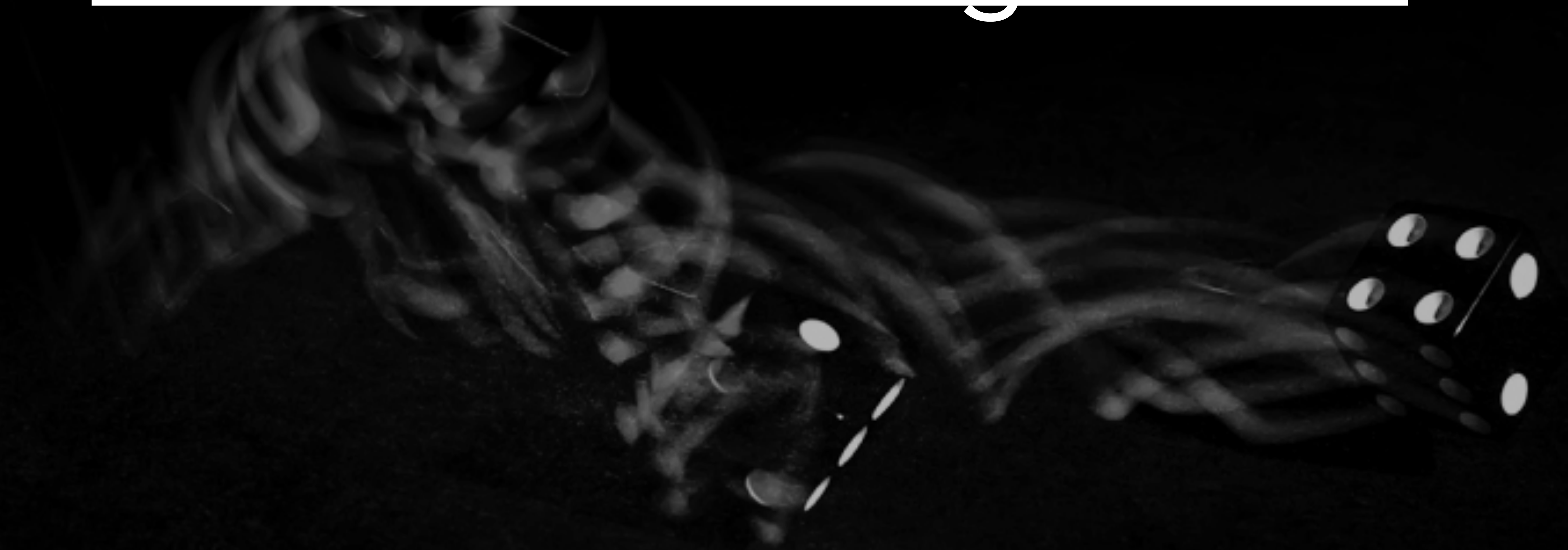
Subthreshold Anomaly Detection & Merged Bins



D. Jason Koskinen
koskinen@nbi.ku.dk

Advanced Methods in Applied Statistics
Feb - Apr 20XX

Photo by Howard Jackman

Info

- This ‘guest’ lecture is about an interesting statistical topic, but it won’t be on the exam

- New statistical tool that people can think about adding to their analysis and research

Background Story

- There is a neutrino oscillation anomaly that has existed in a few experiments since ~1995
  - If the anomaly has a particle physics motivation, it can be tested by experiments
    - Not all neutrino oscillation experiments see the anomaly
    - Particle physics motivation implies a new fundamental particle, i.e. a sterile neutrino

The Big Table of Disagreement

*arXiv:1803.10661

Analysis	$\chi^2_{\text{min,global}}$	$\chi^2_{\text{min,app}}$	$\Delta\chi^2_{\text{app}}$	$\chi^2_{\text{min,disapp}}$	$\Delta\chi^2_{\text{disapp}}$	$\chi^2_{\text{PG}}/\text{dof}$	PG
Global	1120.9	79.1	11.9	1012.2	17.7	29.6/2	3.71×10^{-7}
Removing anomalous data sets							
w/o LSND	1099.2	86.8	12.8	1012.2	0.1	12.9/2	1.6×10^{-3}
w/o MiniBooNE	1012.2	40.7	8.3	947.2	16.1	24.4/2	5.2×10^{-6}
w/o reactors	925.1	79.1	12.2	833.8	8.1	20.3/2	3.8×10^{-5}
w/o gallium	1116.0	79.1	13.8	1003.1	20.1	33.9/2	4.4×10^{-8}
Removing constraints							
w/o IceCube	920.8	79.1	11.9	812.4	17.5	29.4/2	4.2×10^{-7}
w/o MINOS(+)	1052.1	79.1	15.6	948.6	8.94	24.5/2	4.7×10^{-6}
w/o MB disapp	1054.9	79.1	14.7	947.2	13.9	28.7/2	6.0×10^{-7}
w/o CDHS	1104.8	79.1	11.9	997.5	16.3	28.2/2	7.5×10^{-7}
Removing classes of data							
$\bar{\nu}_e$ dis vs app	628.6	79.1	0.8	542.9	5.8	6.6/2	3.6×10^{-2}
$\bar{\nu}_\mu$ dis vs app	564.7	79.1	12.0	468.9	4.7	16.7/2	2.3×10^{-4}
$\bar{\nu}_\mu$ dis + solar vs app	884.4	79.1	13.9	781.7	9.7	23.6/2	7.4×10^{-6}

PG = Parameter goodness-of-fit,
i.e. the probability that the
selected data agree

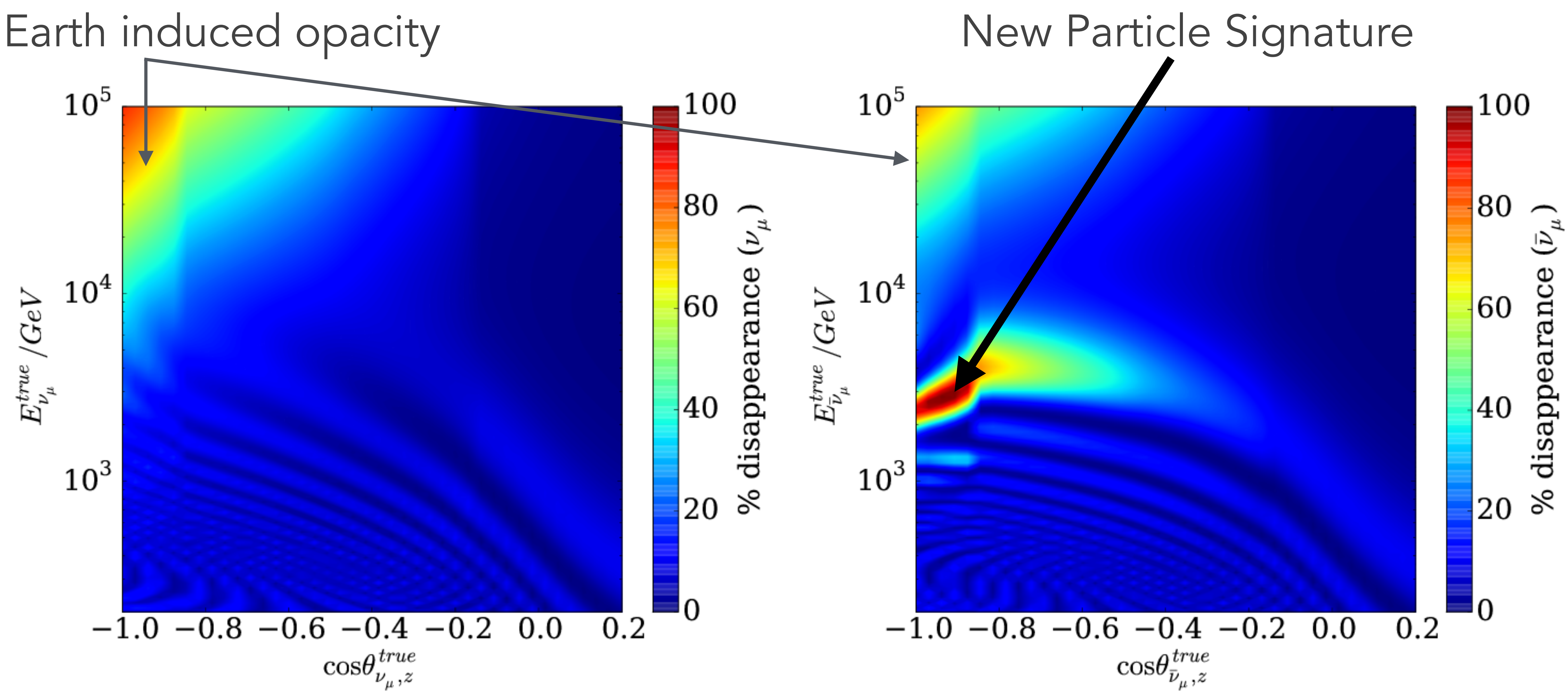
Background Story

- There is a neutrino oscillation anomaly that has existed in a few experiments since ~1995
    - If the anomaly has a particle physics motivation, it can be tested by experiments
        - Not all neutrino oscillation experiments see the anomaly
        - Particle physics motivation implies a new fundamental particle, i.e. a sterile neutrino

- Two separate analyses in 2015/2016 by IceCube looked for the sterile neutrino signature if it exists (it doesn't!)
    - arXiv:1605.01990
    - Analyses are extremely difficult and sensitive to systematic uncertainties

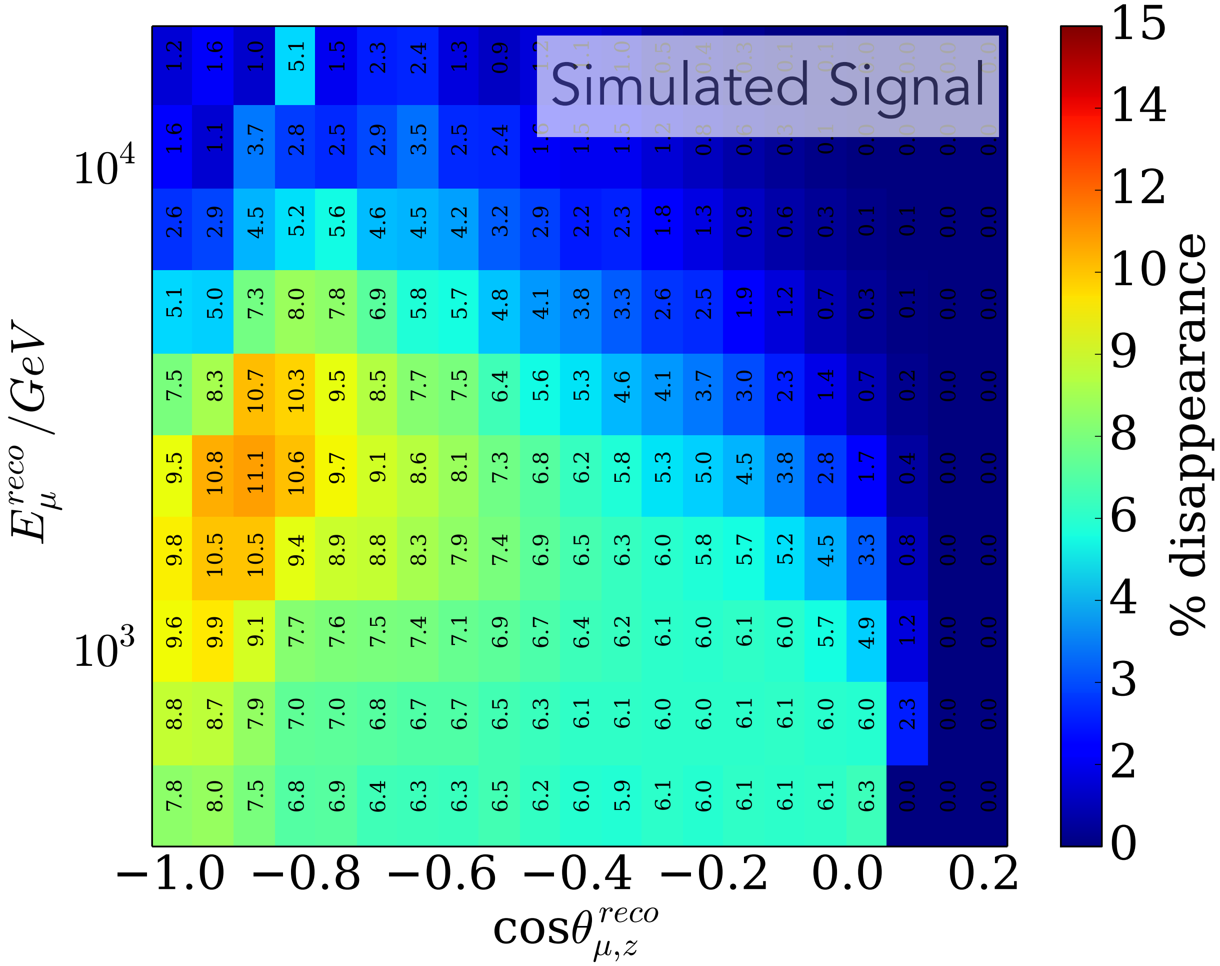
IceCube Experimental Signature

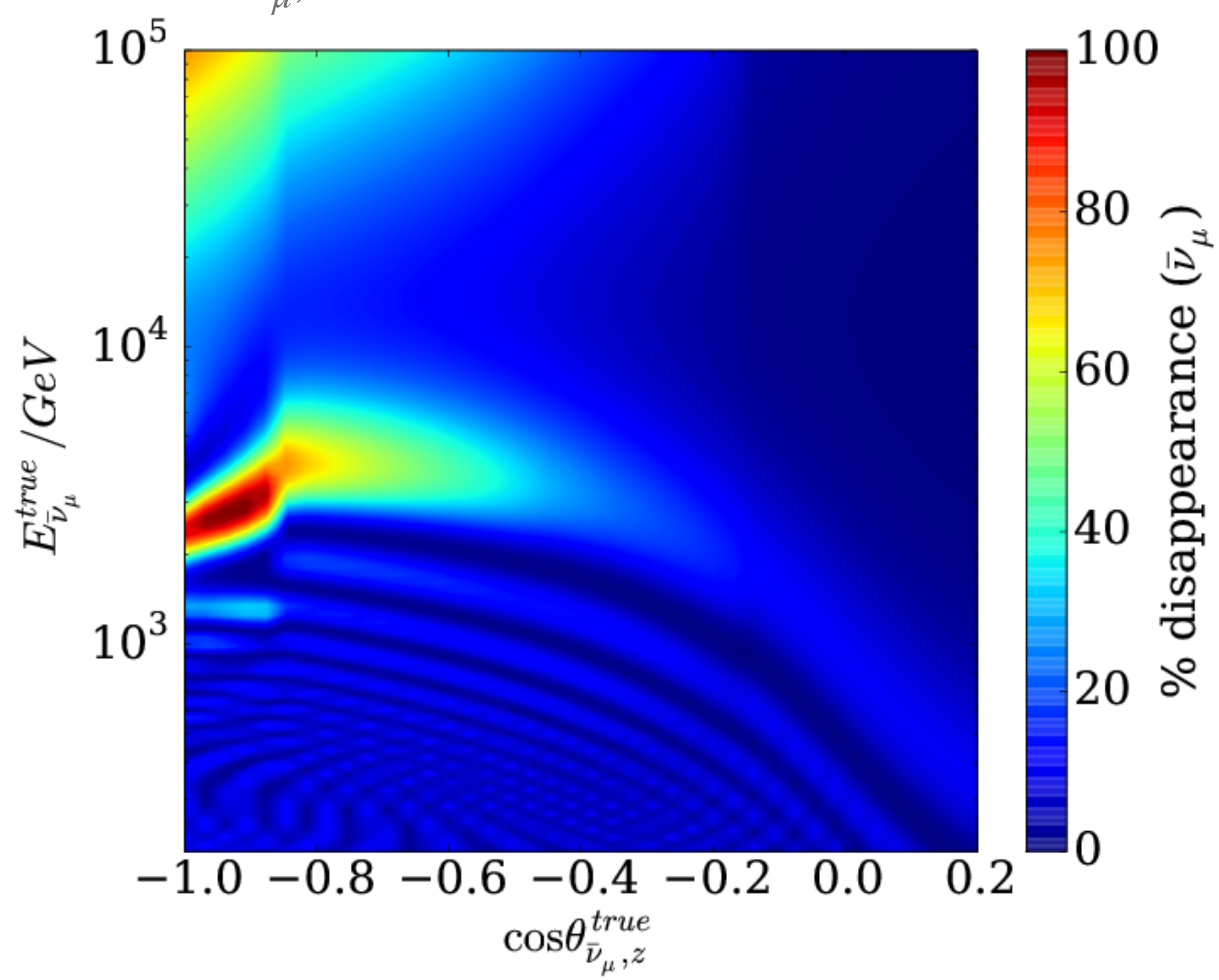
- Neutrino rate changes for, mostly, muon antineutrinos ($\bar{\nu}_\mu$)
  - Y-axis is neutrino energy
  - X-axis is the cosine of the incoming neutrino zenith direction



Sterile Search Approach

- Two separate diffuse ν_μ event selections of 1-year livetime were used to search for a sterile neutrino signal
- The pronounced sterile neutrino feature is smeared out by:
  - Reconstructed energy estimator - E_μ^{reco}
  - Reconstructed direction estimator - $\cos\theta_{\nu_\mu,\ z}^{reco}$





Residual/Pulls

"Everything you always wanted to know about pulls"
http://physics.rockefeller.edu/luc/technical_reports/cdf5776_pulls.pdf

- Analysis is done by fitting Monte Carlo simulation to match the observed data
  - Fit systematic uncertainty parameters
  - Fit physics parameters

- After fitting, check that the best-fit expectation 'matches' the data by looking at the pulls (also known as residuals)

$$pull_{i,j} = \frac{Exp_{i,j} - Obs_{i,j}}{\sigma_{i,j}}$$

i=bin of neutrino energy
j=bin of cos(zenith)
$\sigma_{i,j}$=uncertainty

  - Expected number of neutrinos in each bin compared to observed
  - Distribution should be 'mostly' Gaussian-like if the fit is good
  - For statistical-only residuals the uncertainty is sqrt(# Expected events)
  - Each pull value can be loosely treated as a 'sigma' value

Exercise 1

- Some PDF gives expected number of events in our 2D phase-space of (γ, ρ)
  - 3 bins in γ and 3 bins in ρ
  - PDF can either be from a model w/ a best-fit or predefined

- Create the following pull plots using the data
  - 2D pull plot; can be just color-based representing the pull value
  - 1D projection; histogram of all the individual residuals

Expectation

γ \ ρ			
	99	99	99
γ	99	100	99
	98	100	99

ρ

Observation

γ \ ρ			
	102	90	101
γ	97	103	92
	83	111	96

ρ

Exercise 1 - Extra

- If you have time, you can generate pseudo-experiments by statistically fluctuating the expectation bin counts and produce multiple pull distributions
  - In these pseudo-experiments, the ‘observed’ is the Monte Carlo statistically fluctuated number of events in each bin
  - Uncertainty is still sqrt(# of Expected events)

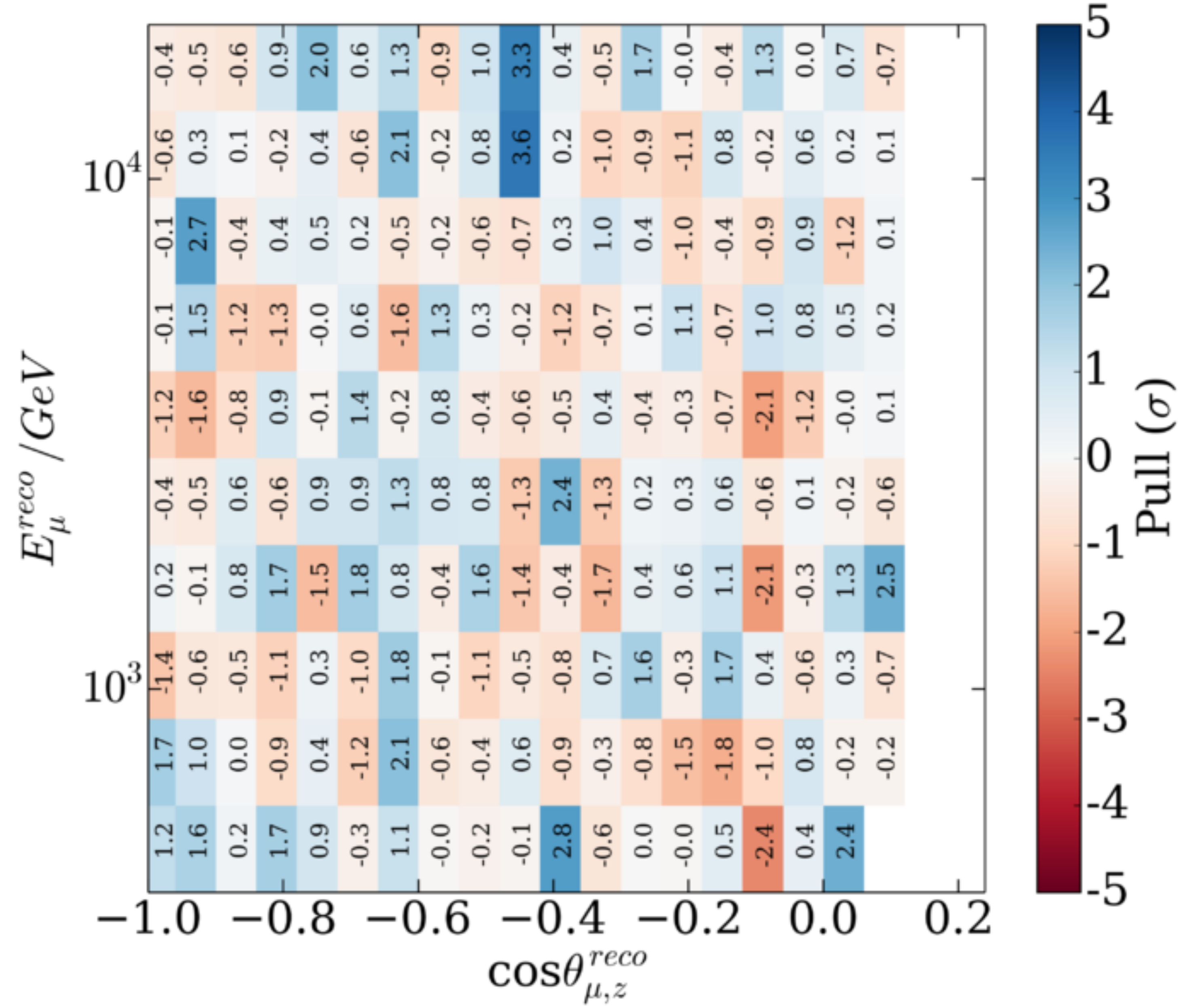
Expectation

γ \			
	99	99	99
γ	99	100	99
	98	100	99

ρ

Pull Plot - Analysis A

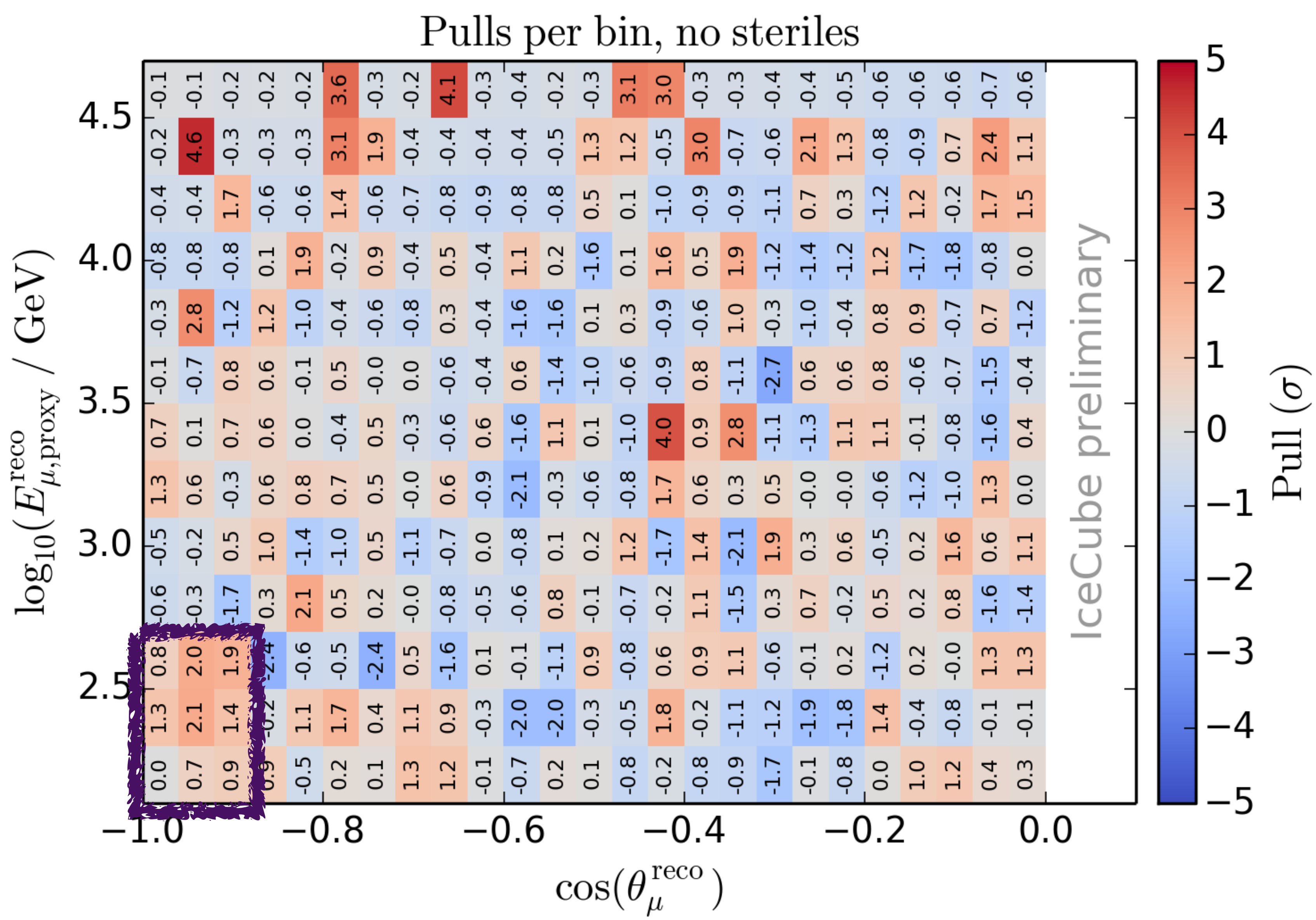
- Each bin gets a pull value
  - Low statistics bins can have very large or small pulls that are non-gaussian
  - $>10^4$ GeV is low-statistics region
- The pulls can be used as a test-statistic
  - Histogram either the pulls or the abs(pulls)
  - Compare to Monte Carlo pseudo-experiments to get a p-value
- Using only statistical uncertainty, the p-value for the shown pull plot is 17%



Pull Plot - Analysis B

IceCube Internal
Not for Reproduction

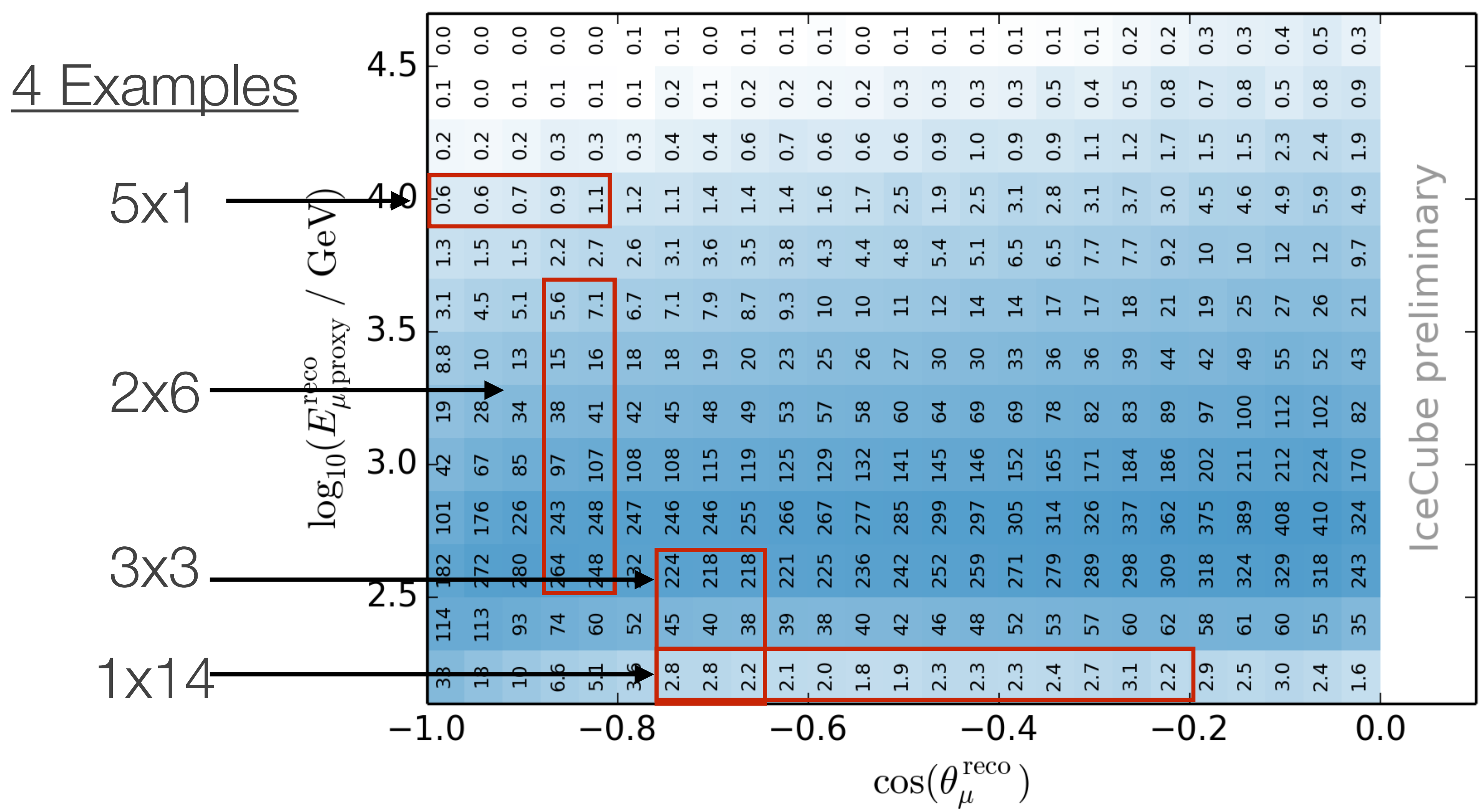
- 1D pull distribution looked 'okay'
    - 200+ bins with reasonable statistics
    - Uncertainty was statistical-only



- Bin widths and ranges are set by analyzers

The Method

- Take the 2D histograms of data and expectation events, and compare all possible combinations of adjacent bins and scan over the full region of phase space



Backup

Back in a bit, getting
some coffee &
stretching legs!
-Jason