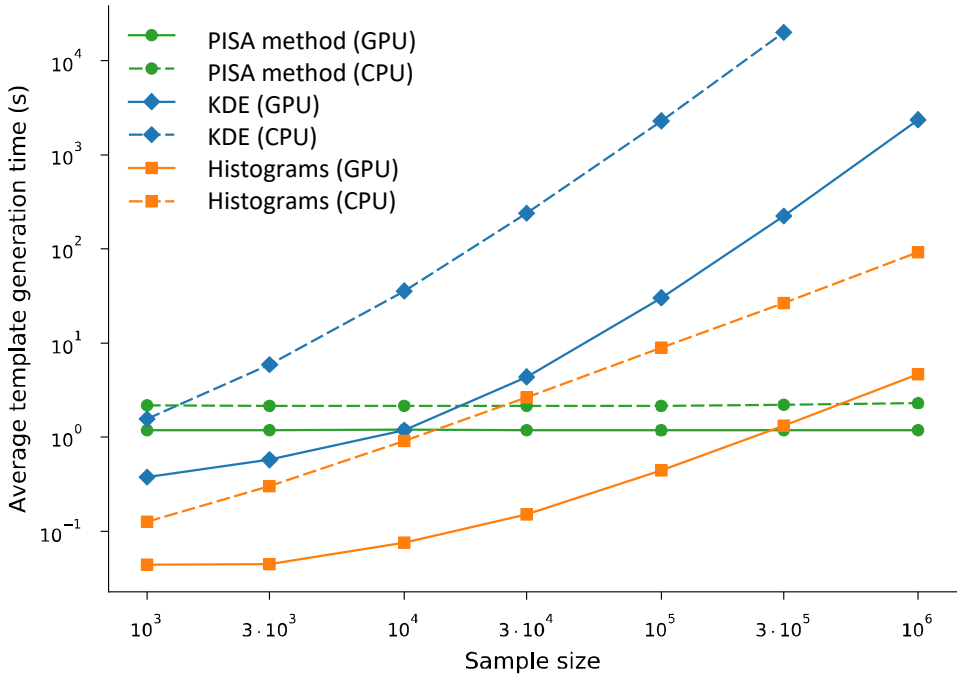
Effect of low MC statistics

- Example analysis with true result = 2.7σ
- Small sample sizes can lead to (grossly) wrong results
- PISA method virtually independent of input MC size
  - -> Can save **orders of magnitude** of MC



GPU acceleration

- Also speed is an issue
  - We accelerated all mentioned methods with Nvidia CUDA to run on GPUs
- Example:
  - Simple Histogram method took 16h to run a full analysis
  - After GPU and other optimizations: ~2 minutes
  - Facilitates develop-ment of analysis
  - Allows to run more intensive statistical procedures (e.g. Feldman-Cousins confidence Interval calculation)



THE END