Accounting for MC statistics in T2K ND fit

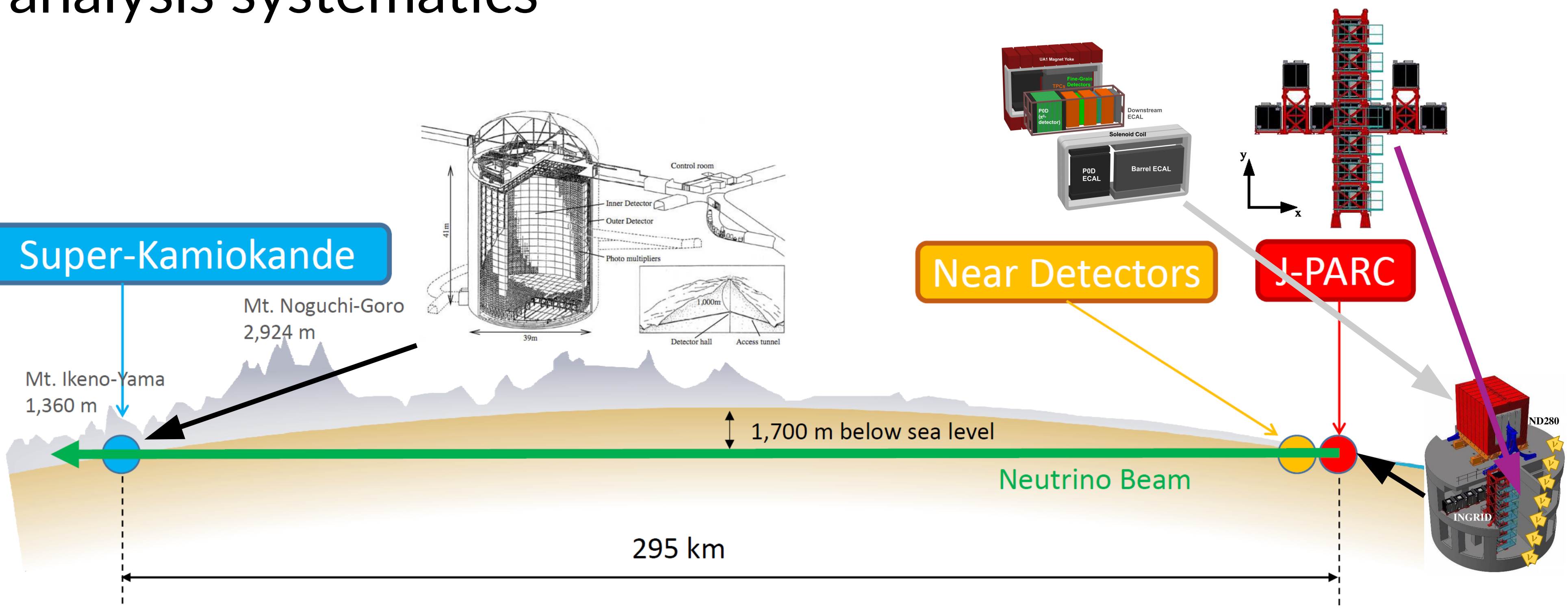
Hopefully useful to other analyses too!

Clarence Wret

Rochester long update
4 November 2019

Overall introduction

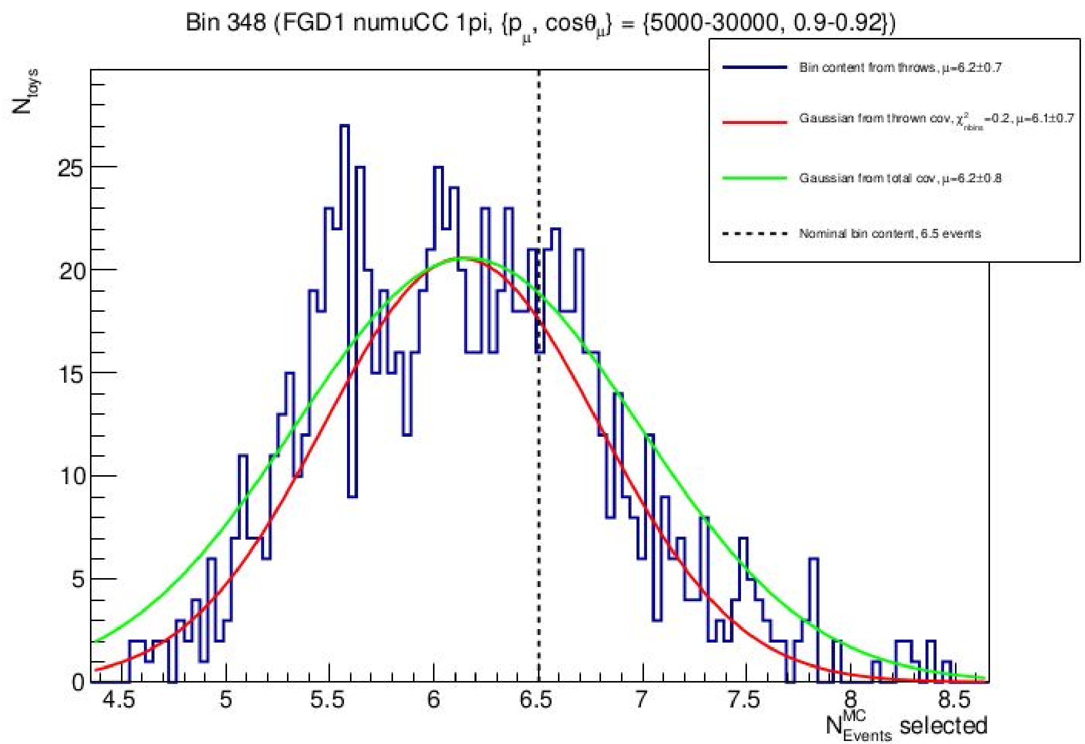
- T2K uses ND280 (near detector) to constrain oscillation analysis systematics



- Systematics: neutrino flux (100), neutrino cross-sections (~30), ND280 detector parameters (~600), SK detector parameters (~30)
- Yuck, so many ND280 parameters

Introduction

- We make a covariance matrix for the ND280 systematic parameters
  - Vary detector parameters simultaneously, making "toy experiments"
- Look at how bin content changes for the different universes
- Merge bins with similar responses to systematic variations
  - Brings down nbins from ~4100 bins to 1076 in my thesis



Introduction

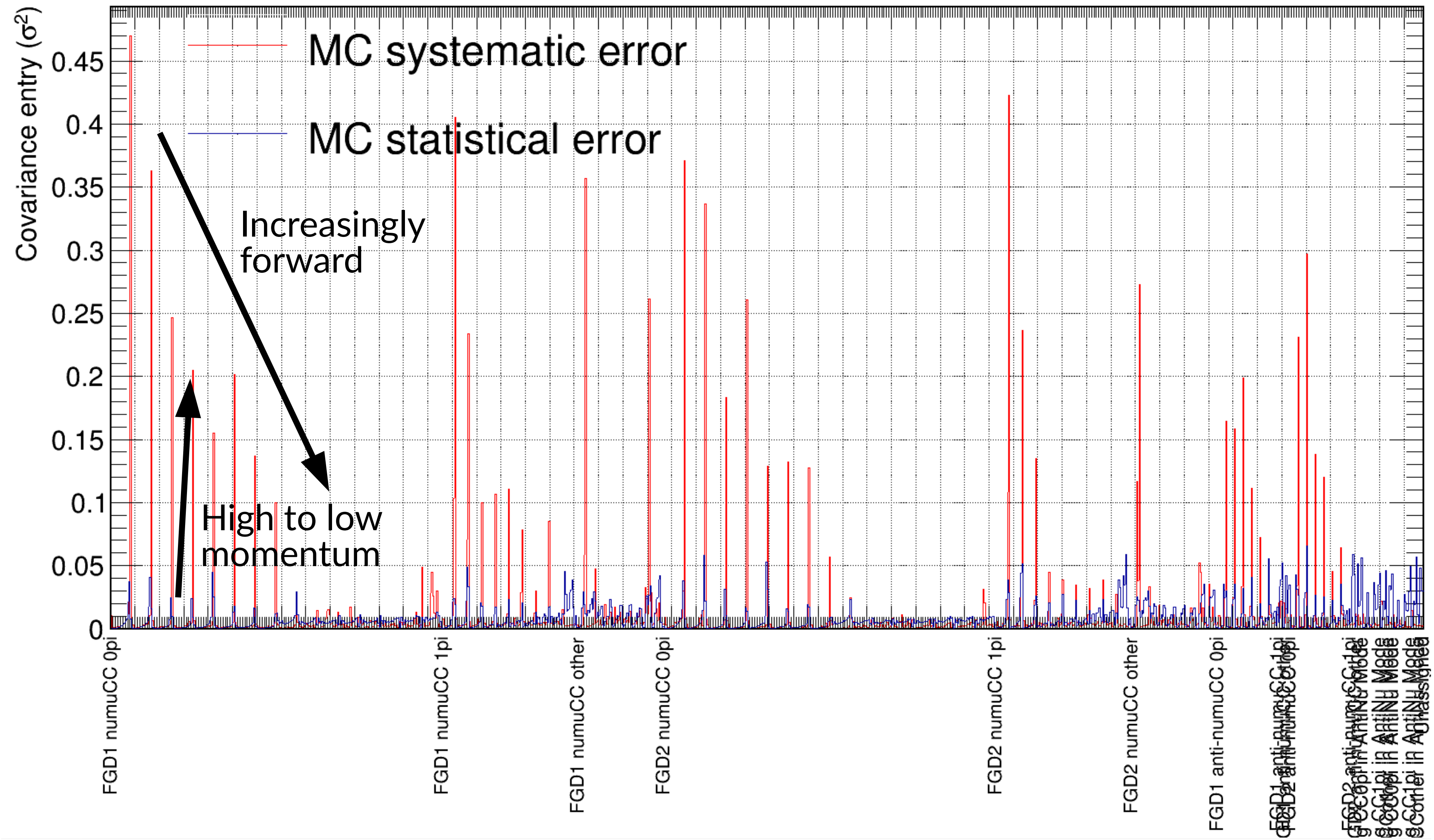
- We then add in MC statistics to create our total covariance matrix

$$V = V_{ND280} + V_{MC\ stat}$$

- MC statistics added as 1/(ηn) in every bin, where (ηn) is the **unscaled nominal weight** generated MC
  - η is the POT scaling, n is the number of POT scaled events (that we actually use as a prediction in a bin)
  - Essentially, MC stat accounts for that we didn't generate infinitely much MC → Uncertainty
- Adding covariances like this might be bad for PCA and bin merging: very different sources of covariance
  - Using merged binning for MC statistics uncertainty
- **Can we separate out MC statistics in a better way?**

Inspecting the current ND280 cov

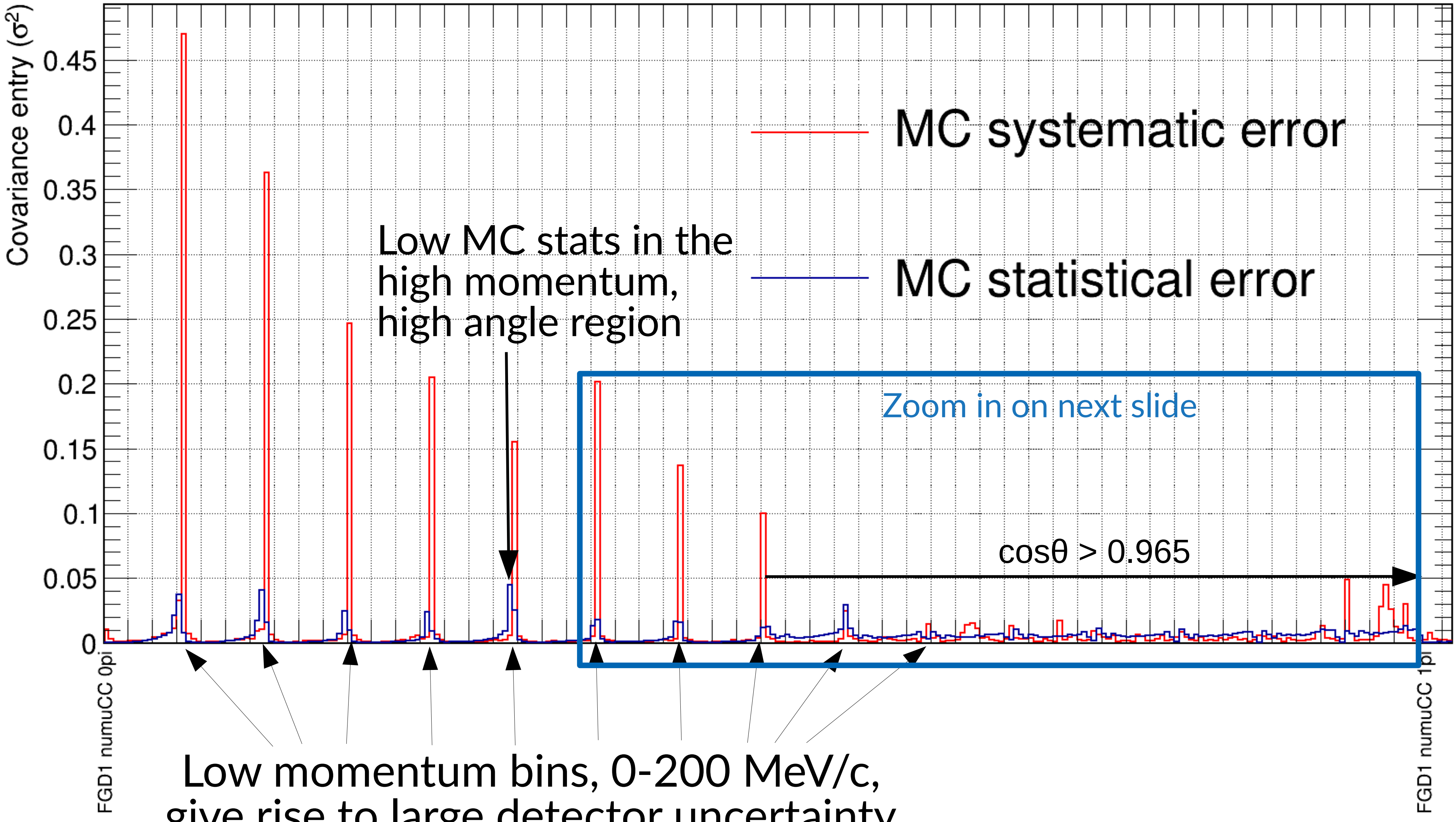
- Looking at the diagonal elements of frac. cov. for bin number



- Clear structure from low momentum and high angle bins

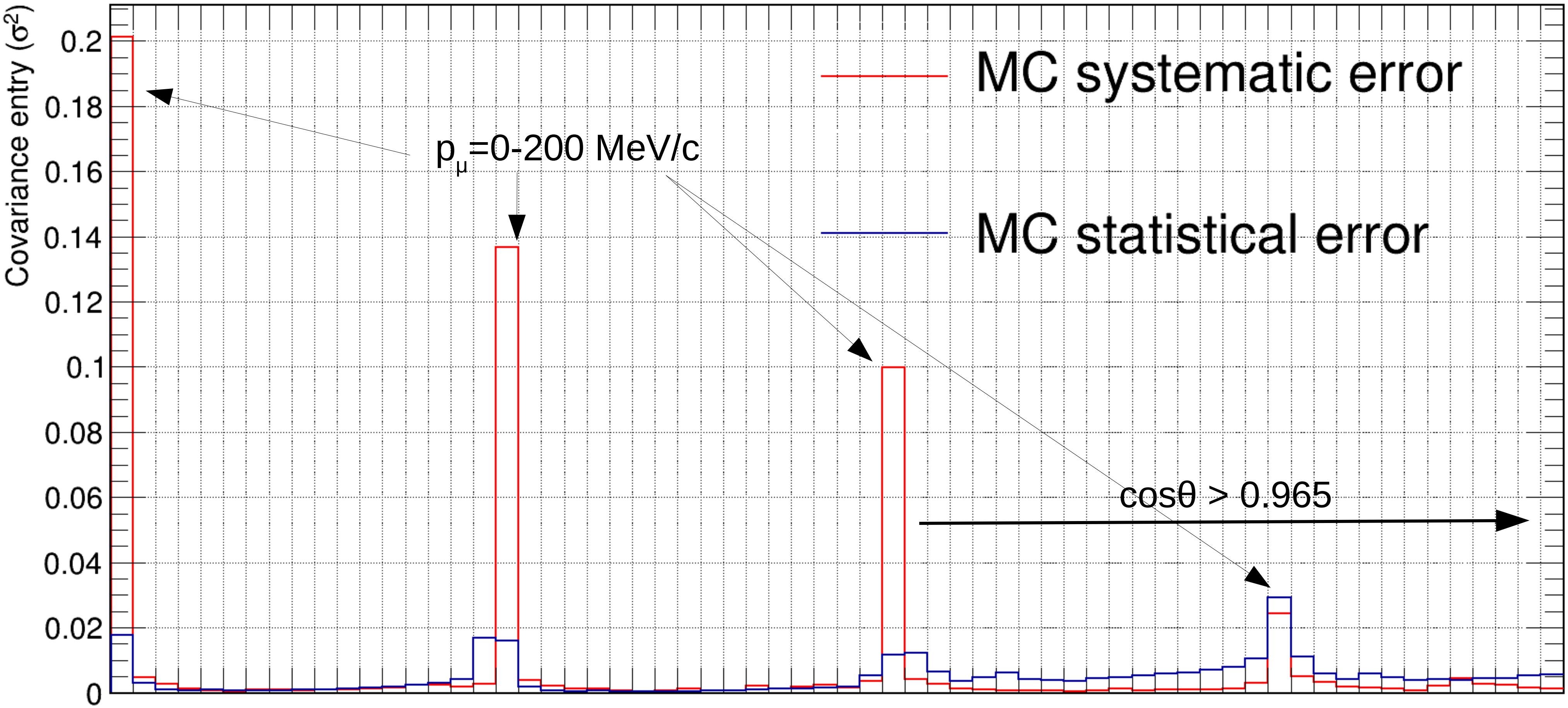
Inspecting the current ND280 cov

- Isolate FGD1 CC0π selection; structure clearer



Inspecting the current ND280 cov

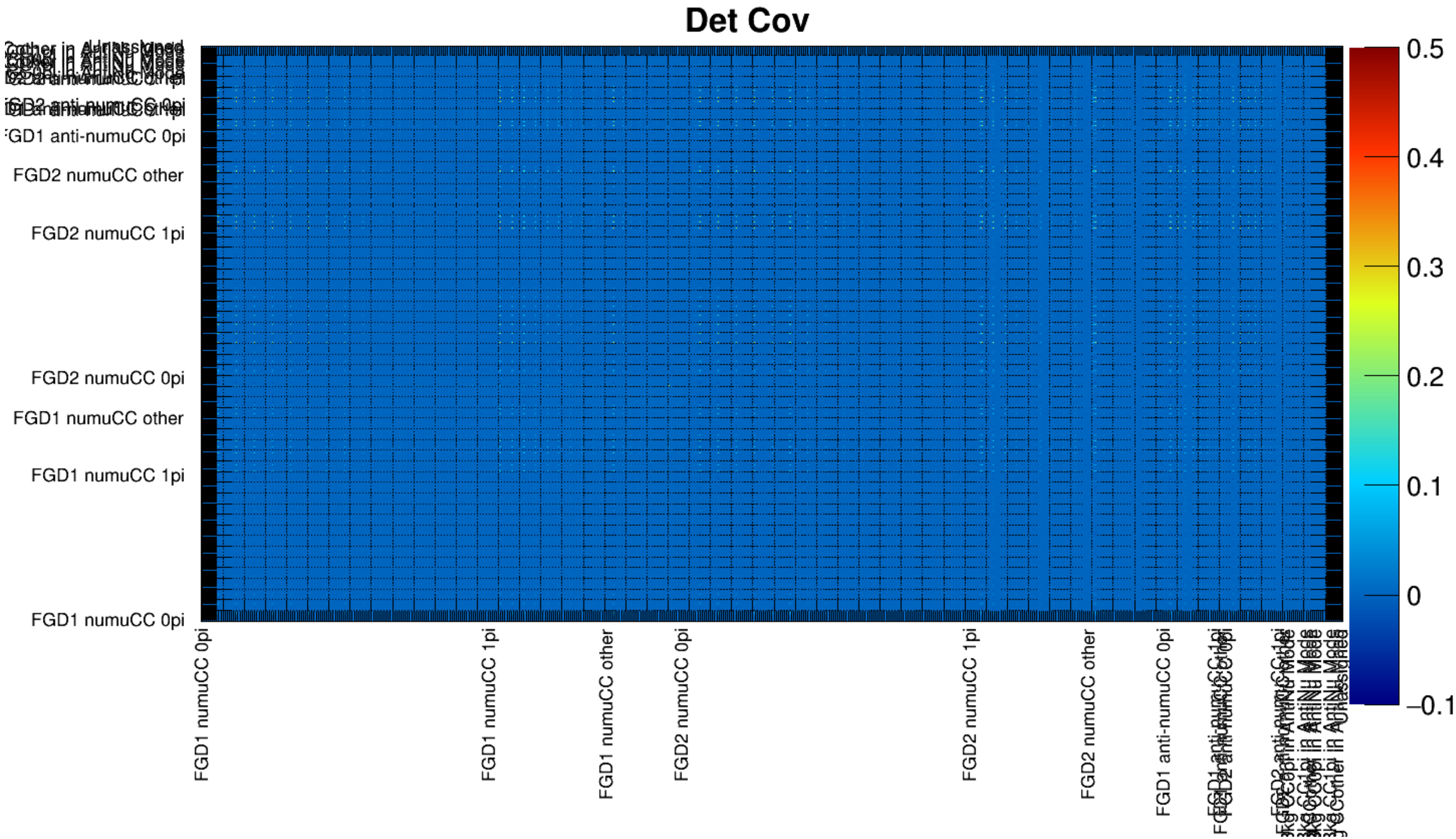
- Look at specific part where MC stat becomes dom.



MC stat is a large contribution in the forward region across the whole momentum range

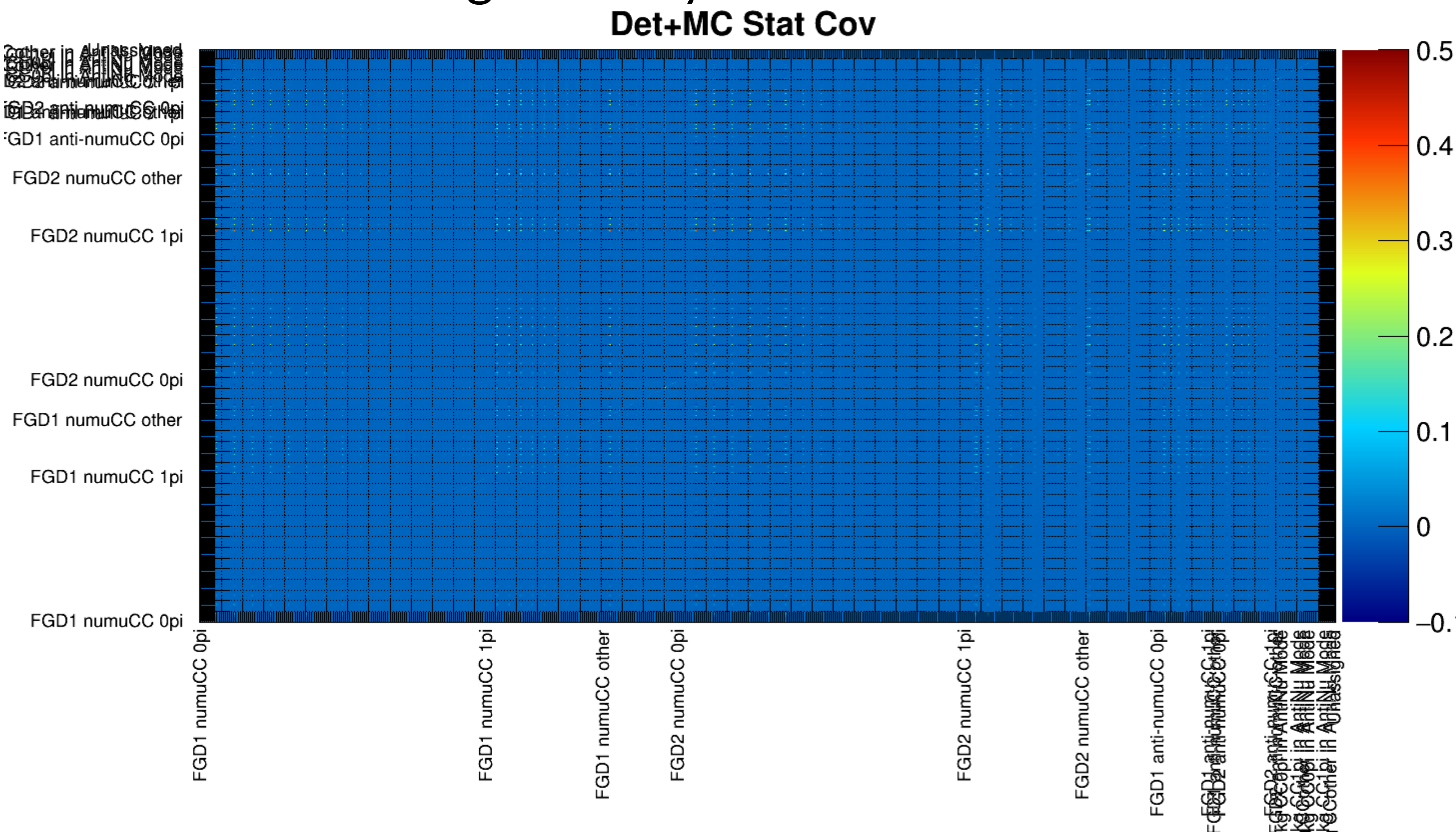
Inspecting the current ND280 cov

- Covariance matrix from detector only



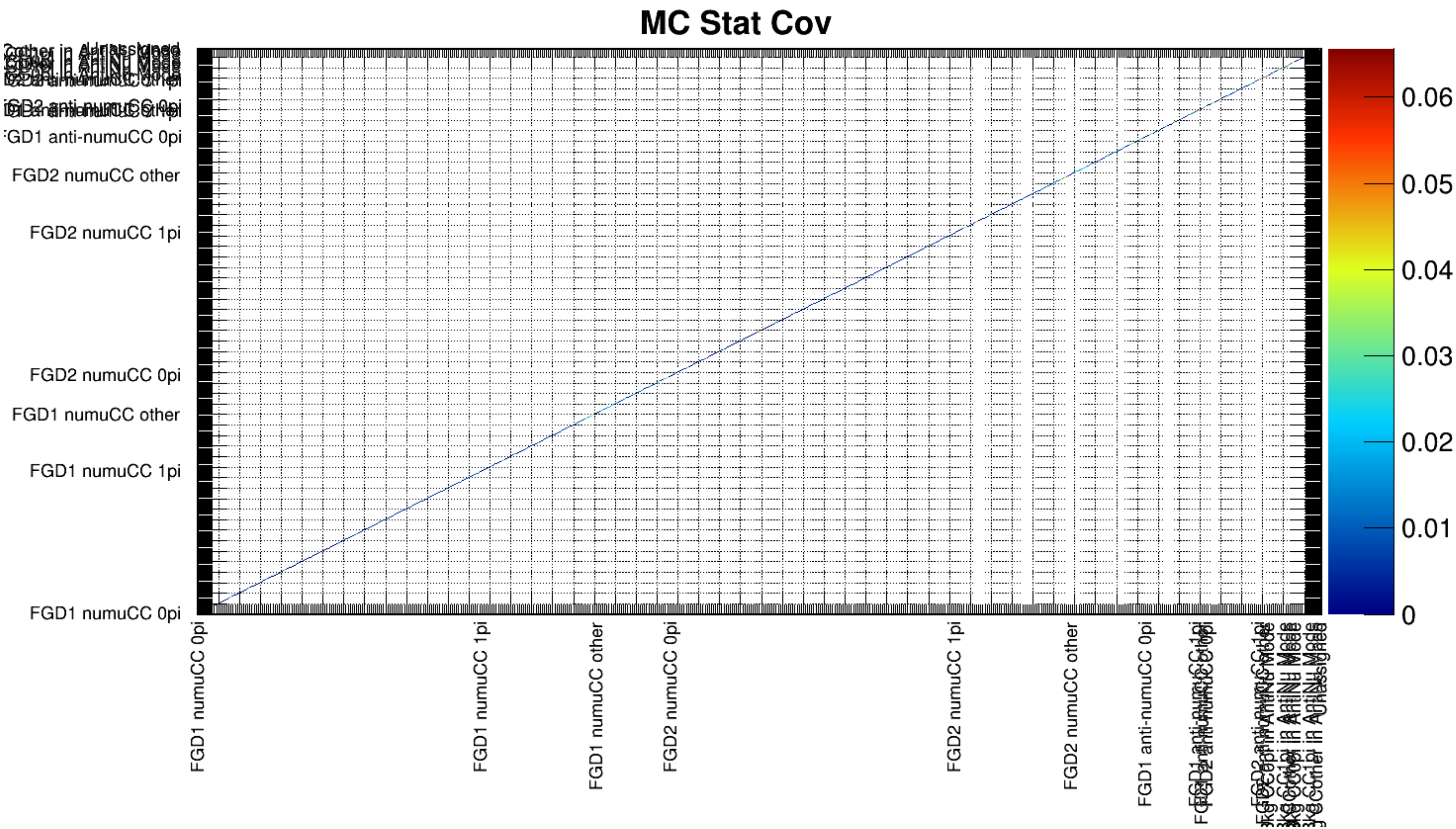
Inspecting the current ND280 cov

- Covariance matrix from det+MC stats only
  - Inflates the diagonal only



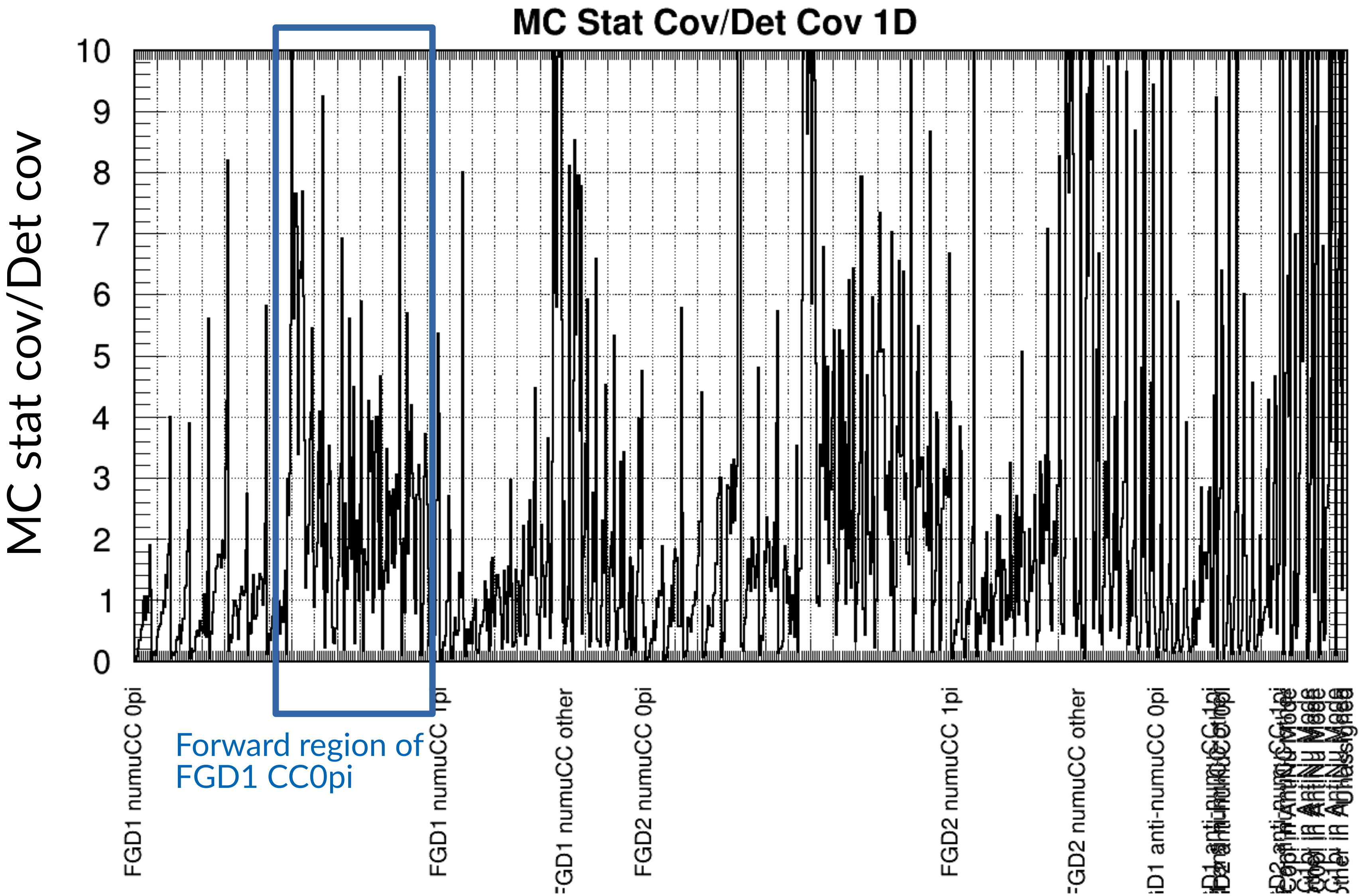
Inspecting the current ND280 cov

- Inspect MC stat covariance: diagonal as expected



Inspecting the current ND280 cov

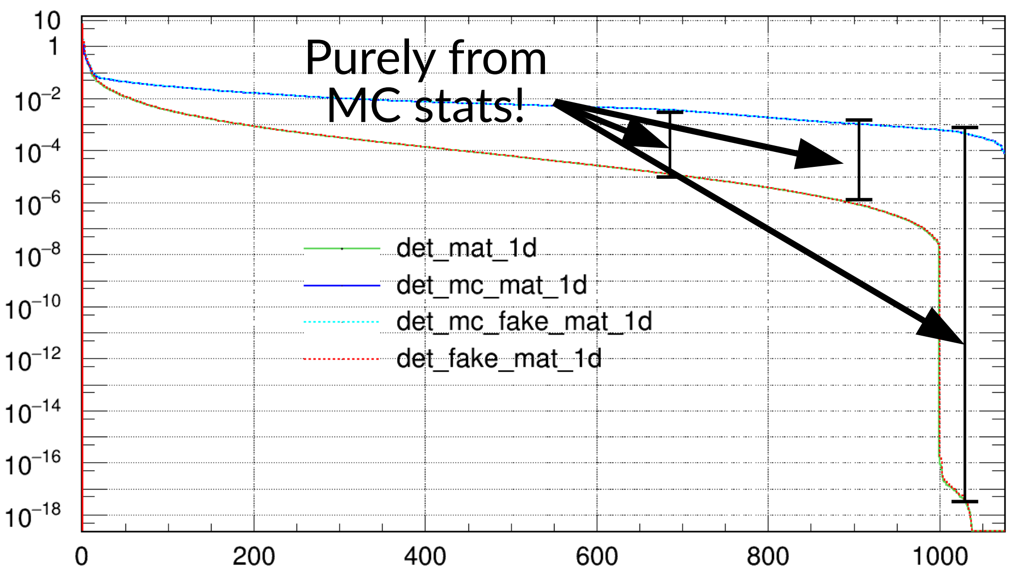
- Compare size of MC stat to detector systematics

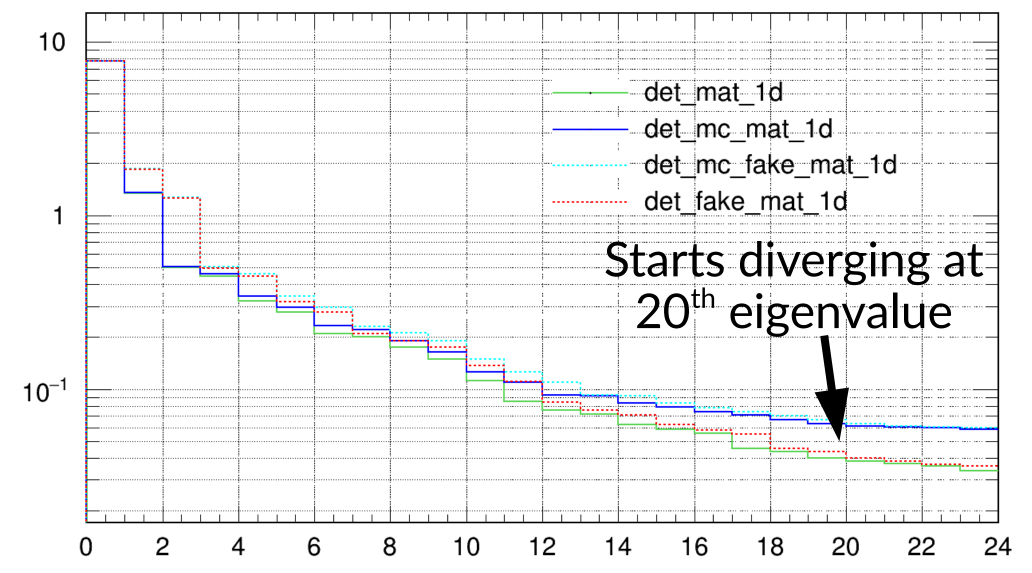


- Same conclusions as previous slides: large in forward region

Inspecting the PCA of covariance

- Look at the size of eigenvectors for
  - Full matrix (ND280+1p1h+MC stats): det_mc_fake_mat_1d
  - Reduced matrix (ND280+MC stats): det_mc_mat_1d
  - Reduced matrix (ND280+1p1h): det_fake_mat_1d
  - ND280 only matrix: det_mat_1d





- Removing MC stats from covariance lessens eigenvalues by O(100) → Seems worthwhile

Introduction

- Need to modify χ^2 to account for MC statistics
- Most straightforward is Pearson with penalty
  - Assumes data distributed around MC as Gaussian

$$\chi^2 = (Data\text{-}MC)^2/(Data)$$

  - Can additionally assume observed unscaled MC (ηn) lays around true unscaled MC with uncertainty $\Delta MC = \text{sqrt}(MC)$

$$\chi^2 = (Data\text{-}MC)^2/(Data+\eta n)$$

- Or use a Poisson for data and MC

$$\chi^2 = 2(MC\text{-}Data+Data\ \ln(Data/MC))$$

Introduction

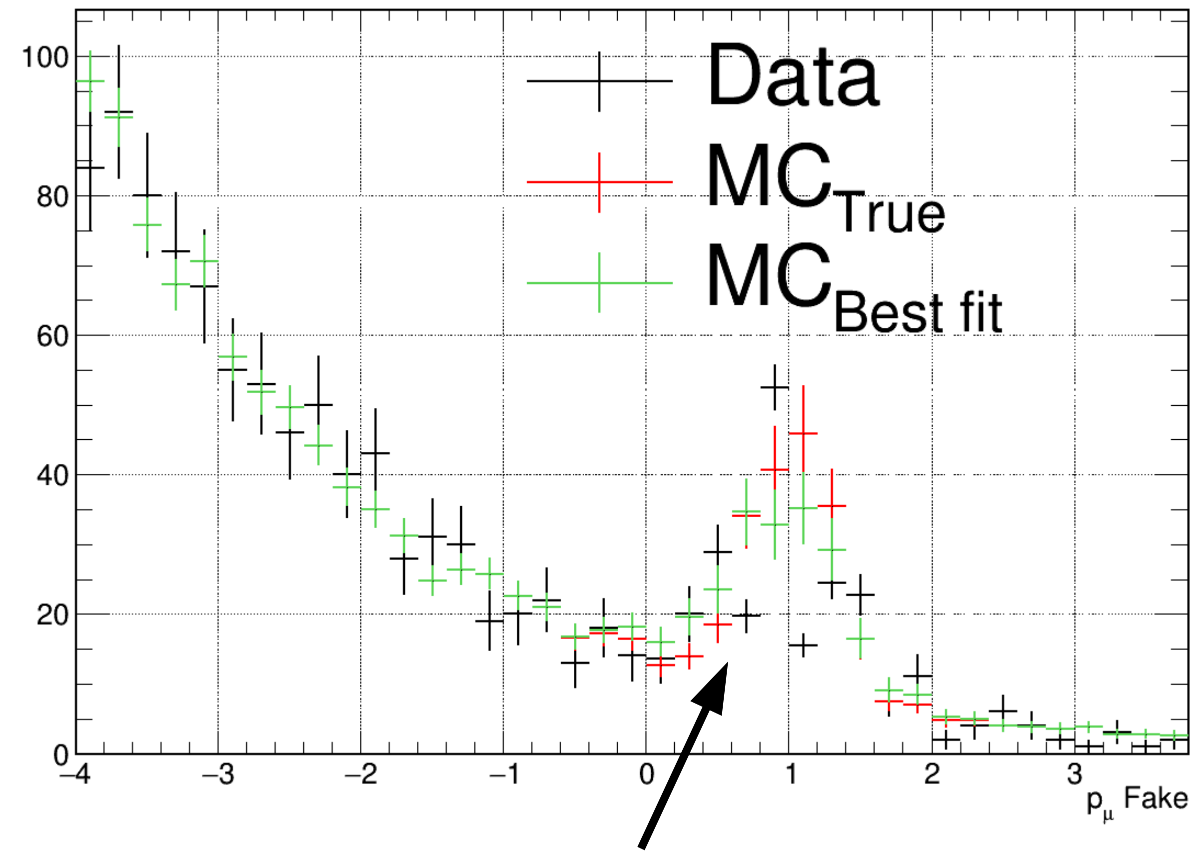
- Neither method estimates true generated MC value
- Barlow-Beeston tries to do this: MC depends on systematic parameters and MC scaling parameters

$$MC^i_{true} = \lambda^i MC^i_{observed}$$

- Add this uncertainty into the likelihood contribution, modifying λ^i as to maximise the likelihood, taking updated MC^i_{true} into account in Poisson likelihood
- Easiest assumption is λ^i are Gaussian distributed around 1
  - Gives an analytic solution for every bin, solvable for every iteration of a fit → No new parameters
- Full Barlow-Beeston requires introducing nbins parameters!
- IceCube authors have an additional likelihood, designed to do something similar, which I don't yet understand...

Testing the likelihoods

- Devised some testing scenarios using different likelihoods
- Make up some event distribution, randomise it, fit it!
- Here an exponentially falling distribution with Gaussian



Fit a Gaussian's central value and width to match MC to data
MC_{True} is MC reweighted with the true values of Gaussian

Rinse, repeat study for 500 toys, with different data statistics and MC/data ratios

Testing the likelihoods

- 100 data events, MC scaling of 2.0 (1.0 didn't converge)
- Surprisingly, Pearson does best by far
  - 30% fit failure but when succeeds has decent coverage
- But essentially uninteresting for us (ND280 doesn't have low stats and low MC stats)

Poisson
μ=4/500 (0.8%)
σ=11/500 (2.2%)
n_{bad}=10/500 (2.0%)

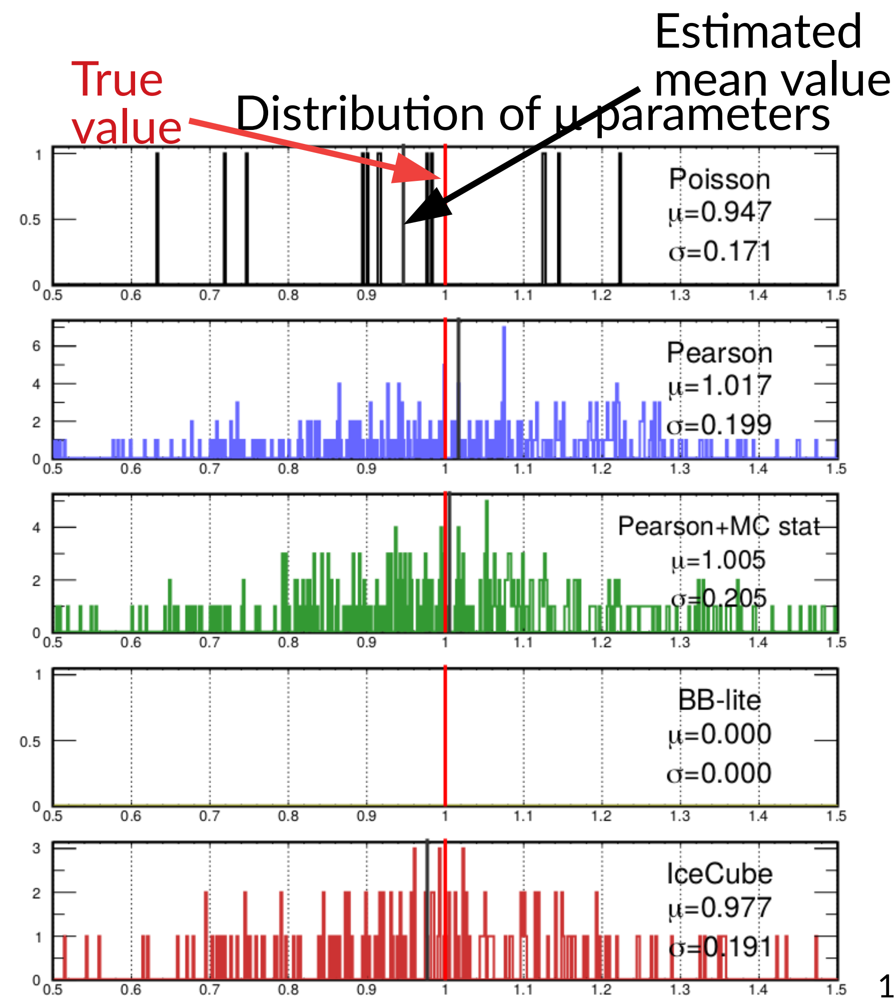
Pearson
μ=202/500 (40.4%)
σ=257/500 (51.4%)
n_{bad}=51/500 (10.2%)

Pearson+MC stat
μ=246/500 (49.2%)
σ=257/500 (51.4%)
n_{bad}=152/500 (30.4%)

BB-lite
μ=0/500 (0.0%)
σ=0/500 (0.0%)
n_{bad}=0/500 (0.0%)

IceCube
μ=106/500 (21.2%)
σ=156/500 (31.2%)
n_{bad}=63/500 (12.6%)

Accounts for MC stat in different ways



Testing the likelihoods

- 100 data events, MC scaling of 10.0
- Increasing MC statistics helps, but still pretty poor performance throughout
  - Need more data events to constrain two parameters

Poisson

μ=125/500 (25.0%)

σ=205/500 (41.0%)

n_{bad}=93/500 (18.6%)

Pearson

μ=235/500 (47.0%)

σ=235/500 (47.0%)

n_{bad}=95/500 (19.0%)

Pearson+MC stat

μ=175/500 (35.0%)

σ=171/500 (34.2%)

n_{bad}=220/500 (44.0%)

BB-lite

μ=82/500 (16.4%)

σ=125/500 (25.0%)

n_{bad}=51/500 (10.2%)

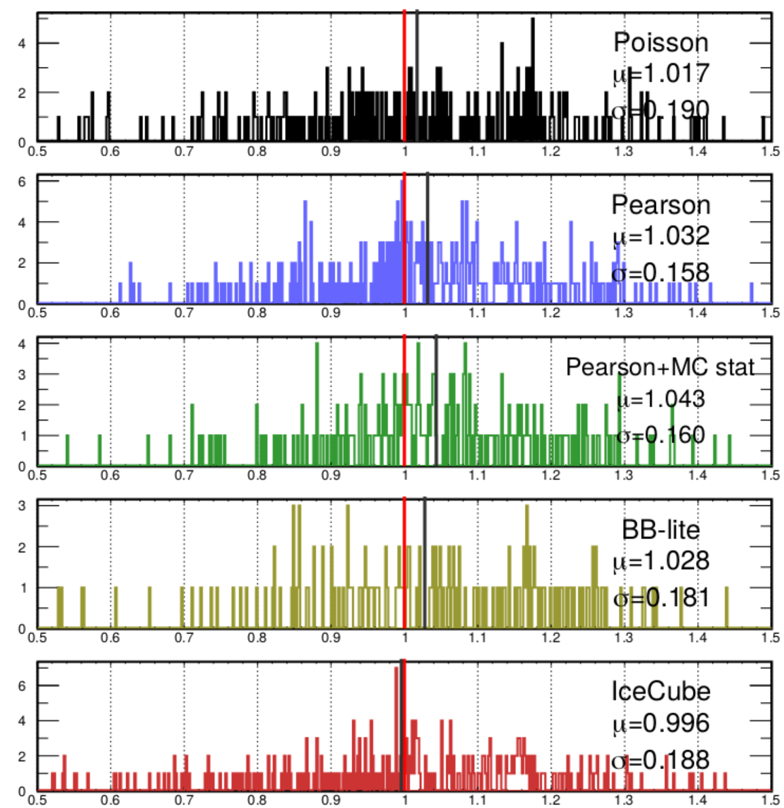
IceCube

μ=155/500 (31.0%)

σ=246/500 (49.2%)

n_{bad}=99/500 (19.8%)

Distribution of μ parameters



Testing the likelihoods

- 500 data events, MC scaling of 1.0
- Pearson+MCStat and IceCube perform well
- Need more for the other methods, including BB

Poisson
μ=23/500 (4.6%)
σ=22/500 (4.4%)
n_{bad}=12/500 (2.4%)

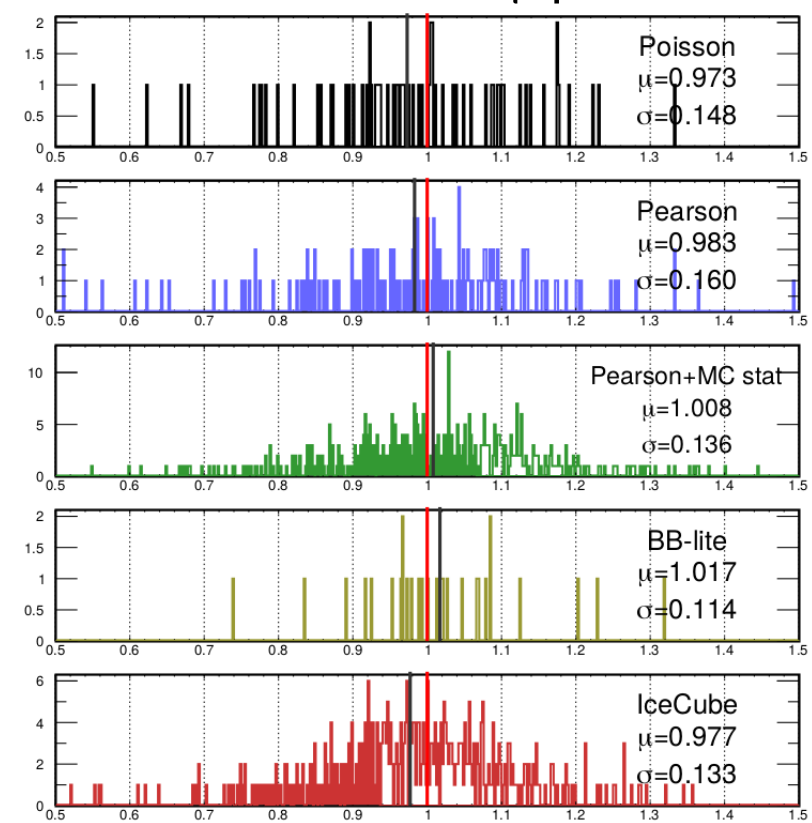
Pearson
μ=78/500 (15.6%)
σ=84/500 (16.8%)
n_{bad}=20/500 (4.0%)

Pearson+MC stat
μ=253/500 (50.6%)
σ=311/500 (62.2%)
n_{bad}=16/500 (3.2%)

BB-lite
μ=19/500 (3.8%)
σ=17/500 (3.4%)
n_{bad}=1/500 (0.2%)

IceCube
μ=230/500 (46.0%)
σ=243/500 (48.6%)
n_{bad}=44/500 (8.8%)

Distribution of μ parameters



Testing the likelihoods

- 500 data events, MC scaling of 10.0
- Barlow-Beeston and IceCube are by far best

Poisson	Pearson
μ=191/500 (38.2%)	μ=121/500 (24.2%)
σ=226/500 (45.2%)	σ=153/500 (30.6%)
n_{bad}=36/500 (7.2%)	n_{bad}=19/500 (3.8%)

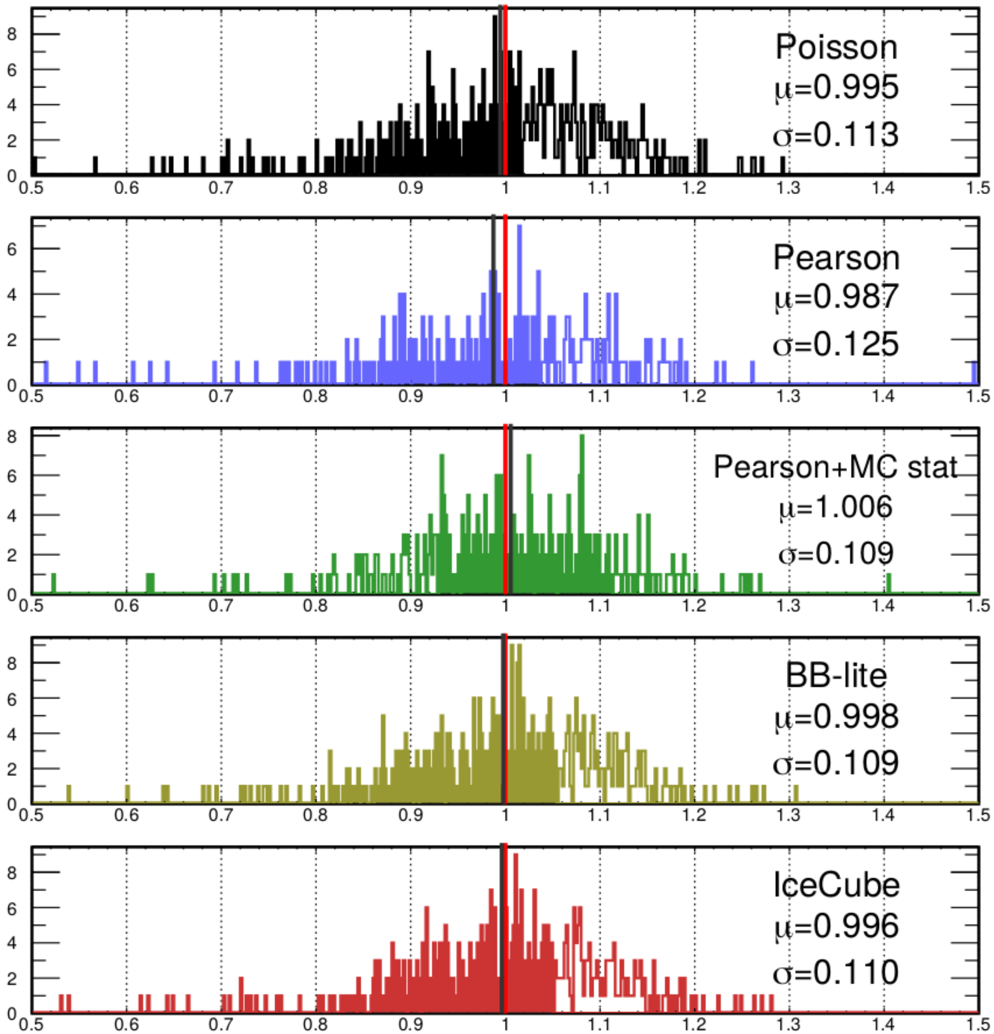
Pearson+MC stat	BB-lite
μ=179/500 (35.8%)	μ=222/500 (44.4%)
σ=196/500 (39.2%)	σ=266/500 (53.2%)
n_{bad}=97/500 (19.4%)	n_{bad}=30/500 (6.0%)

IceCube
μ=217/500 (43.4%)
σ=264/500 (52.8%)
n_{bad}=28/500 (5.6%)

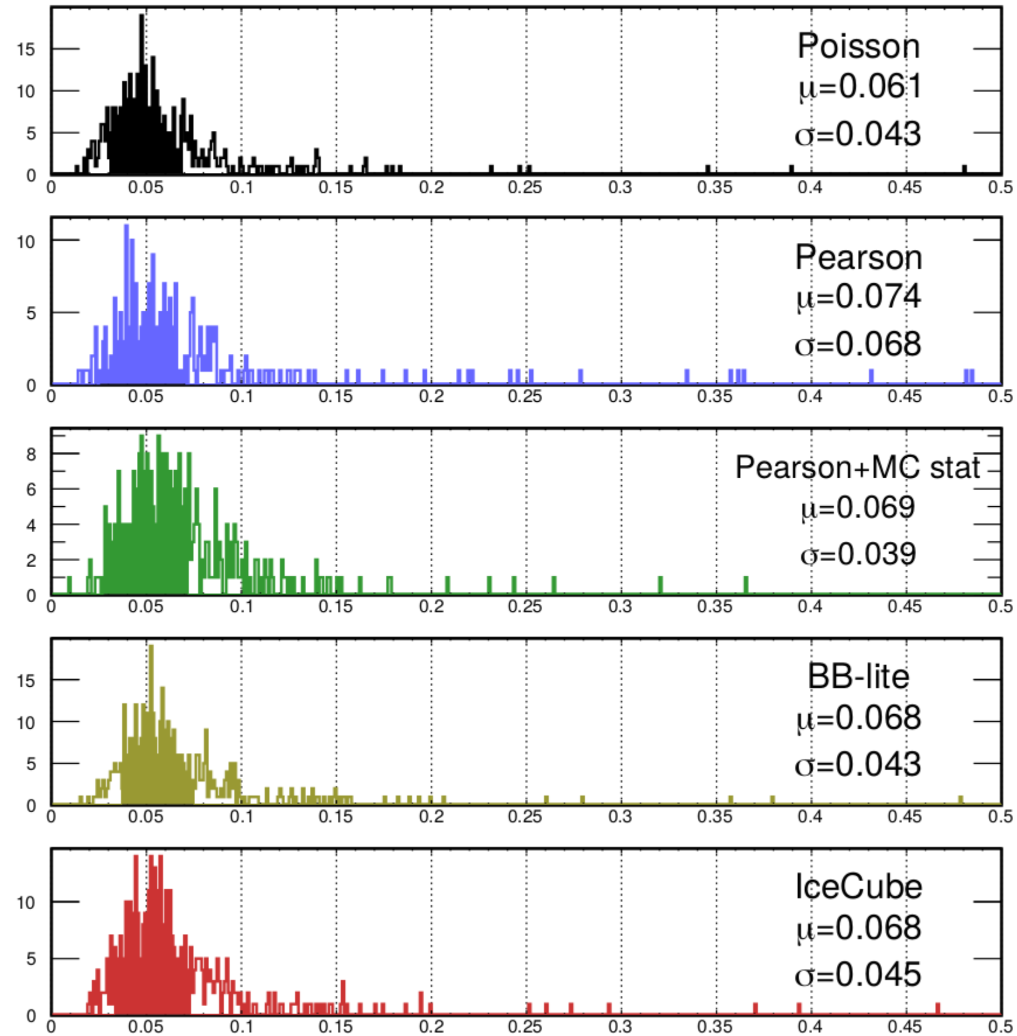
Testing the likelihoods

- 500 data events, MC scaling of 10.0
- Barlow-Beeston and IceCube are by far best

Distribution of μ parameters



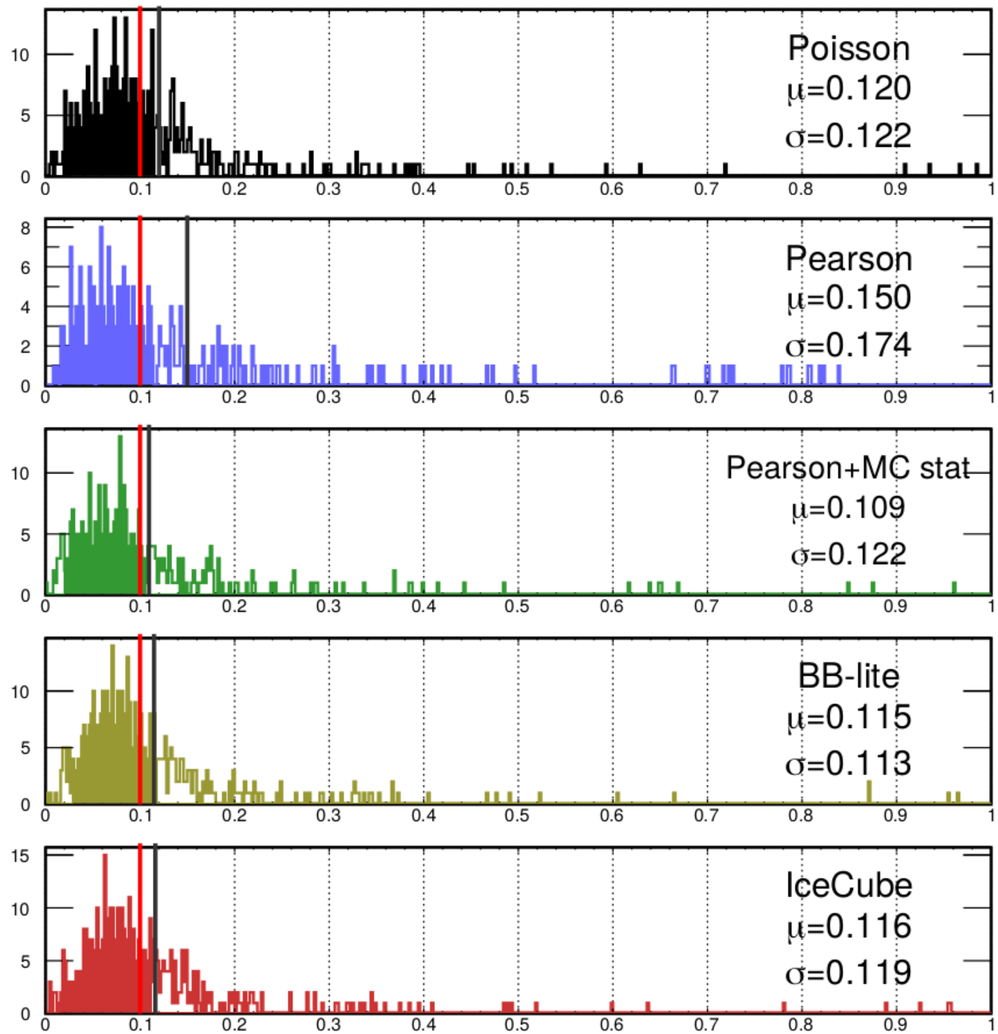
Uncertainty of μ parameters



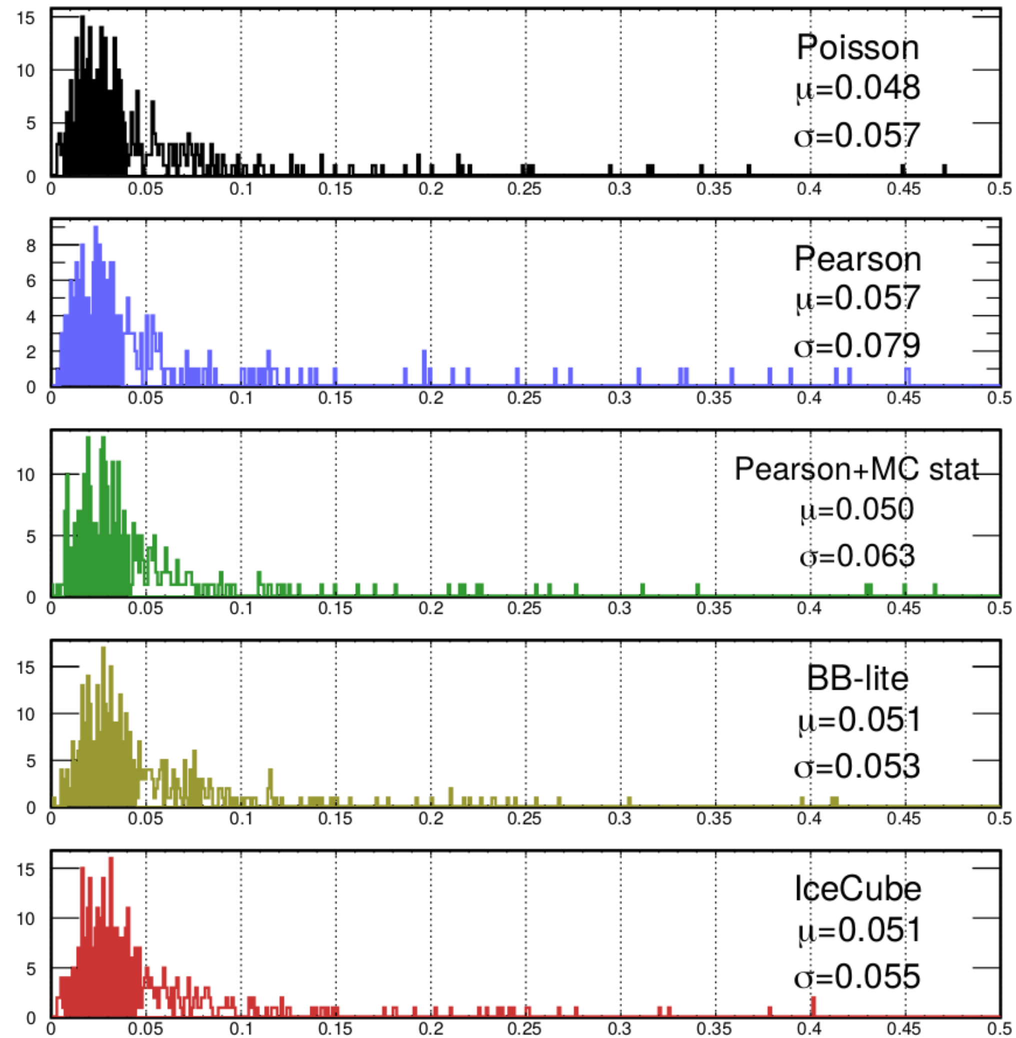
Testing the likelihoods

- 500 data events, MC scaling of 10.0
- Barlow-Beeston and IceCube are by far best

Distribution of σ parameters



Uncertainty of σ parameters



Testing the likelihoods

- 5000 data events, MC scaling of 1.0
- Methods that account for MC stats are similar

Poisson	Pearson
μ=142/500 (28.4%)	μ=122/500 (24.4%)
σ=147/500 (29.4%)	σ=144/500 (28.8%)
n_{bad}=28/500 (5.6%)	n_{bad}=21/500 (4.2%)

Pearson+MC stat	BB-lite
μ=269/500 (53.8%)	μ=263/500 (52.6%)
σ=322/500 (64.4%)	σ=313/500 (62.6%)
n_{bad}=1/500 (0.2%)	n_{bad}=16/500 (3.2%)

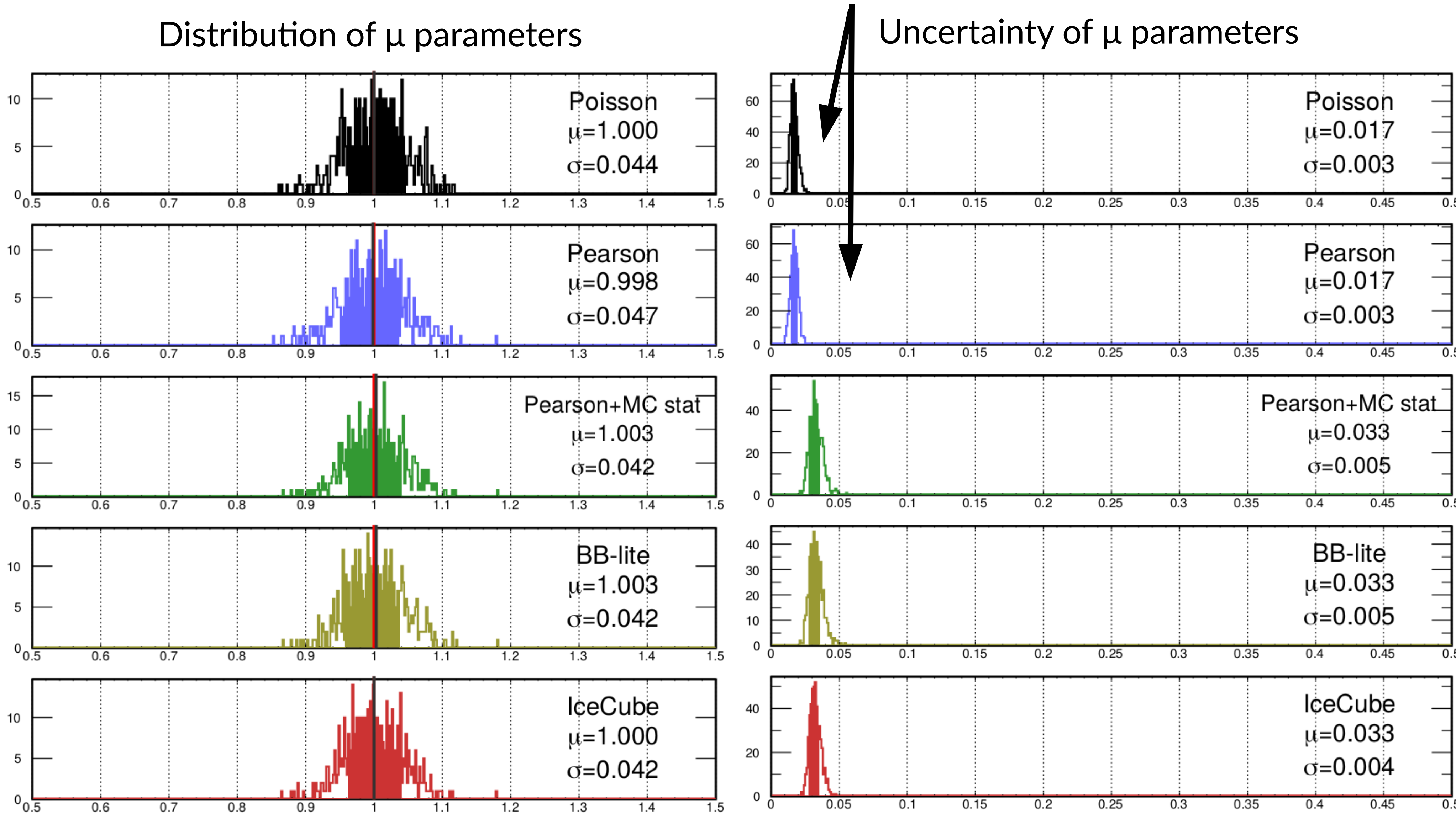
IceCube
μ=265/500 (53.0%)
σ=303/500 (60.6%)
n_{bad}=13/500 (2.6%)

Testing the likelihoods

- 5000 data events, MC scaling of 1.0
- Without MC stats estimates half the error!

Distribution of μ parameters

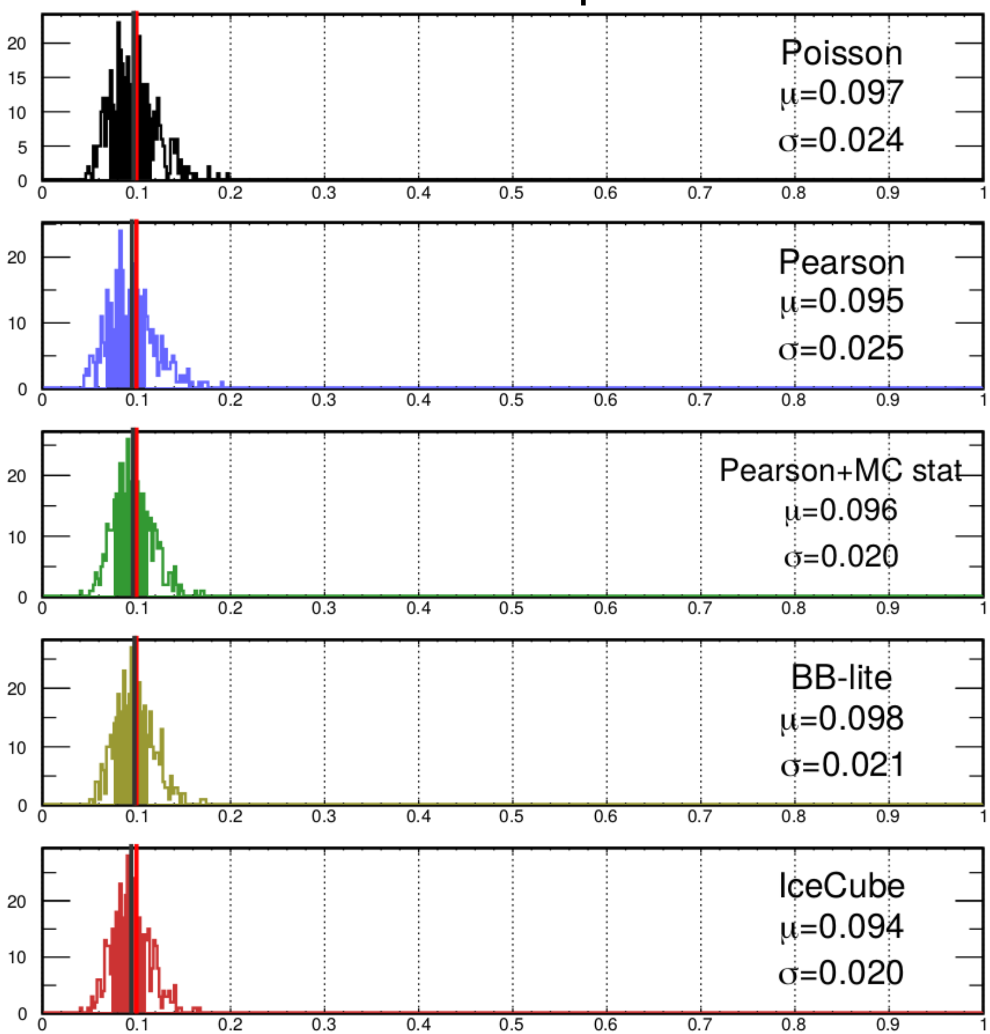
Uncertainty of μ parameters



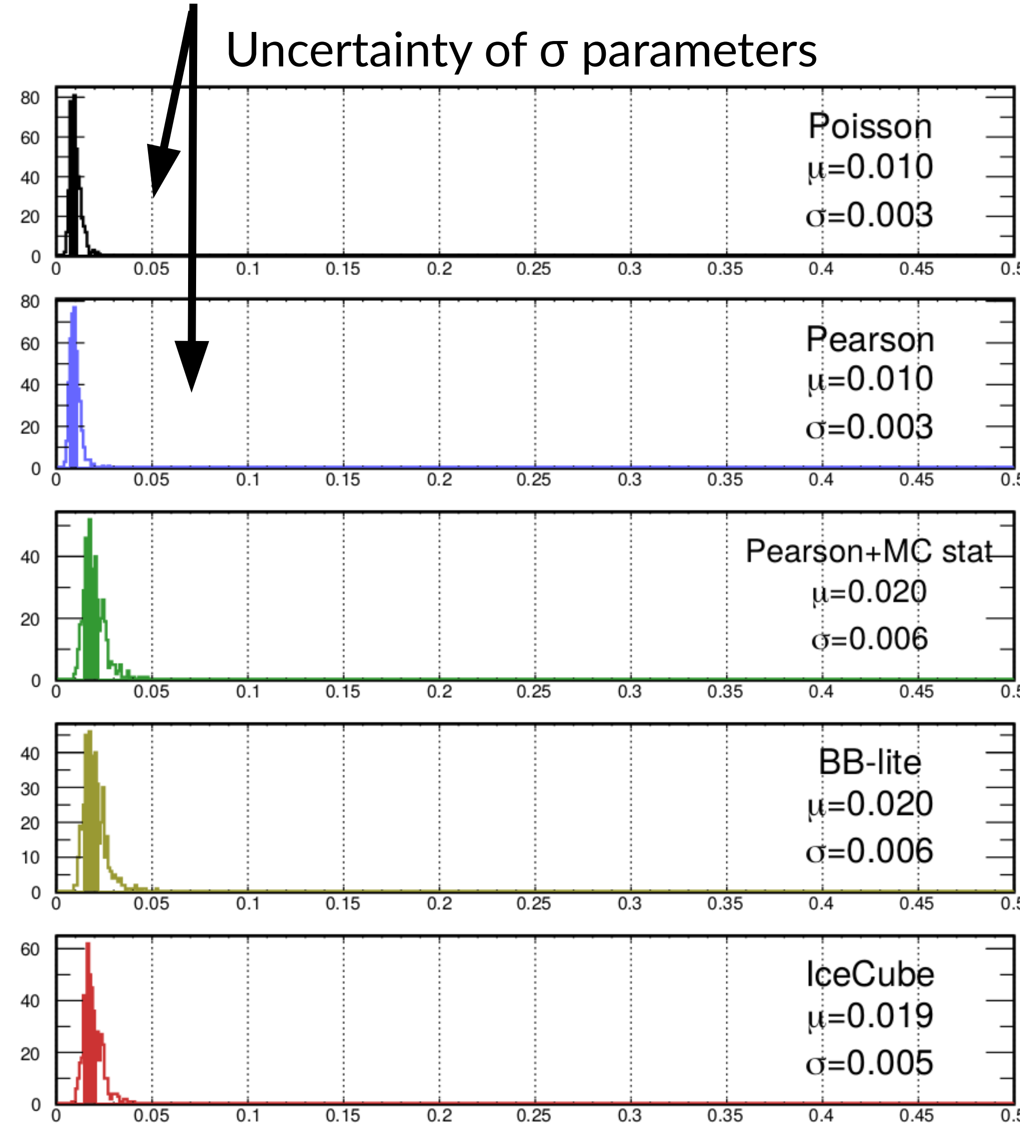
Testing the likelihoods

- 5000 data events, MC scaling of 1.0
- Without MC stats estimates half the error!

Distribution of σ parameters



Uncertainty of σ parameters



Testing the likelihoods

- 5000 data events, MC scaling of 10.0
- Methods that account for MC stats are similar
  - Other methods catching up

Poisson	Pearson
μ=215/500 (43.0%)	μ=183/500 (36.6%)
σ=213/500 (42.6%)	σ=201/500 (40.2%)
n_{bad}=15/500 (3.0%)	n_{bad}=25/500 (5.0%)

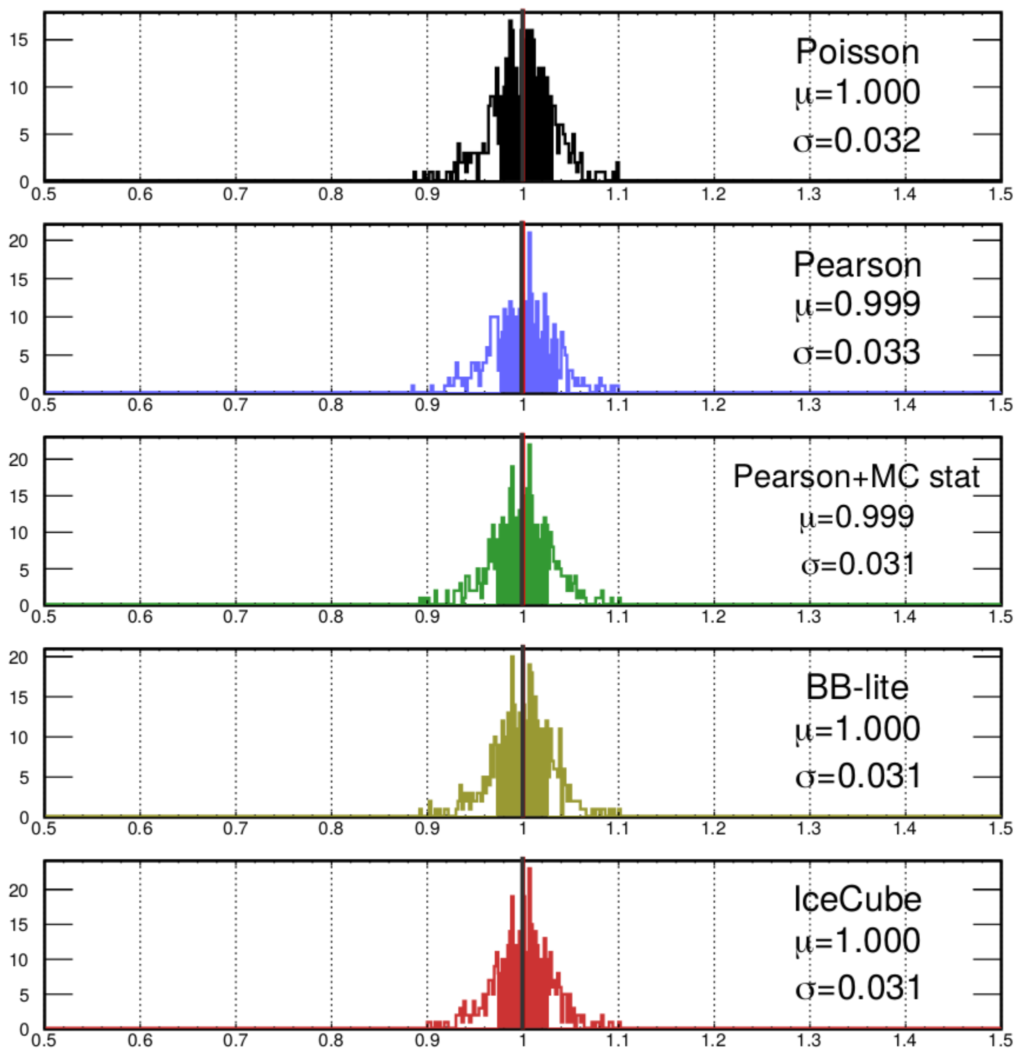
Pearson+MC stat	BB-lite
μ=241/500 (48.2%)	μ=246/500 (49.2%)
σ=253/500 (50.6%)	σ=252/500 (50.4%)
n_{bad}=1/500 (0.2%)	n_{bad}=10/500 (2.0%)

IceCube
μ=244/500 (48.8%)
σ=249/500 (49.8%)
n_{bad}=16/500 (3.2%)

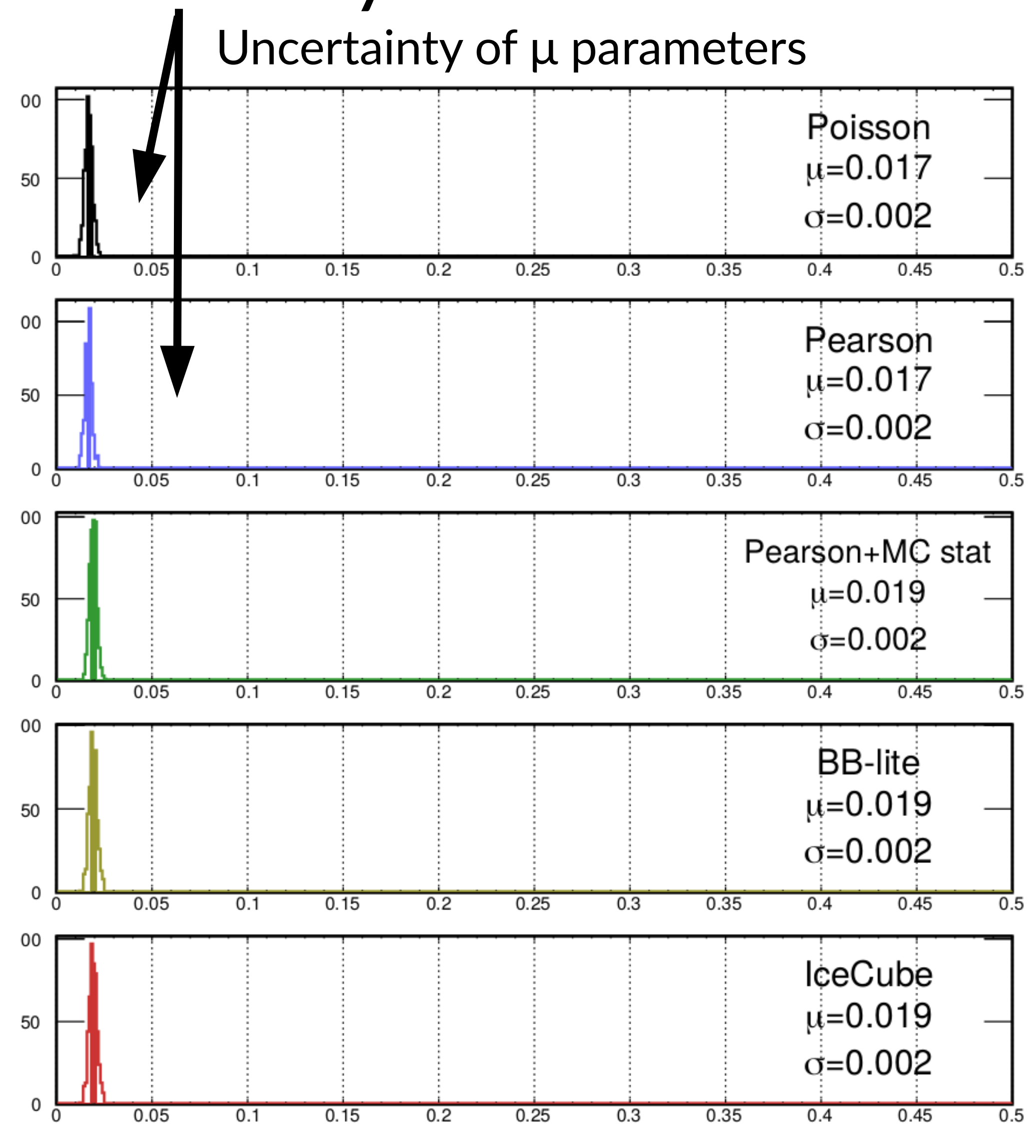
Testing the likelihoods

- 5000 data events, MC scaling of 10.0
- Only small difference in uncertainty now

Distribution of μ parameters



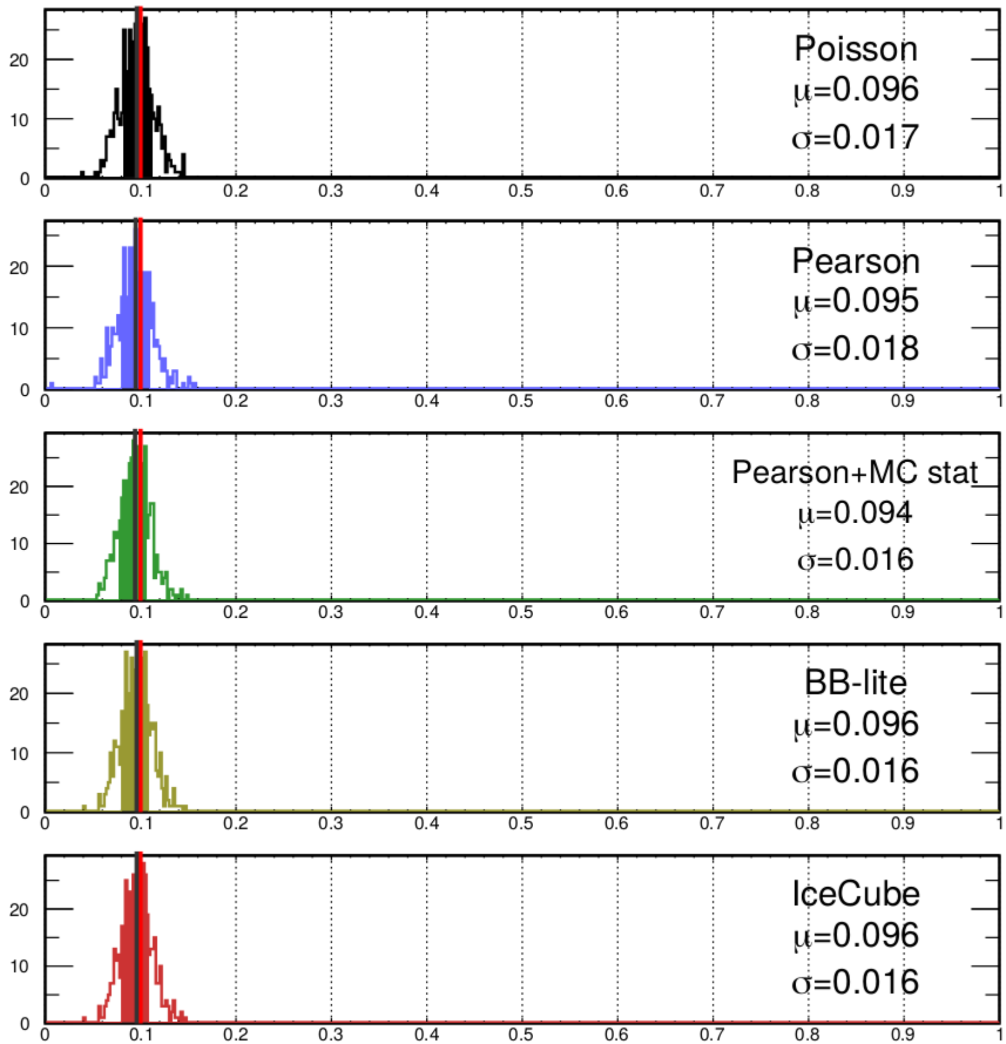
Uncertainty of μ parameters



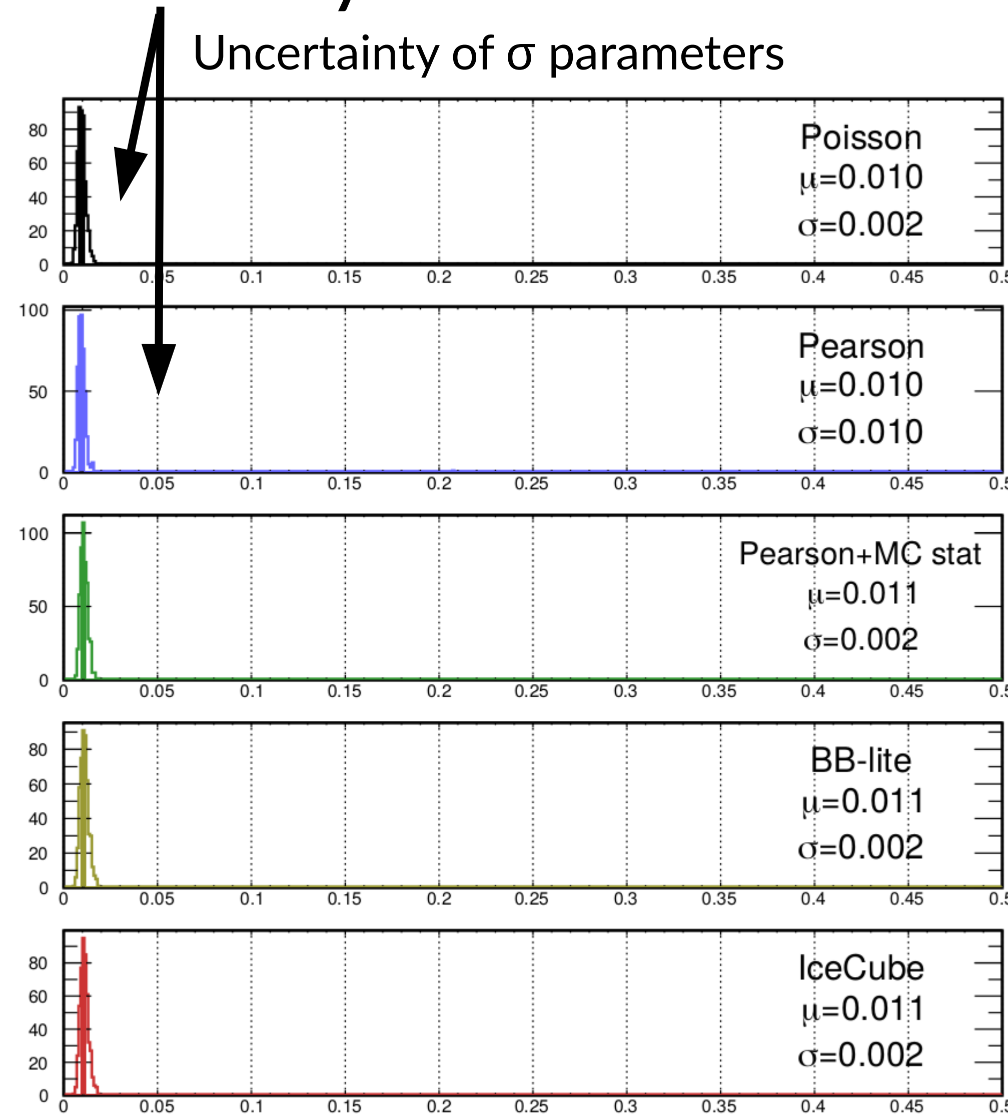
Testing the likelihoods

- 5000 data events, MC scaling of 10.0
- Only small difference in uncertainty now

Distribution of σ parameters



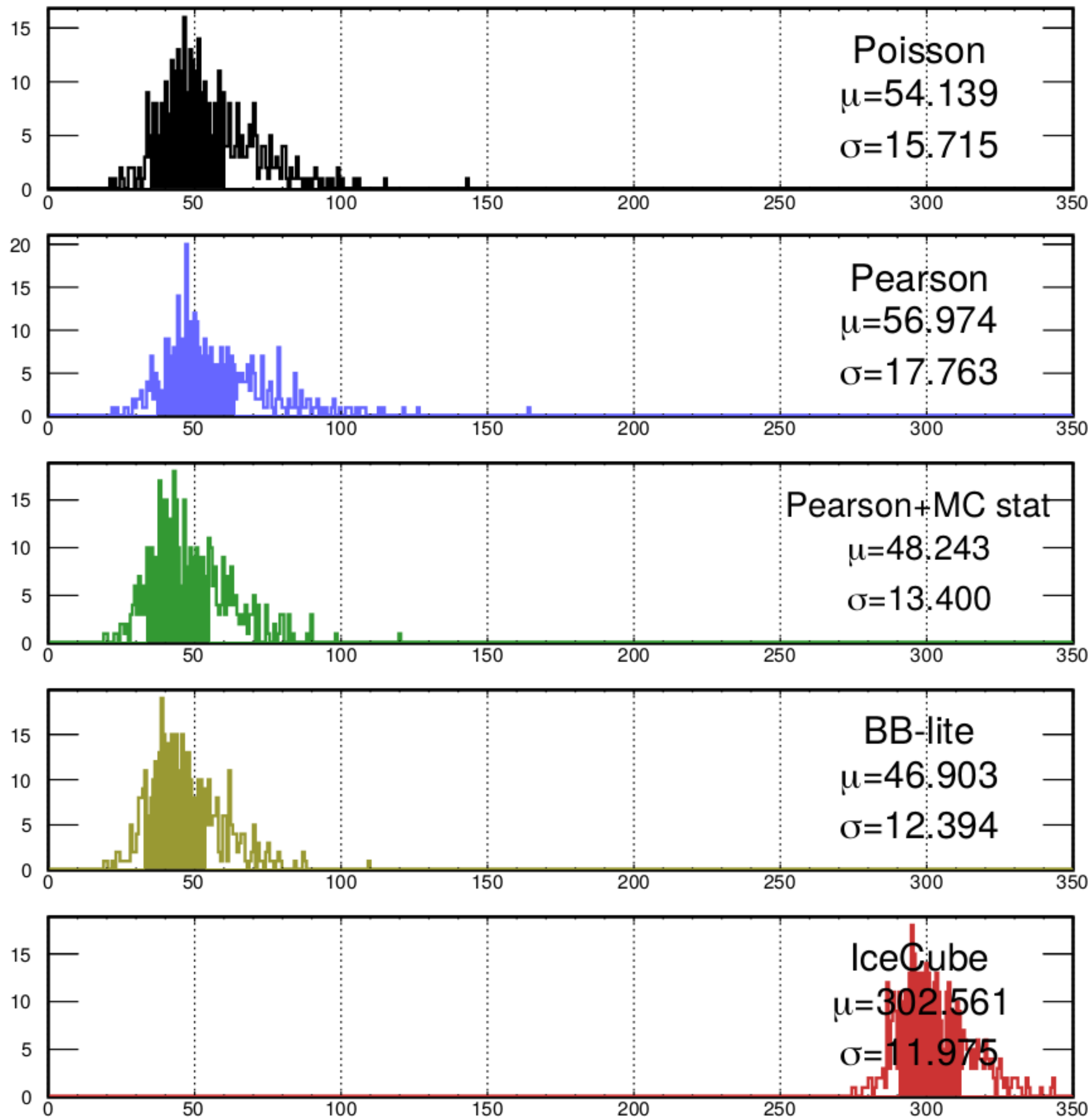
Uncertainty of σ parameters



Testing the likelihoods

- 5000 data events, MC scaling of 10.0
- Also look at the likelihood distributions

Distribution of best-fit -2lnL



Poisson and Pearson similar: neither take MC stats into account

Poisson and Pearson with MC stat accounted for (in different ways): similarly lower -2lnL

IceCube -2lnL always much higher than the other methods: ???

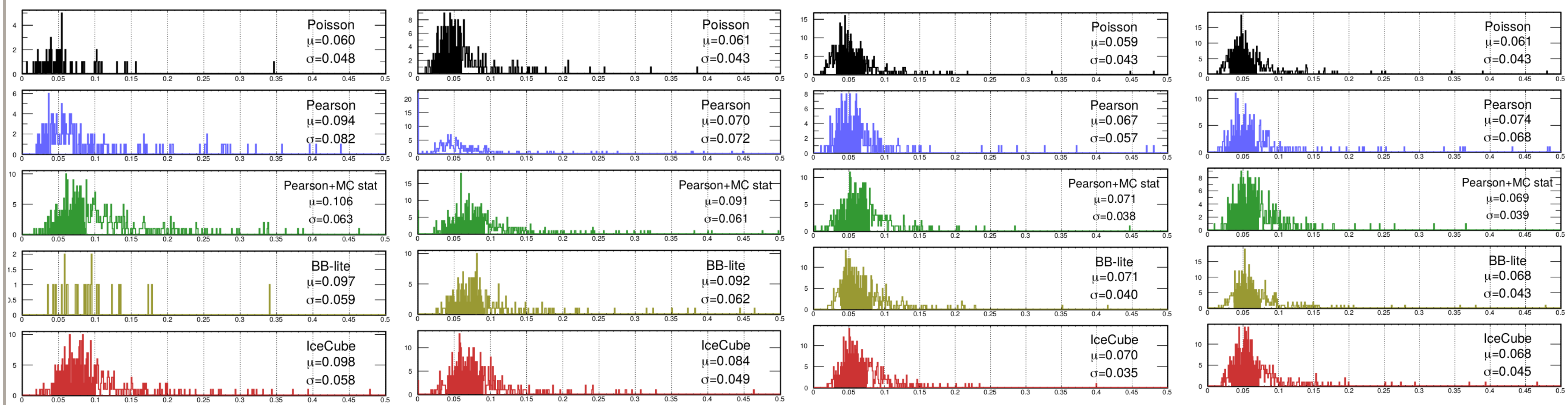
UNIVERSITY of ROCHESTER

Testing the likelihoods

- Check how uncertainty develops with scaling

500 data events

1.0 scaling | 2.0 scaling | 5.0 scaling | 10.0 scaling



- Poisson estimates are constant, Pearson wildly changing → Both are unreliable
- MC statistics methods have similar estimates
- MC statistics methods' estimate decrease with more generated MC

Summary

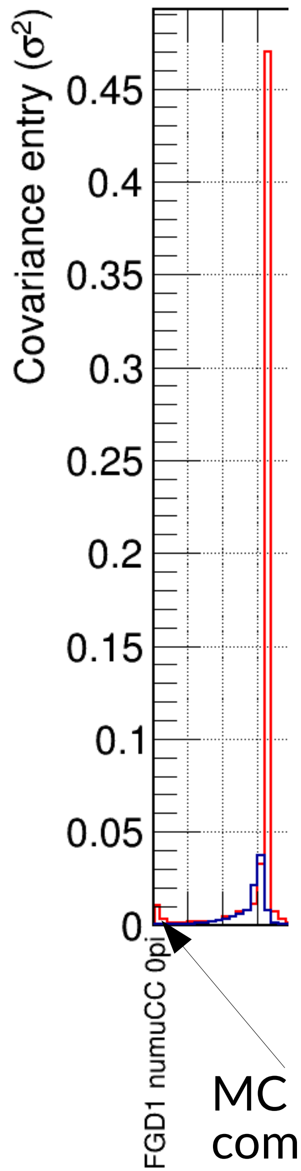
- Trying to reduce number of ND parameters in T2K ND fit
  - Better PCA, better merging, more stable fits/faster convergence
- Accounting for MC stats is central in producing unbiased intervals
  - **Can bias 1σ intervals by 100% in this test case**
- IceCube likelihood seems best, but strangely large χ^2
  - Collaborators can reproduce this in his cross-section analysis, and authors say this is intended?! Not sure about this...
- BB-lite struggles in ultra-low MC stats: likely due to Gaussian assumption in scaling parameters
  - Seems to fail to converge instead of giving unreliable estimate
- Pearson+MC stats struggles at low MC stats
- My preference from these studies:
  - IceCube > BB-lite > Pearson+MC stats
- Caveat: methods introduce changing MC estimate → discontinuous likelihoods? → Minuit problems?

Thanks!

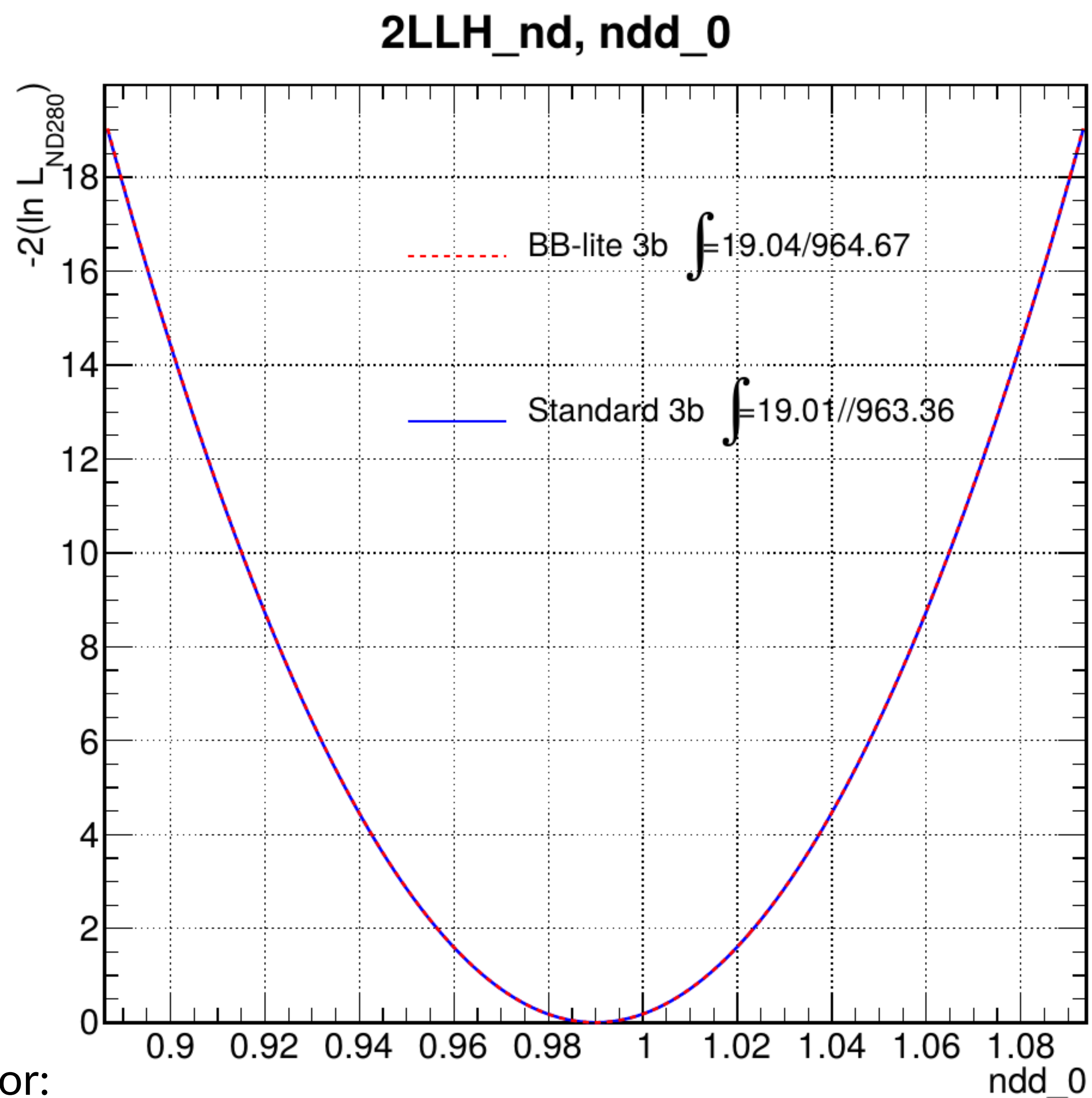
T2K studies

- Used reduced # of runs with/without BB-lite, with corresponding covariance matrix
  - Study should be completed with full MC statistics
- Comparing LLH scans: response to prior should be the same
- However, scan should run in narrower range since it starts at +/- 3 units of 1σ
  - By removing MC stats our 1σ range is smaller
  - Should see effect in regions where MC stat is dominant
- Start with comparing the prior

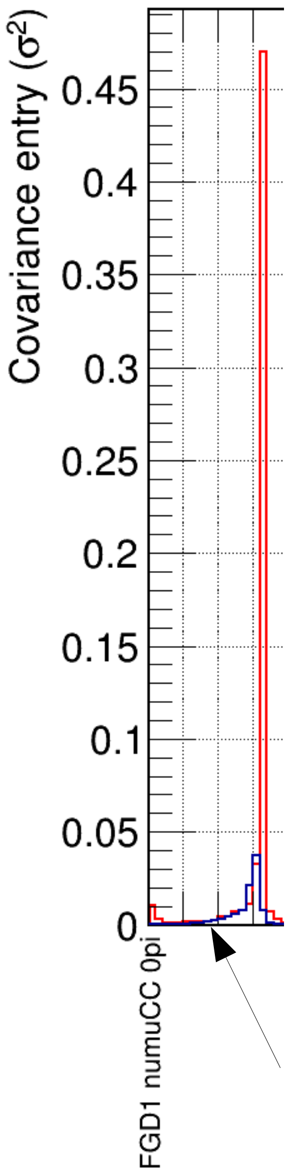
T2K studies



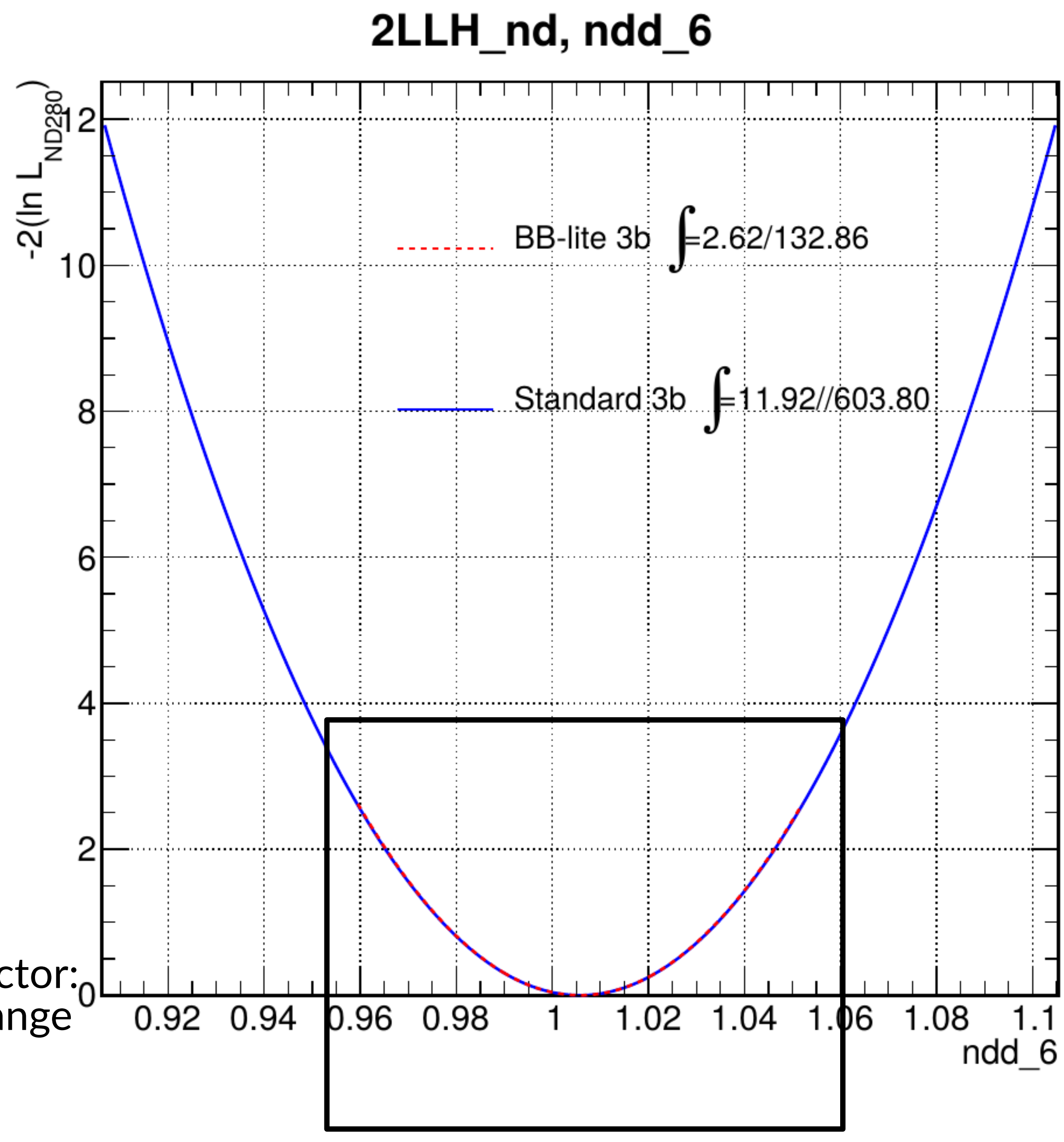
MC stats effect tiny compared to detector: no effect



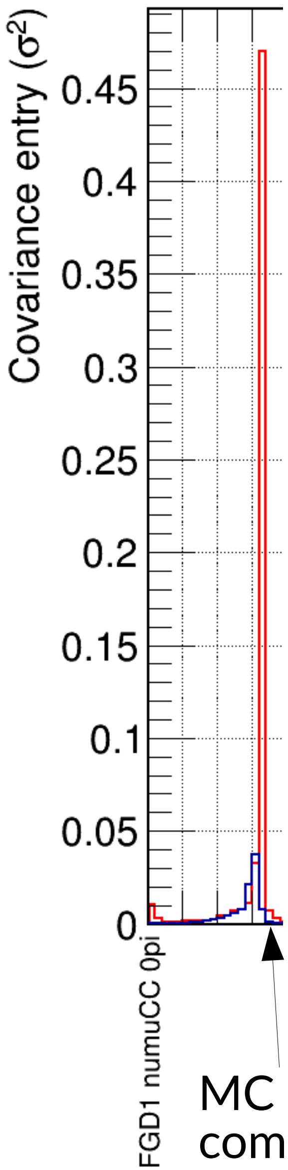
T2K studies



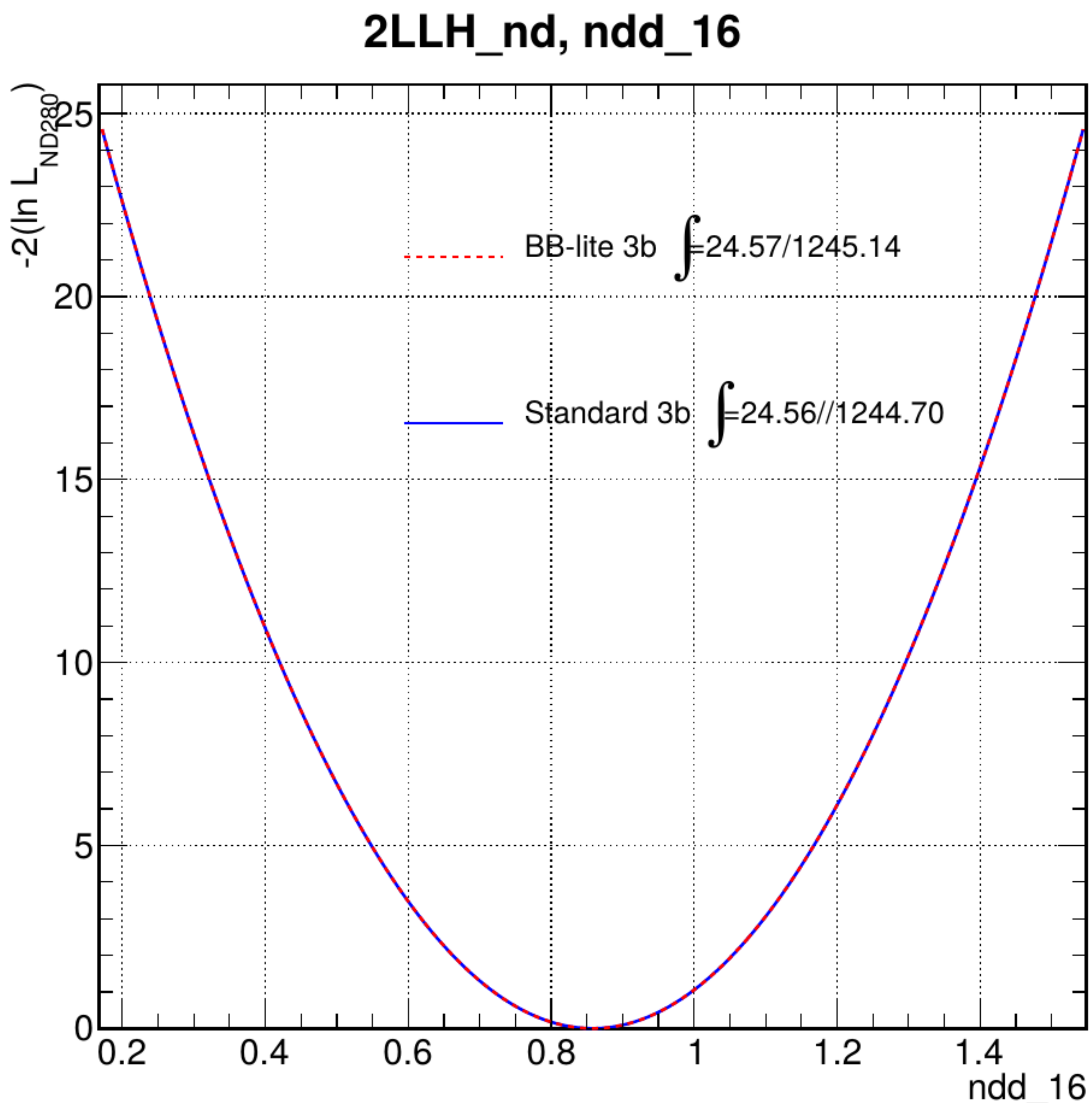
MC stats effect comparable to detector: large reduction in range (direct reflection of smaller 1σ)



T2K studies

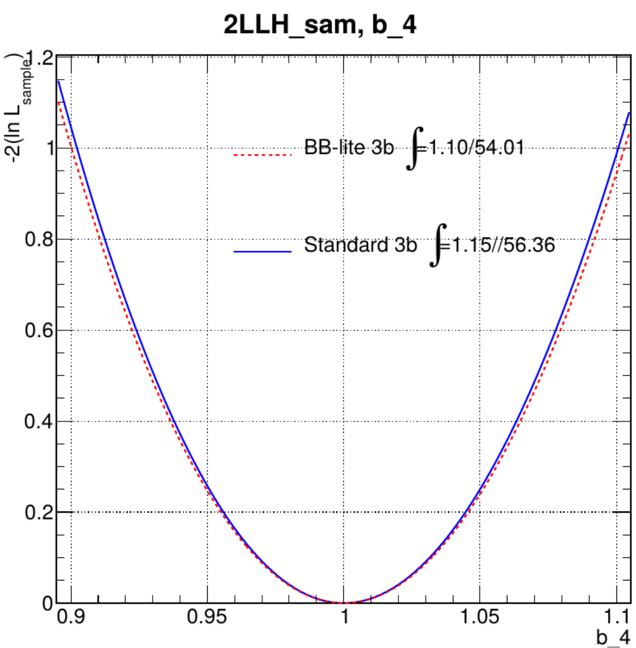


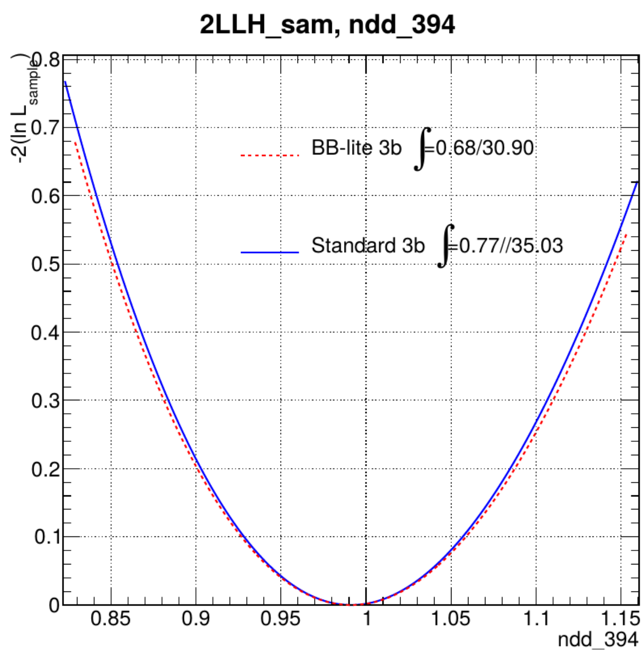
MC stats very small compared to detector: no effect

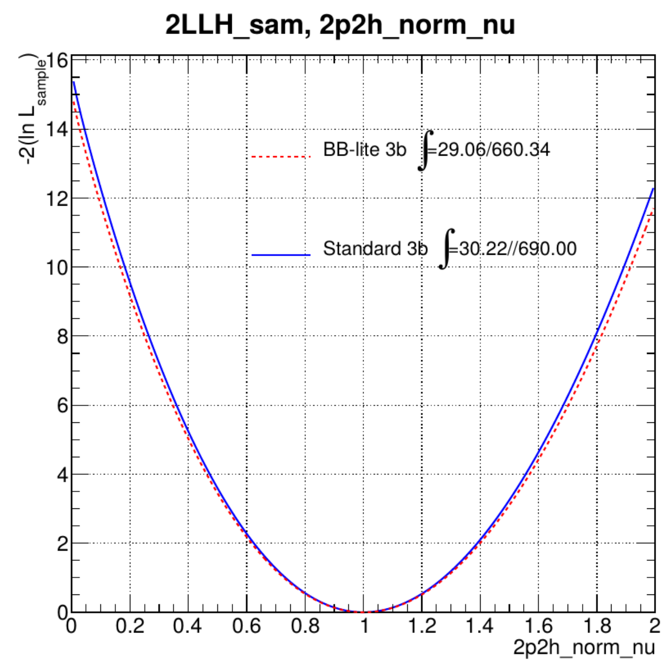


T2K studies

- Now look at the effect on the sample contribution
- Introducing MC statistics into the sample likelihood should lower the overall χ^2







- Seems to be working as intended

Future and Summary

- Repeat studies for full ND280 MC
- Run some Asimov fits
- Figure out why ND280+1p1h Cholesky decomposes but ND280 only doesn't
- Hand over to Will who's producing the covariance matrix for next round
  - BB seems ok for us, but not great in low stat regions
  - Probably because of Gaussian assumption of scaling par.
- Get in touch with IceCube folks about their likelihood: why so large χ^2
  - Looks promising otherwise: much better in ultra low stat and does not require additional parameters
- Maybe OA folks want to port likelihood?