Analysis update: Beam asymmetry in gamma n -> pi- p from the g13b dataset

Daria Sokhan

July 2015

Can the fit-extracted asymmetry be trusted?

Can the fit-extracted asymmetry be trusted?

- Extensive study of the effect different aspects of an asymmetry distribution have on the fit-extracted asymmetry was carried out by N. Zacchariou (CLAS-Note 2012-011): for very large samples (~Million events).

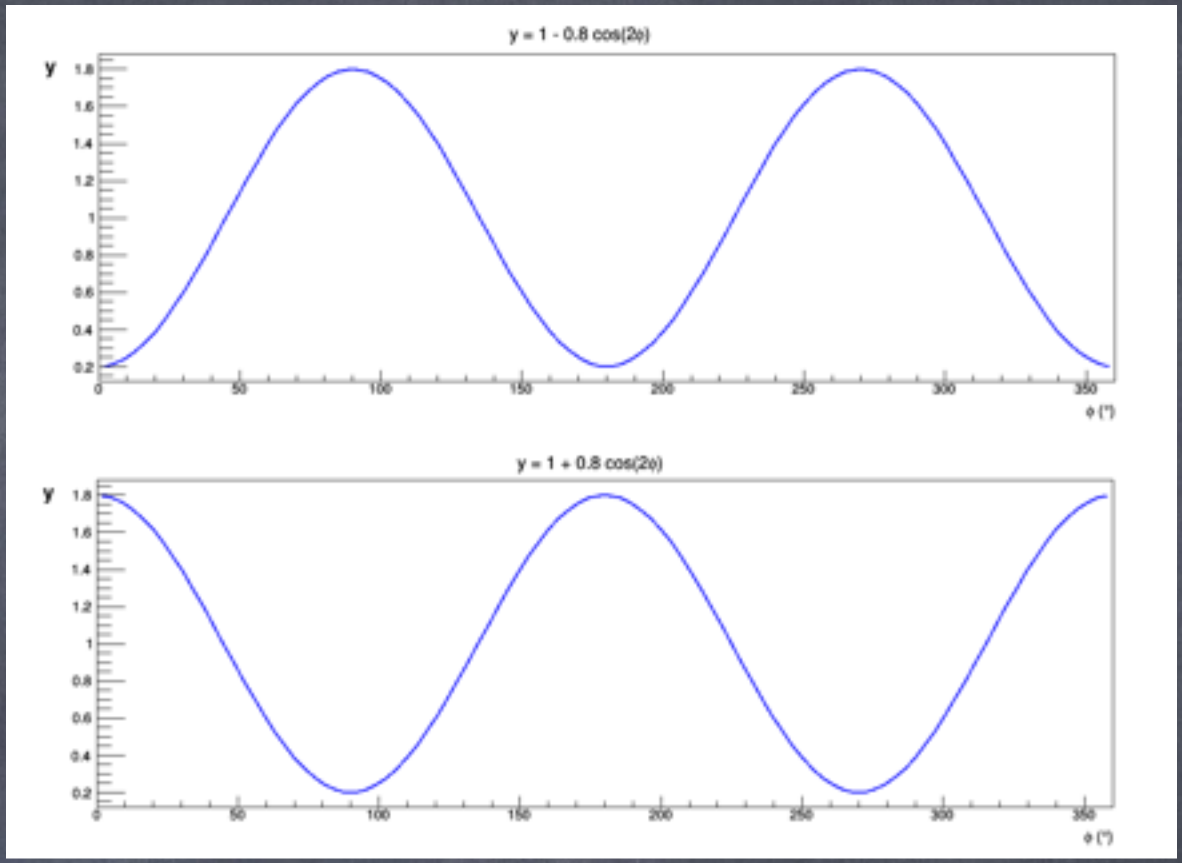
- Found that most aspects have no effect. The width of the phi-bin DOES, but can be corrected for with: $\Sigma_{fit} = \frac{\Delta\phi}{sin(\Delta\phi)}\Sigma_{mean}$

- Are there any systematic effects associated with a smaller statistical sample?

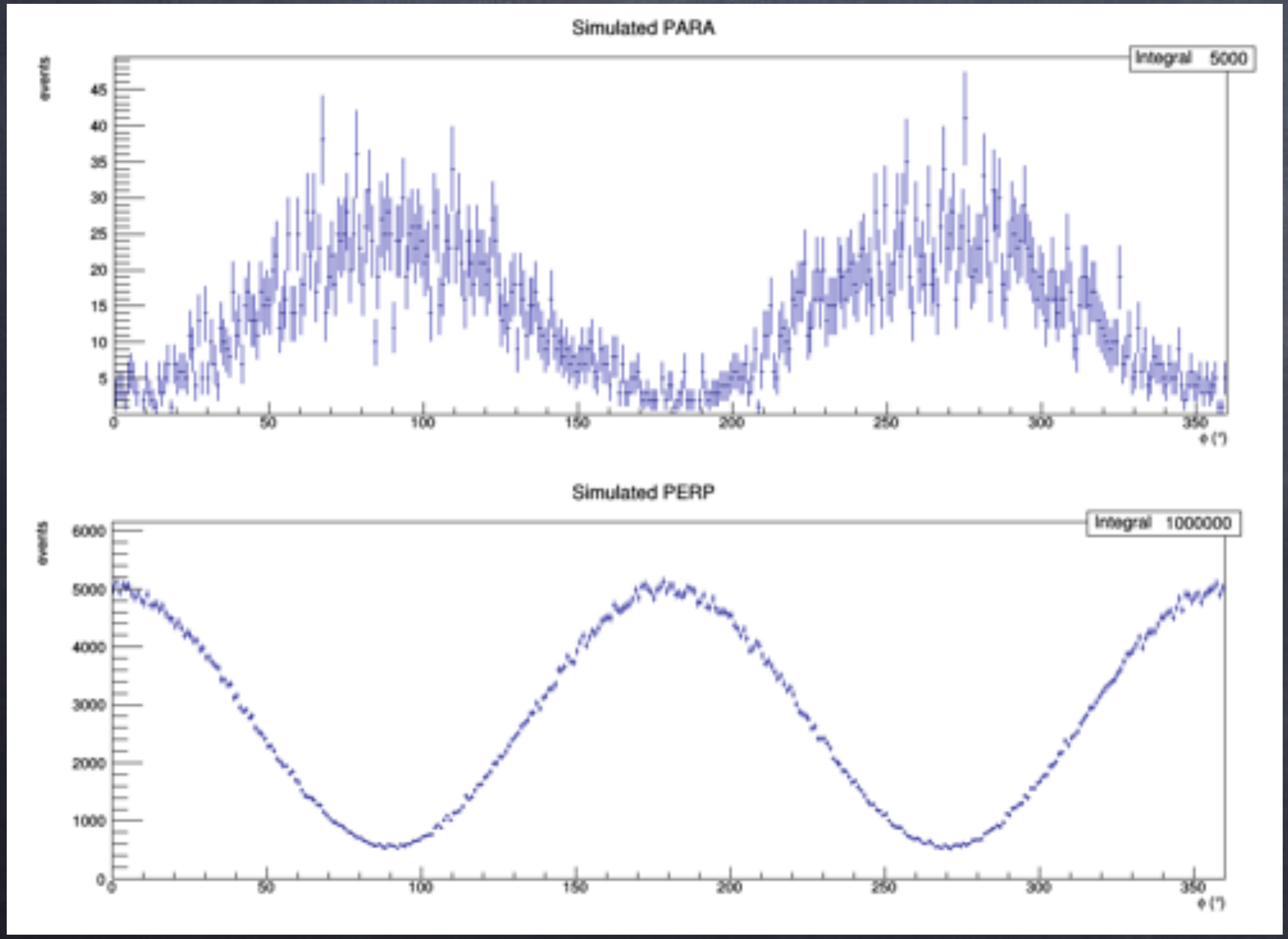
- It turns out that there are. They are typically smaller than the statistical error on the extracted asymmetry, but can reach up to 15 %.

Simulation: technique

- Simulations performed by filling a PARA and PERP distribution randomly according to their perfect distribution functions.



Perfect PARA and PERP distribution functions



Randomly-filled histograms, with 5000 events (PARA) and 1M events (PERP), phi-bin of 1 deg, from the perfect distributions above.

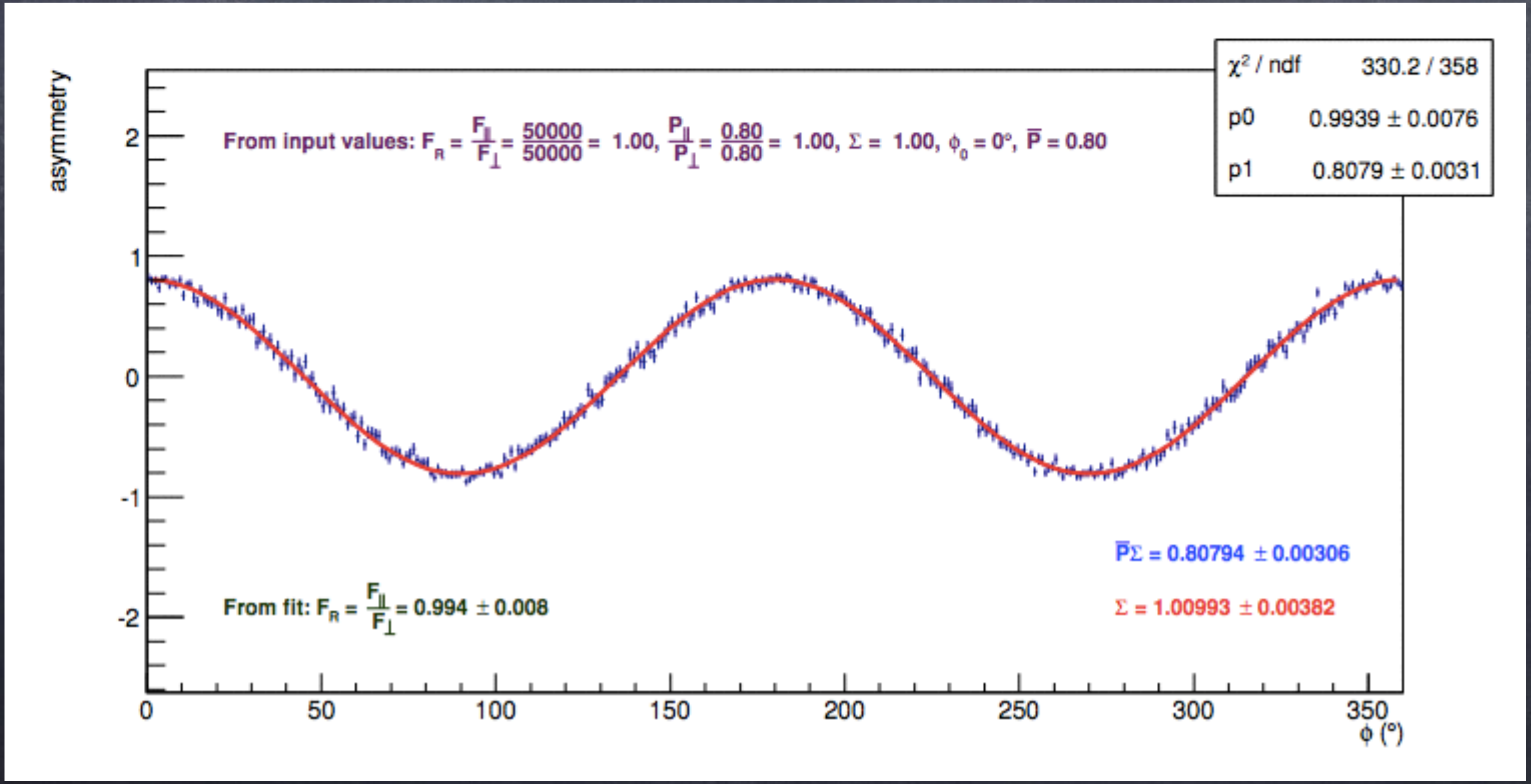
Simulation: technique

- Next, an asymmetry was created and fitted with Ken's famous function:

$$y = \frac{(A-1) + \frac{AB+1}{B+1}\, 2\, C \cos\left(2(x+\phi_0)\right)}{(A+1) + \frac{AB-1}{B+1}\, 2\, C \cos\left(2(x+\phi_0)\right)}$$

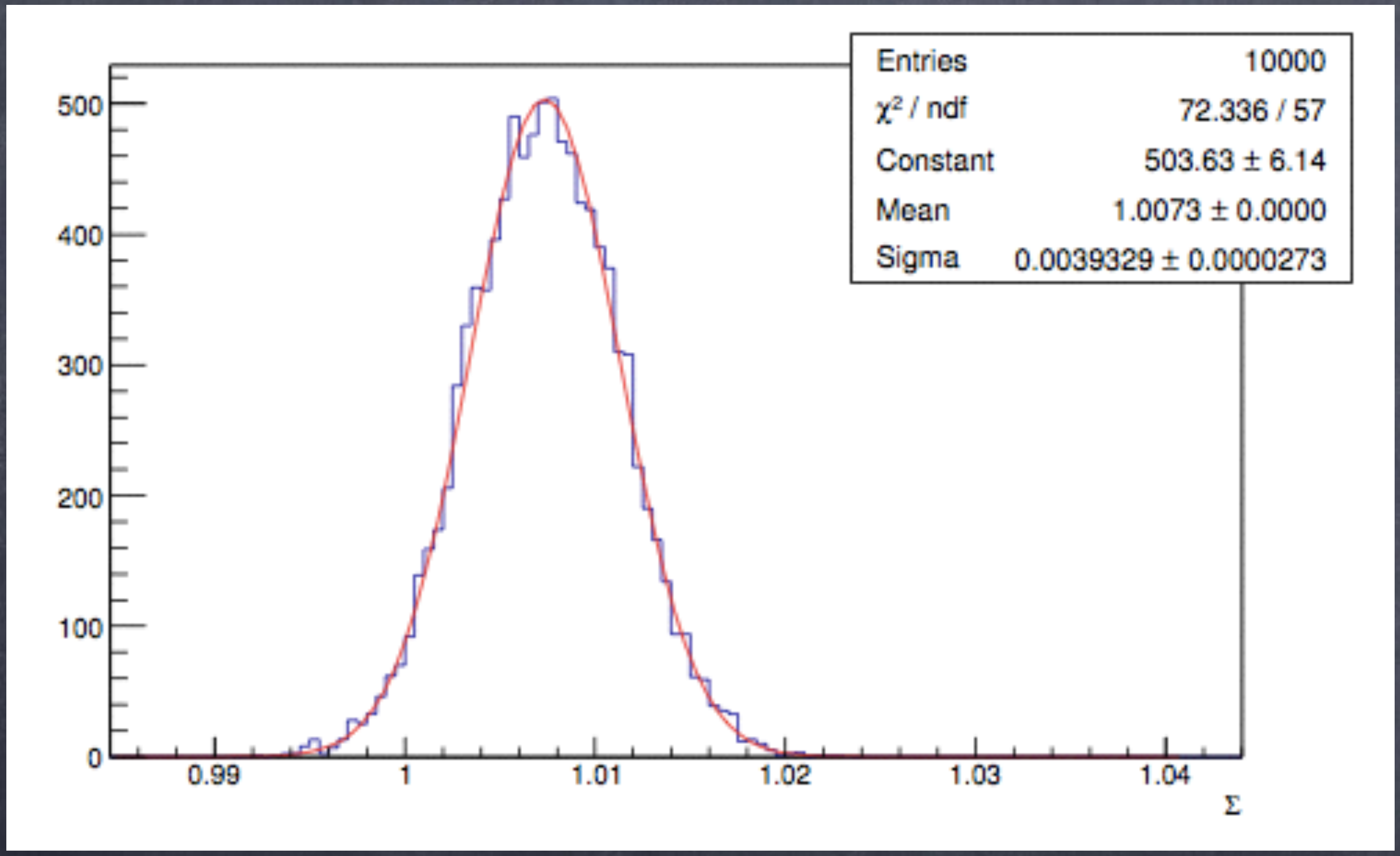
$$\frac{\sigma_\perp - \sigma_\parallel}{\sigma_\perp + \sigma_\parallel} = \frac{(N_R-1) + \frac{N_R P_R+1}{P_R+1}\, 2\, \bar{P}\, \Sigma \cos\left(2(\phi+\phi_0)\right)}{(N_R+1) + \frac{N_R P_R-1}{P_R+1}\, 2\, \bar{P}\, \Sigma \cos\left(2(\phi+\phi_0)\right)}$$

Asymmetry from randomly-generated PARA and PERP distributions, with 50,000 events each.



Simulation: technique

- This was repeated 10,000 times to get a statistical sample of extracted Sigma. The resulting distribution was fitted with a Gaussian to obtain a mean Sigma and the standard deviation was taken as the error on the Sigma.



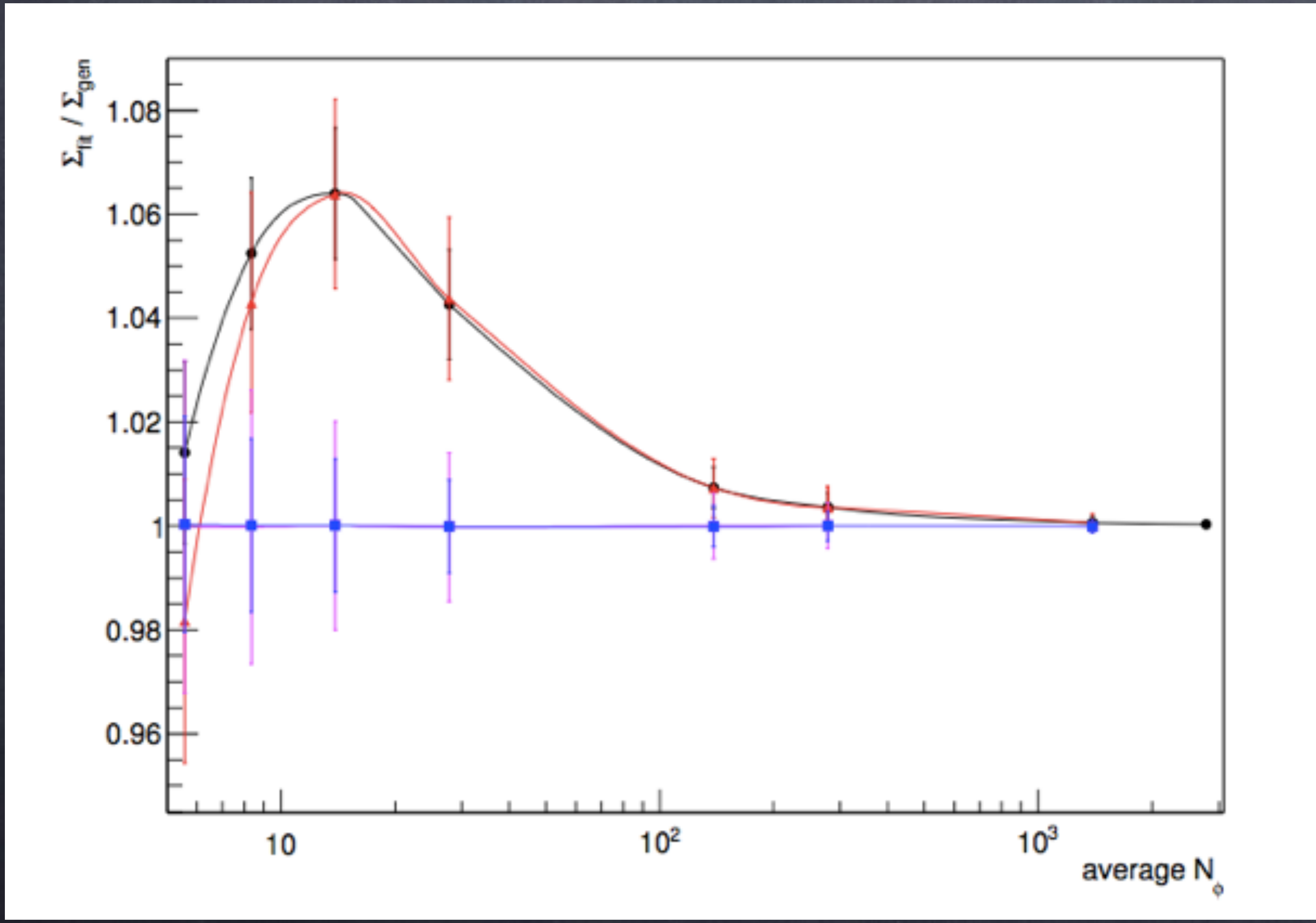
Sigma from the fit to the asymmetry produced from PARA and PERP histograms, each with 50,000 events and a generated Sigma of 1. Result of 10,000 simulation runs.

- The mean Sigma was then corrected for a finite phi-bin width, as in Nick's note:

$$\Sigma_{fit} = \frac{\Delta\phi}{sin(\Delta\phi)}\Sigma_{mean}$$

Simulation: results

- A systematic shift in the mean value of Sigma was observed for cases with low average number of events per phi-bin. This is the result of phi-bins with very low stats in each PARA or PERP distribution, and consequently significantly lower Poisson errors, biasing the fit.



Ratio of mean Sigma extracted from fit and corrected for finite phi-bin width / generated Sigma vs average number of events per phi-bin.

Black dots: from fit to asymmetry.

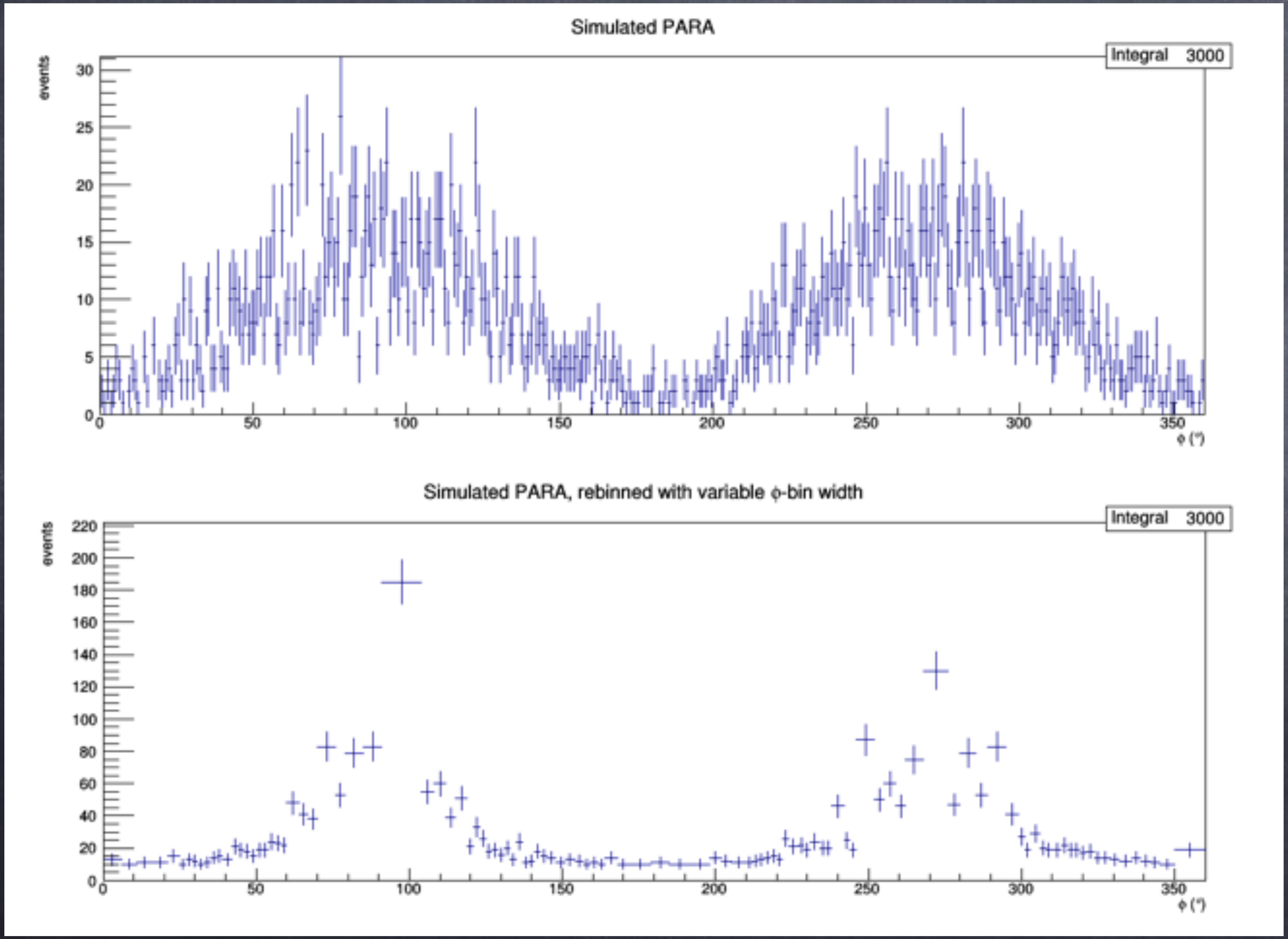
Red triangles: from fit to individual PARA and PERP distributions.

Pink dots: from fit to individual PARA and PERP distributions after artificially adding 1000 events to each phi-bin (to equalise errors somewhat).

Blue dots: from fir to asymmetry where error for each point is artificially set to 0.1 (to equalise errors).

How to reduce the stats-related systematic

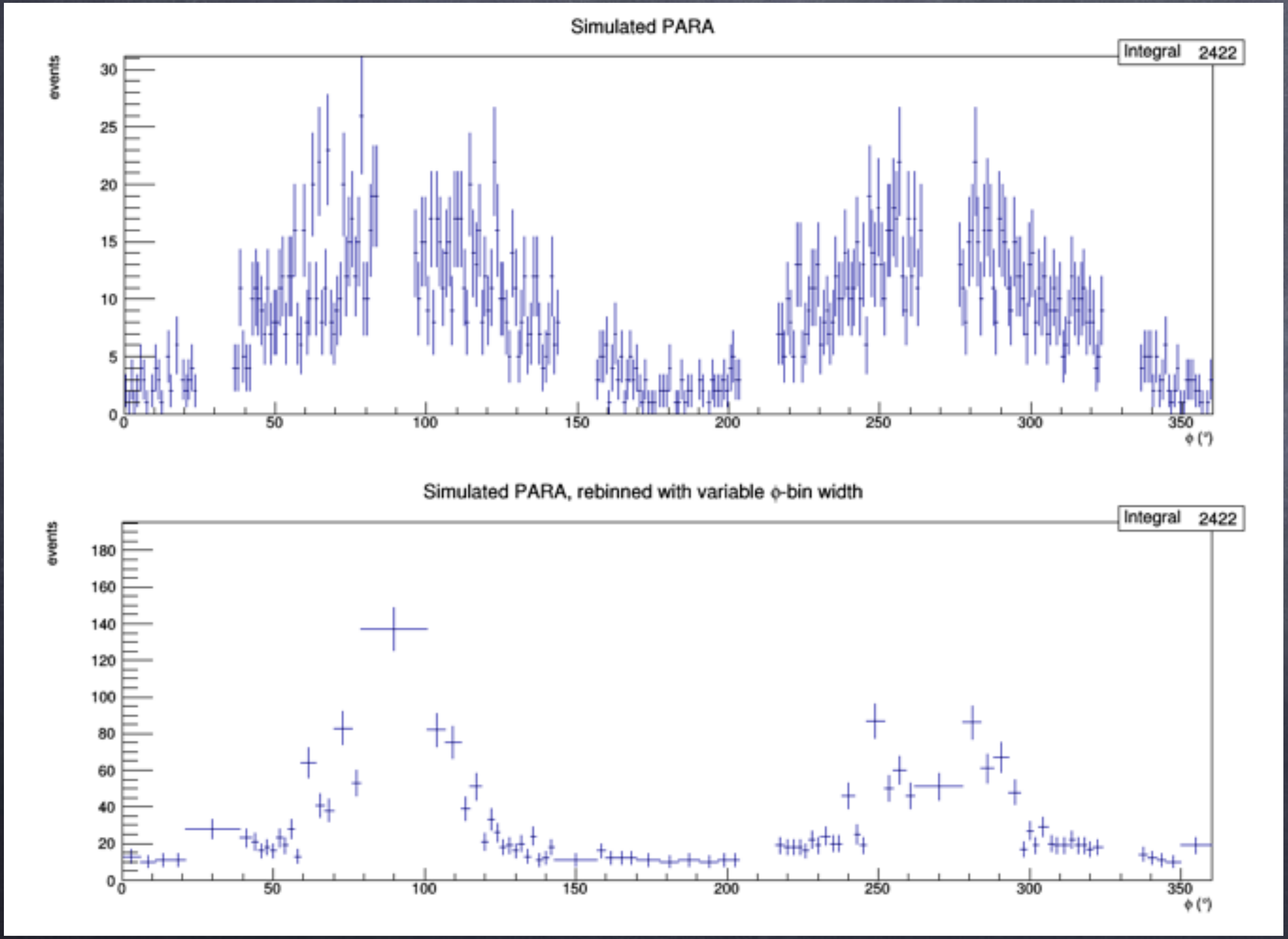
- Use variable phi-bins, requiring a min bin-width of 2deg (to agree with CLAS resolution) and requiring a min number of events per phi-bin (N_phi) in both the PARA and corresponding PERP histograms:



Simulated PARA histogram re-binned with a min N_phi = 10 (also in the corresponding PERP histogram, not shown), and a min phi-bin width of 2deg.

How to reduce the stats-related systematic

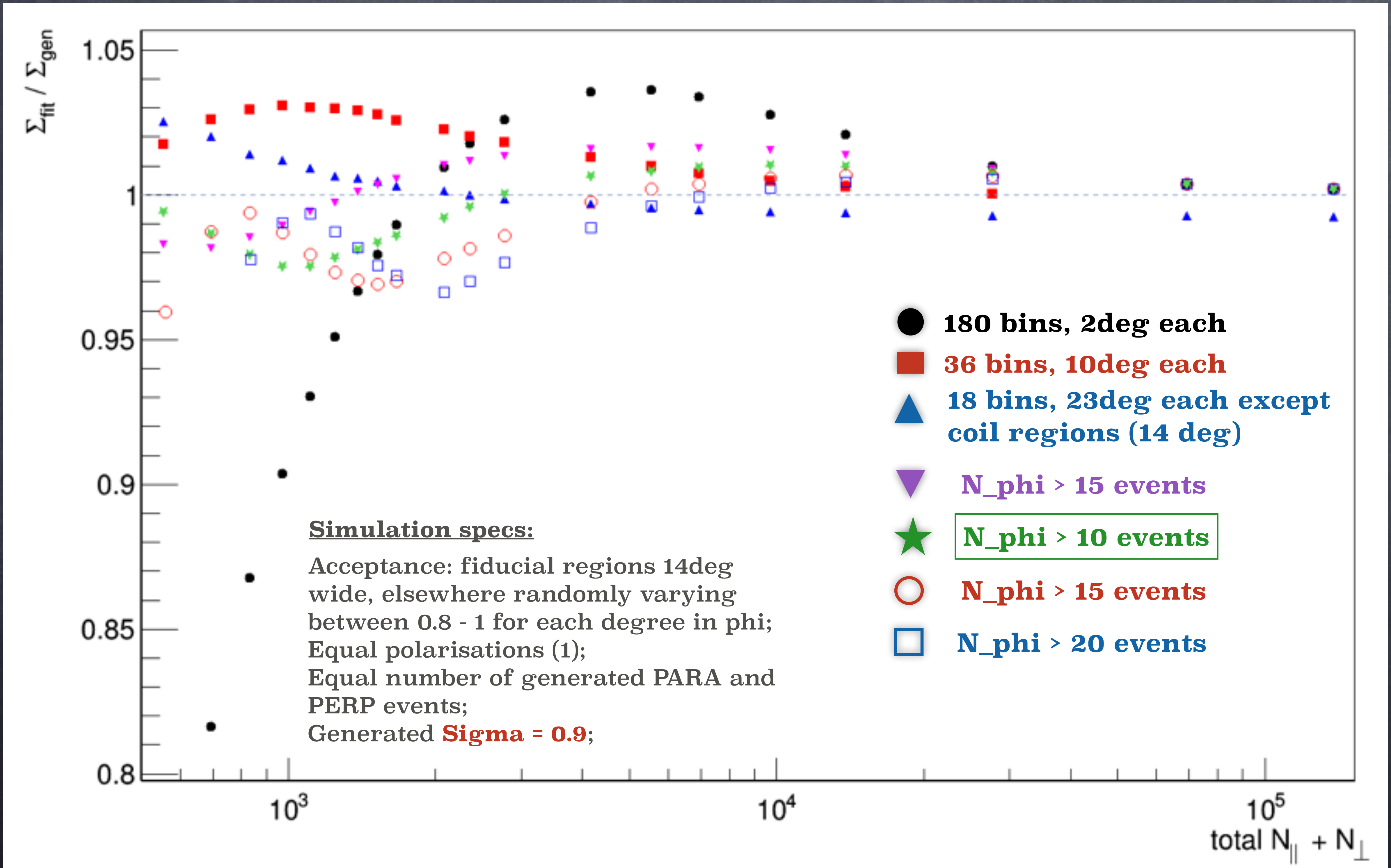
- This is the simulation where an acceptance function was applied to mimic CLAS, removing 6 coil regions, 12-degrees wide each:



Simulated PARA histogram with fiducial regions re-binned with a min N_phi = 10 (also in the corresponding PERP histogram, not shown), and a min phi-bin width of 2deg.

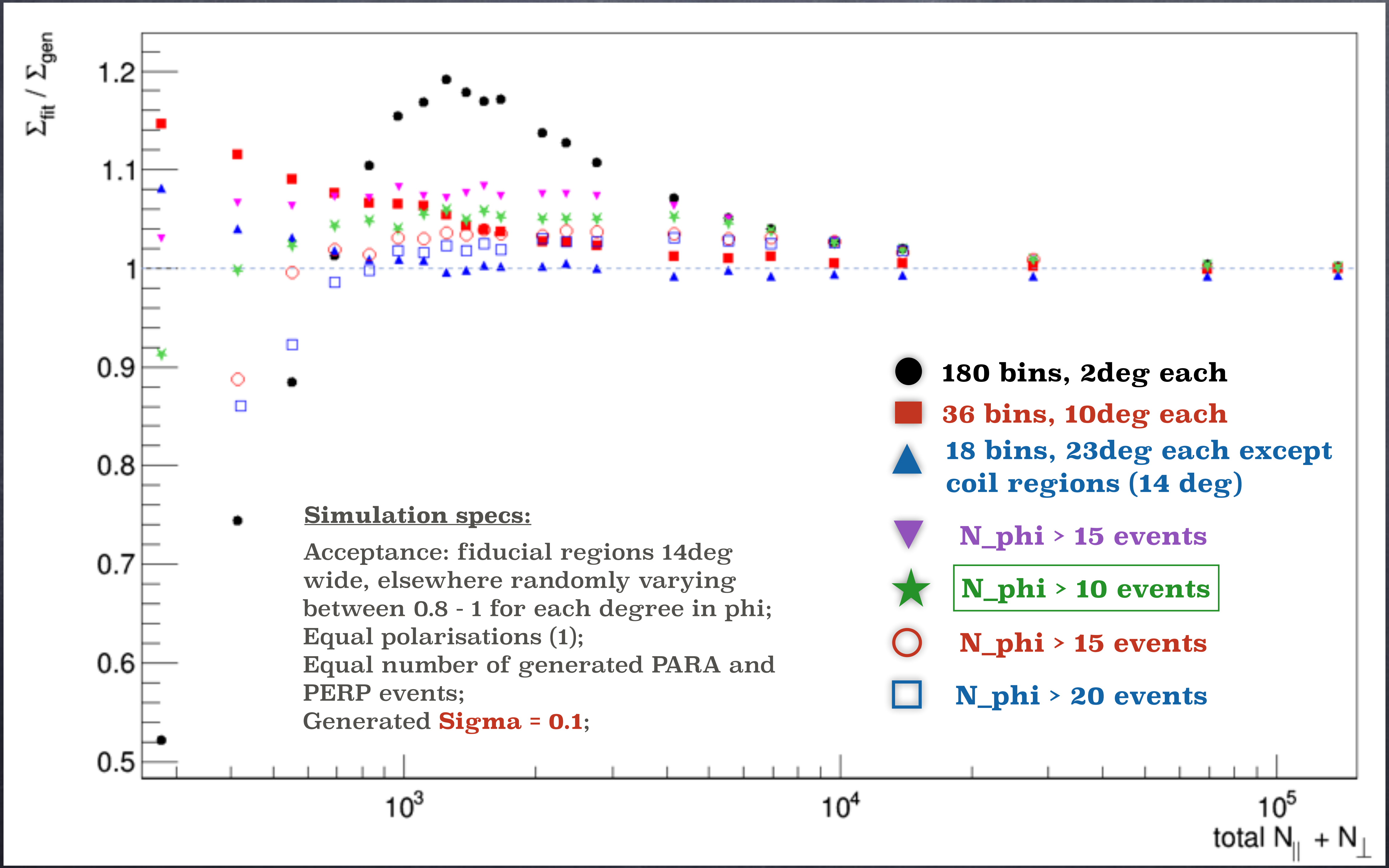
How to reduce the stats-related systematic

- A number of different binnings were tested, for a range of statistics in each simulated PARA and PERP histogram.



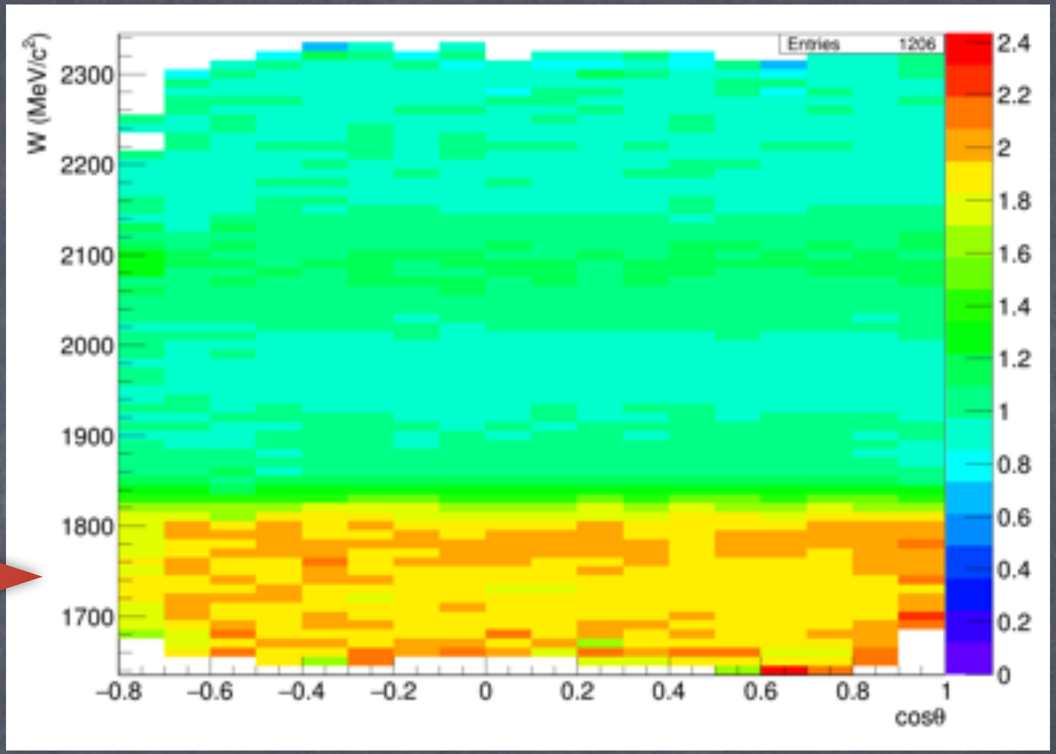
How to reduce the stats-related systematic

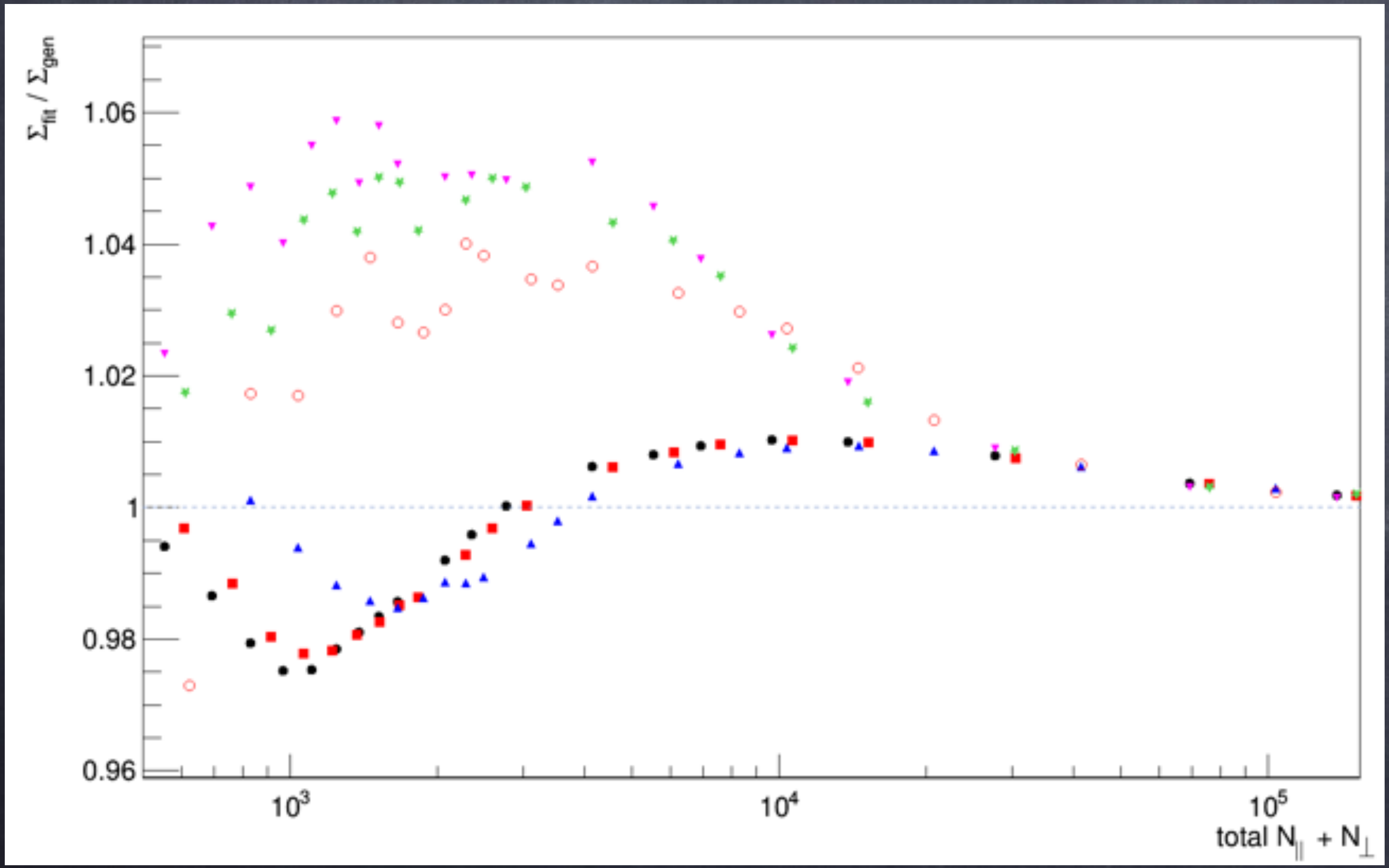
- Same simulation as previous slide, but with Sigma_gen = 0.1. Chosen binning method: variable bin-width, N_phi > 10 events in each, PARA and PERP histogram. Compromise between smallest reliable bin-width and enough bins per histogram.



Effect of flux ratio

- In the experiment, the flux ratio between PARA and PERP data (Nr), was mostly in the range 0.8 - 1.2, except the lower energy settings, where Nr rose to 2.





Simulation specs:

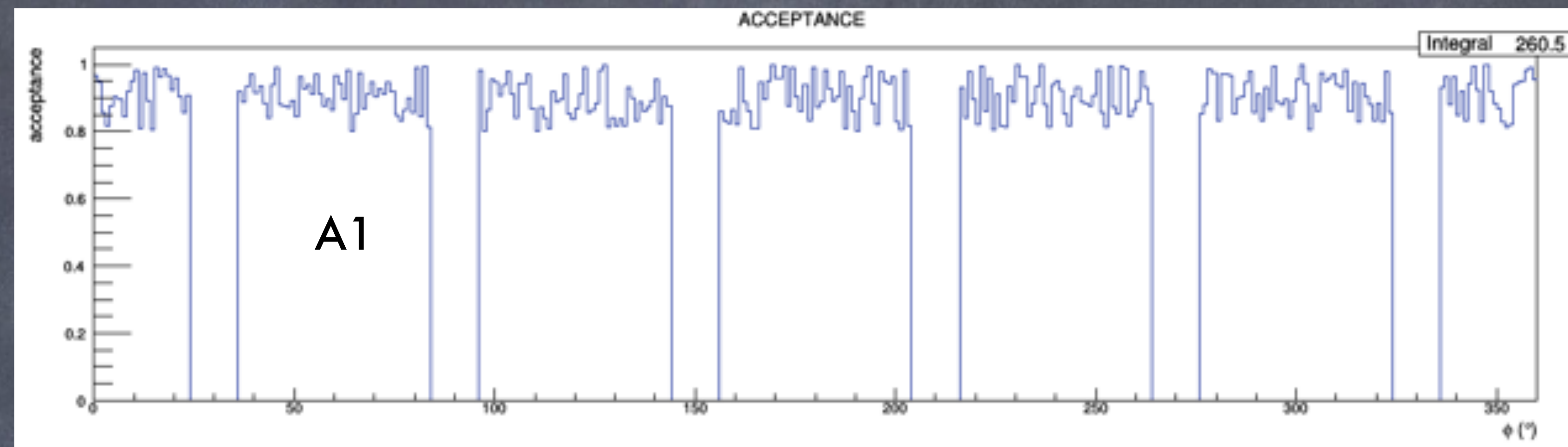
Acceptance: fiducial regions 14deg wide, elsewhere randomly varying between 0.8 - 1 for each degree in phi;
Equal polarisations (1);
N_phi > 10 for each PARA, PERP bin.

- ● Sigma_gen = 0.9, Nr = 1
- ■ Sigma_gen = 0.9, Nr = 1.2
- ▲ Sigma_gen = 0.9, Nr = 2
- ▼ Sigma_gen = 0.1, Nr = 1
- ★ Sigma_gen = 0.1, Nr = 1.2
- ○ Sigma_gen = 0.1, Nr = 2

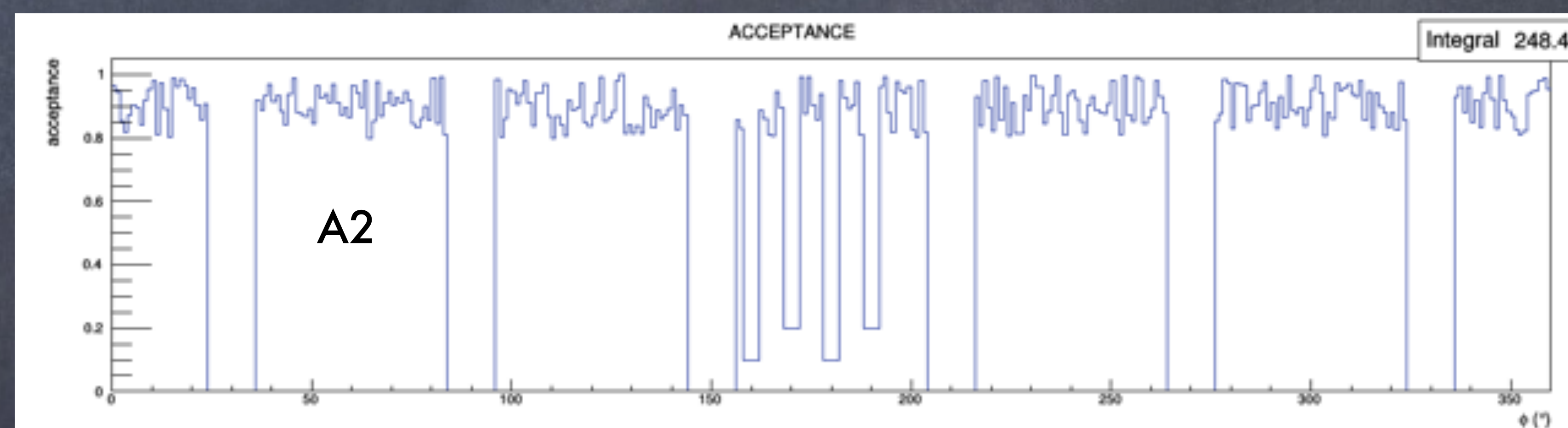
Acceptance functions

- Three different acceptance functions studied, based on the fiducial regions observed in data at different polar angles. In all cases, acceptance value was varied randomly between 0.8 - 1 for each deg in phi.

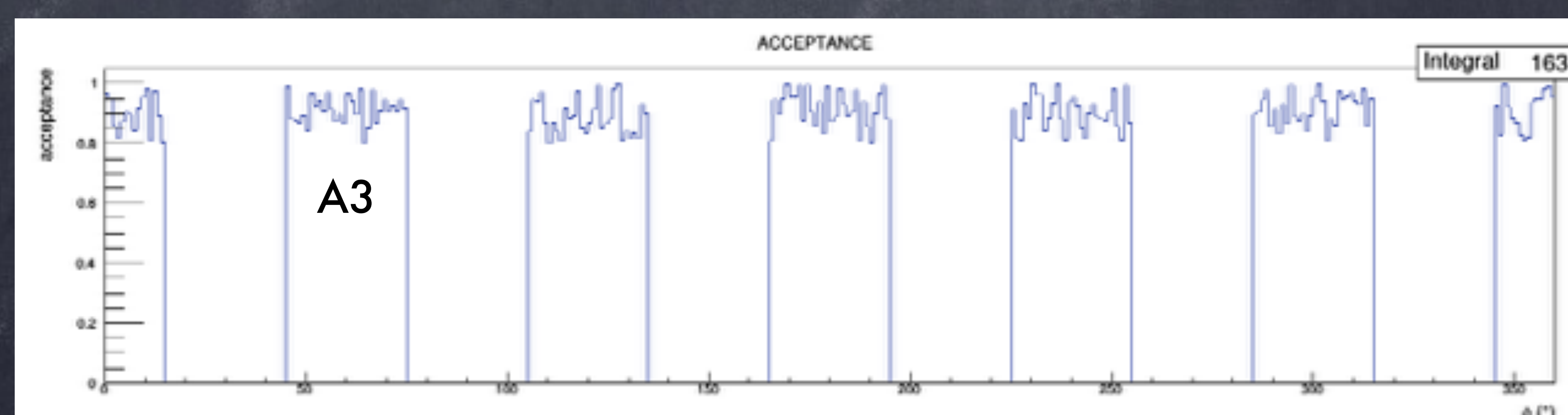
- A1: Fiducial regions 14 deg wide (corresponds to central polar angles in the data)



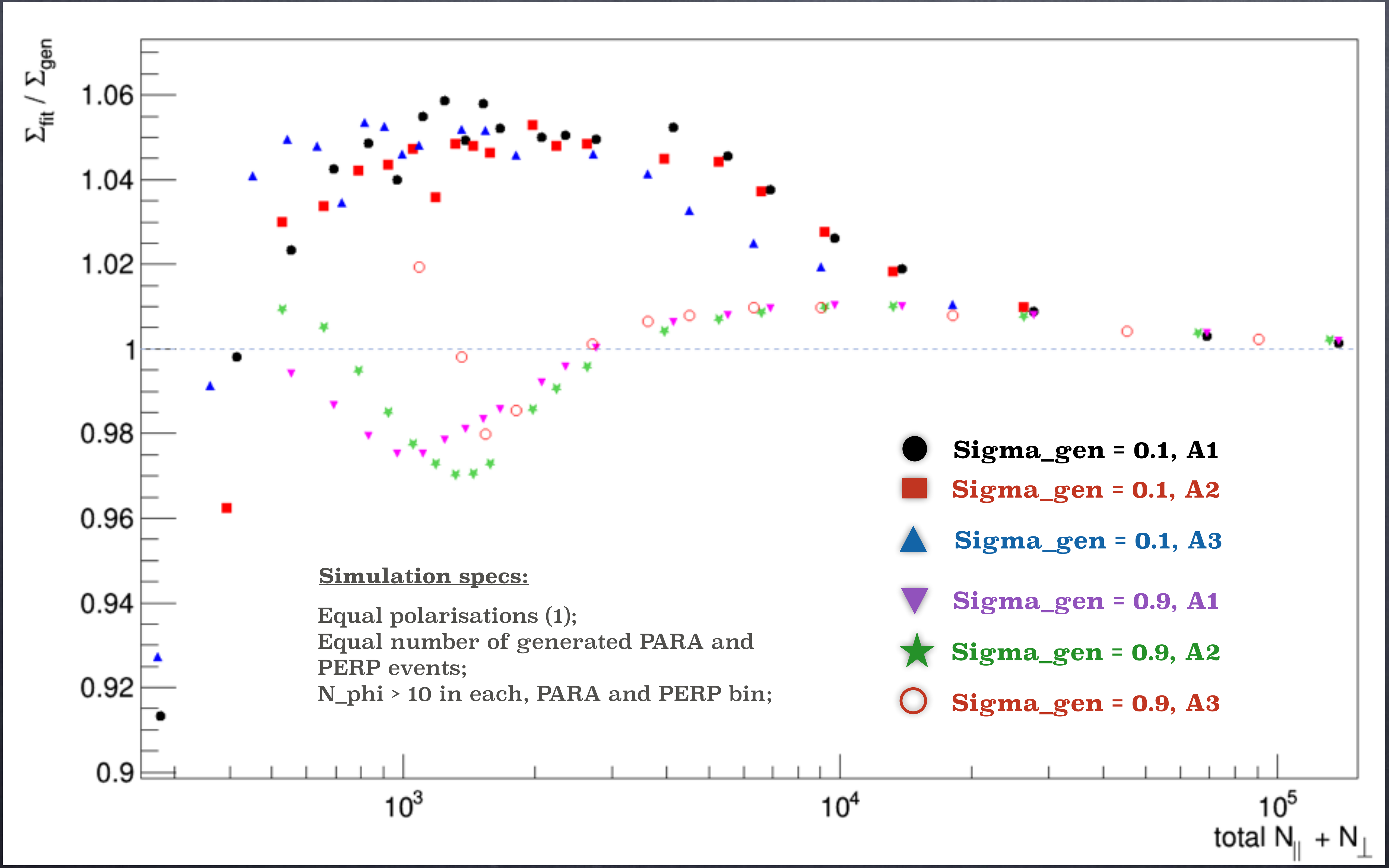
- A2: Fiducial regions 14 deg wide, plus one bad sector.



- A3: Fiducial regions 30 deg wide (corresponds to very forward data)

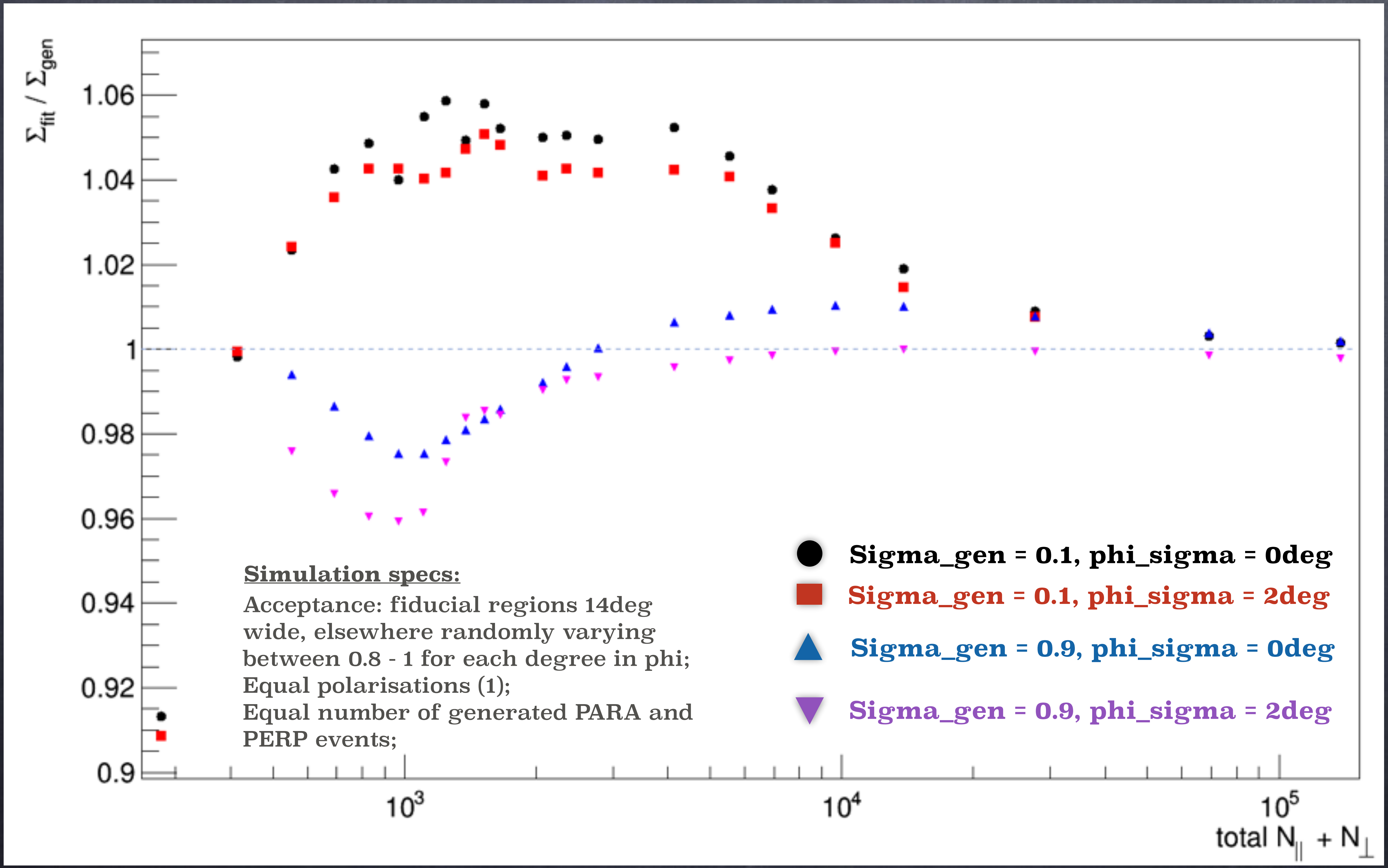


Effect of acceptance function



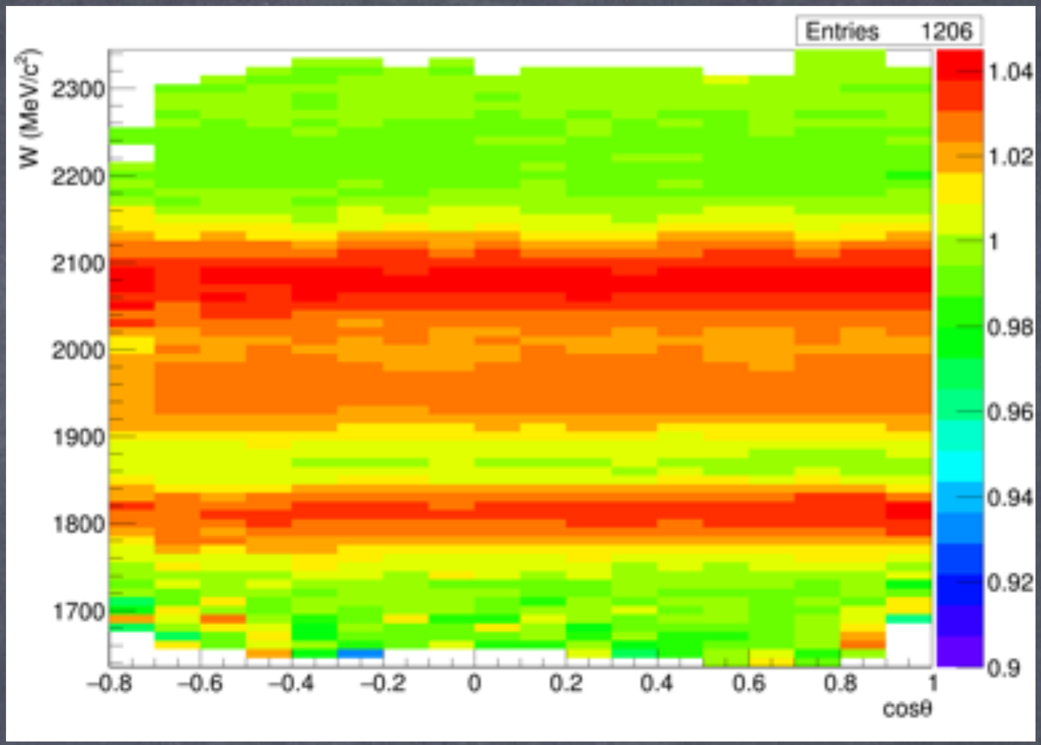
Effect of phi-resolution

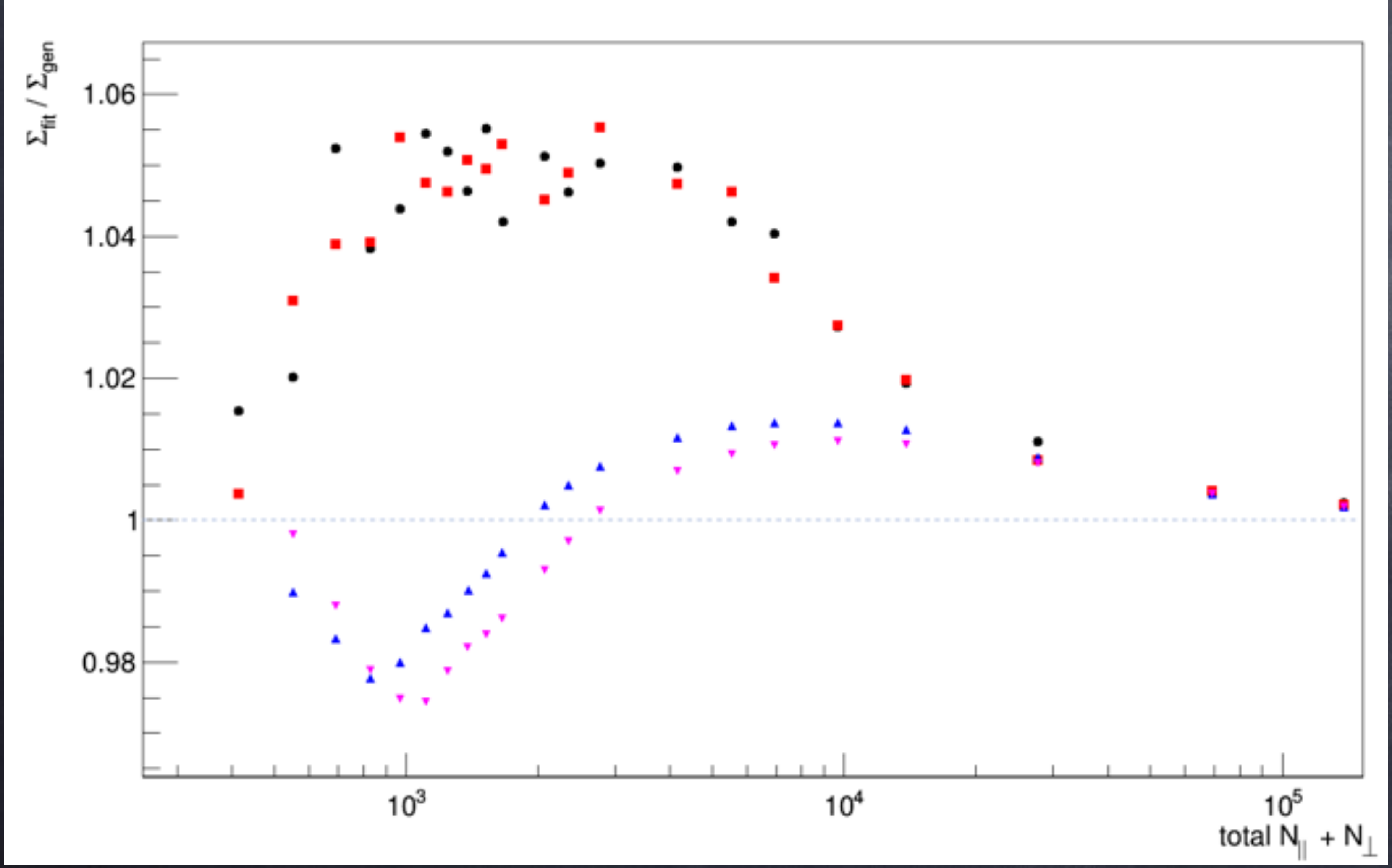
- A phi resolution was introduced by smearing the phi of each generated, acceptance-scaled event with a Gaussian function, sigma = 2deg.



Effect of Polarisation Ratio

- In the experiment, the polarisation ratio for each bin varied by +/- 4%.





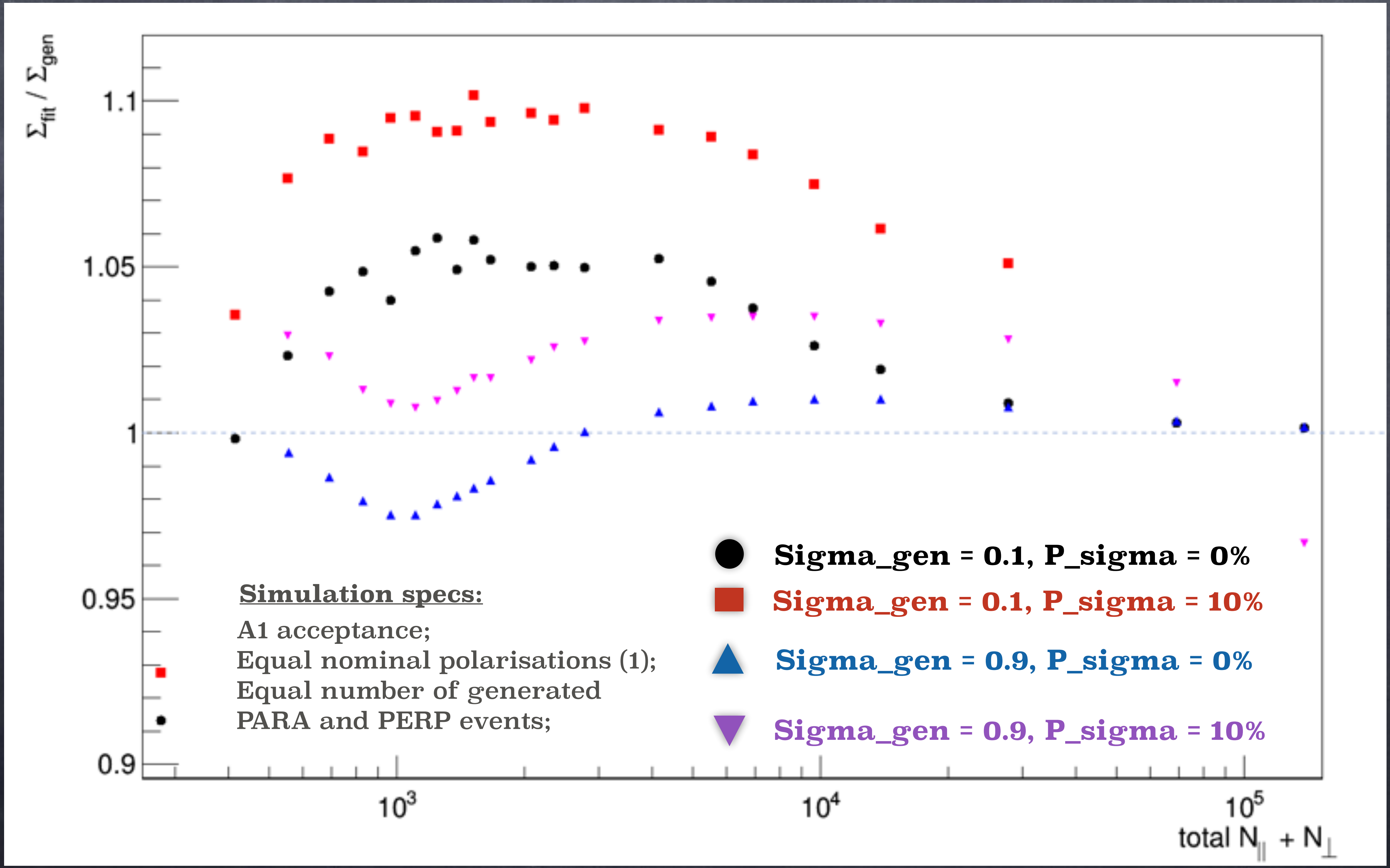
Simulation specs:

Acceptance: fiducial regions 14deg wide, elsewhere randomly varying between 0.8 - 1 for each degree in phi;
Equal number of generated PARA and PERP events;
N_phi > 10 for each PARA, PERP bin;
Average P = 0.95

- ● Sigma_gen = 0.1, Pr = 1
- ■ Sigma_gen = 0.1, Pr = 0.9
- ▲ Sigma_gen = 0.9, Pr = 1
- ▼ Sigma_gen = 0.9, Pr = 0.9

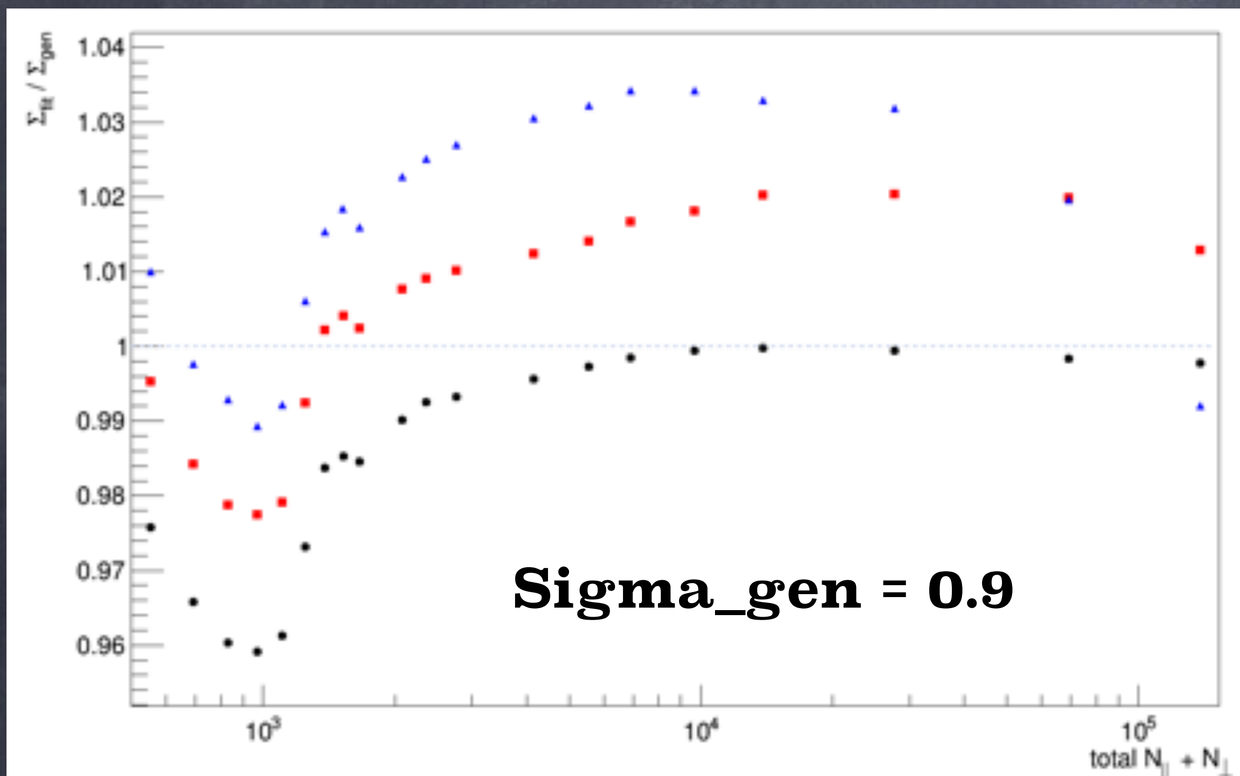
Effect of Polarisation Error

- Depending on the photon energy, polarisation error is expected to vary between a few - 10%.
- Introduced into simulation by smearing the P value of each PARA or PERP event with a Gaussian, sigma = 0.1*nominal P



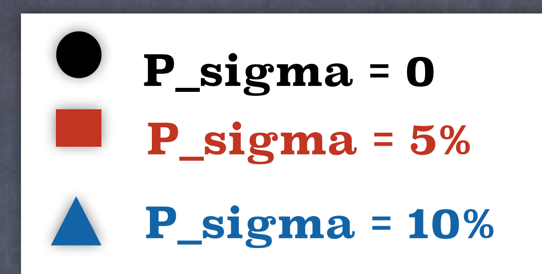
“Quasi-realistic simulation”

- Simulations also carried out with “realistic” conditions, where a smearing was introduced both on phi (2deg resolution) and Polarisation:

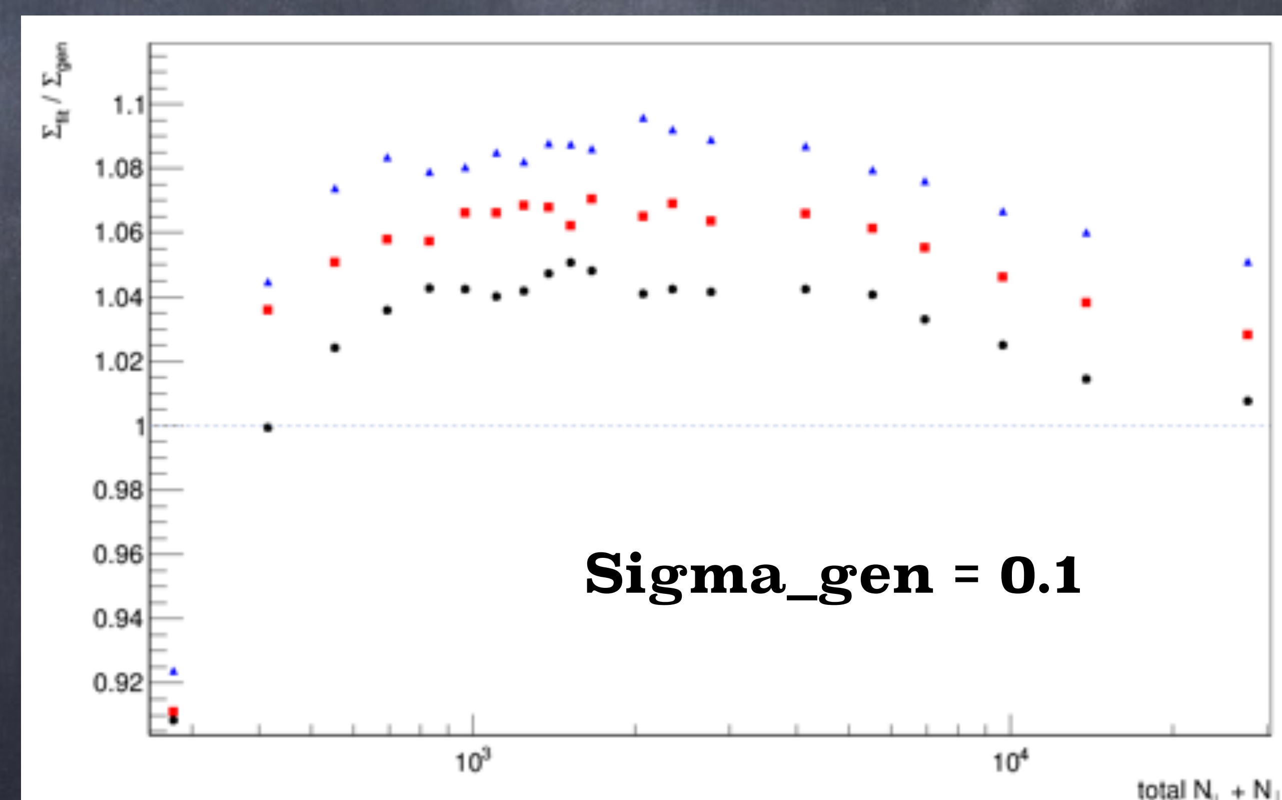


Simulation specs:

Acceptance: fiducial regions 14deg wide, elsewhere randomly varying between 0.8 - 1 for each degree in phi;
Equal polarisations (1);
N_phi > 10 for each PARA, PERP bin.



- Since Pr is a fixed parameter in the fit, an error in P affects the fit itself.



Conclusions from the simulations

- The statistics in the PARA and PERP histograms have a systematic effect on the fit-extracted asymmetry, which can be minimised by adopting a variable binning in phi. Requiring a minimum 10 events in each phi bin offers a compromise between reliable stats in each bin for the Poisson error to apply, and a large number of bins for the histogram.

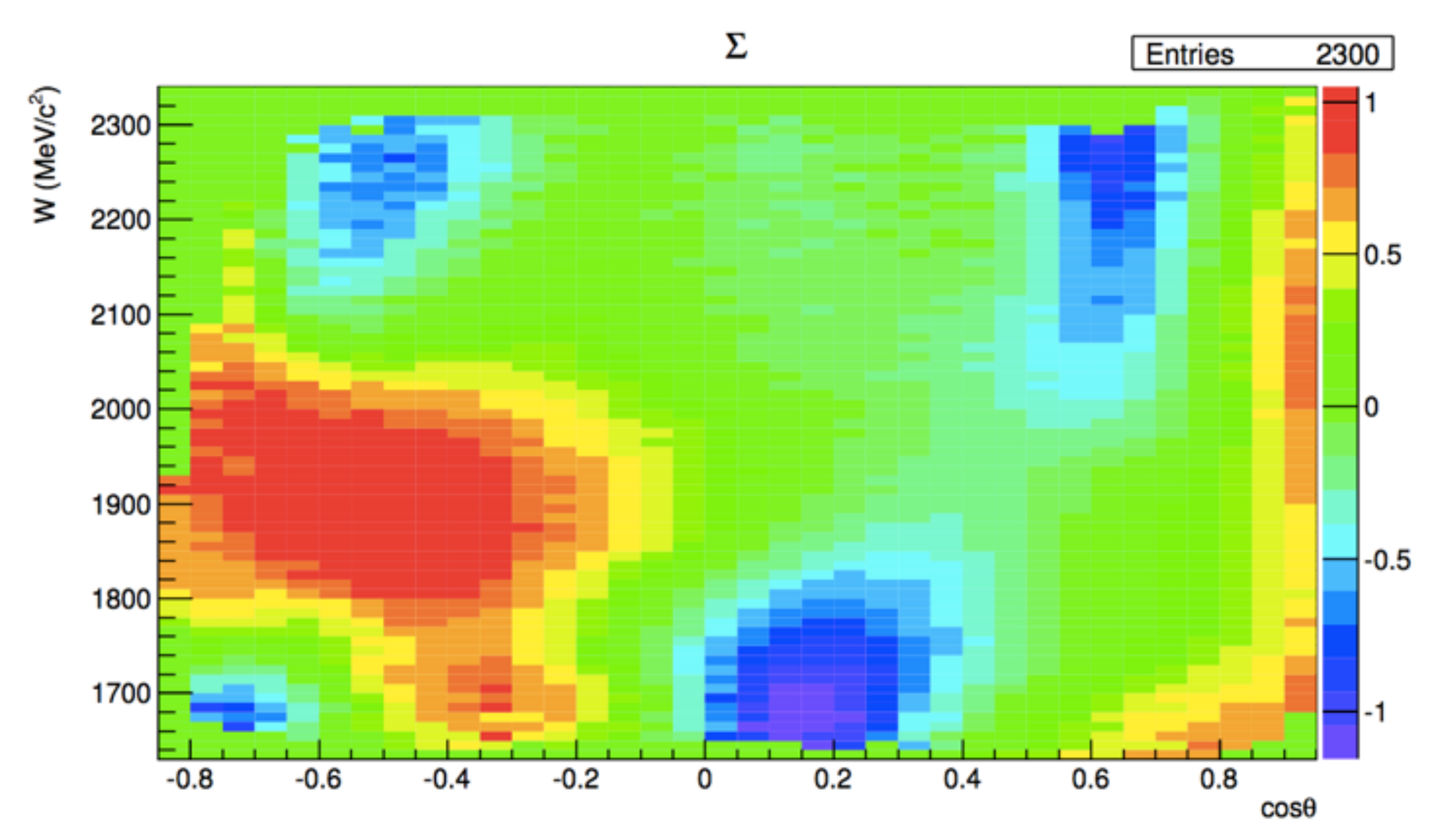
- The systematic effect is a shift on the order of 2% (PSigma = 0.9) to 6% (PSigma = 0.1), which varies by about 2% depending on the conditions: flux ratio, acceptance, polarisation ratio, phi-resolution.

- Polarisation error introduces a bigger systematic shift (up to 4% for PSigma = 0.9, 9% for PSigma = 0.1 for P_sigma = 10%), and an additional error on extracting the observable from the asymmetry.

Phi_0 offset

Why g13 Sigma data is a bitch

- Consider how quickly Sigma changes with cos-theta and W:



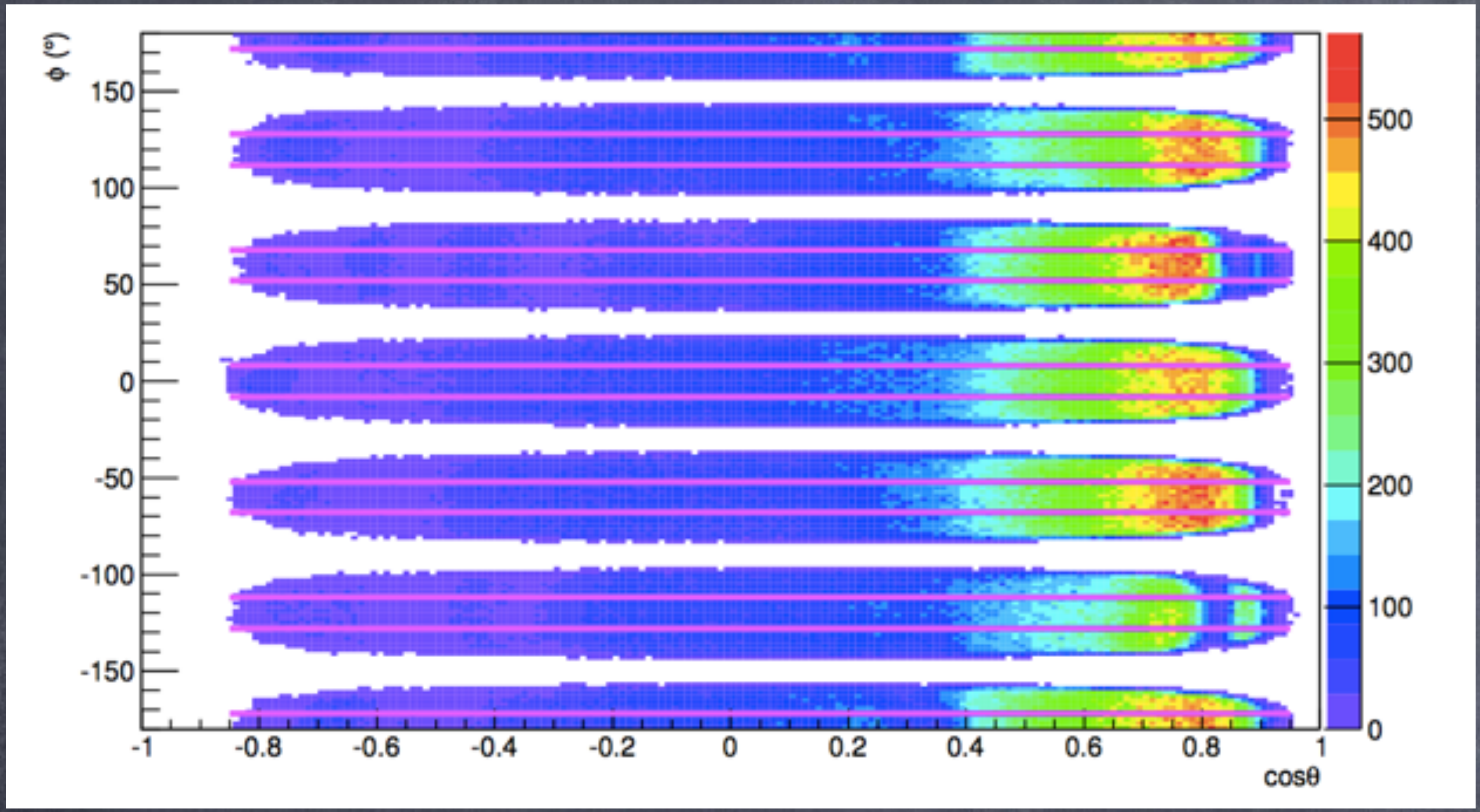
- This makes any tests based on summing across large areas of phase-space really difficult.

For example, to determine phi_0

- In the fit equation: $$\frac{\sigma_\perp - \sigma_\parallel}{\sigma_\perp + \sigma_\parallel} = \frac{(N_R - 1) + \frac{N_R P_R + 1}{P_R + 1} 2\,\bar{P}\,\Sigma\,\cos\left(2(\phi + \phi_0)\right)}{(N_R + 1) + \frac{N_R P_R - 1}{P_R + 1} 2\,\bar{P}\,\Sigma\,\cos\left(2(\phi + \phi_0)\right)}$$
- Phi_0 is a rotation between the PARA / PERP directions and the cartesian axes of CLAS. It is the result of imperfect alignment of the diamond crystal.
- In principle, can be extracted from a fit to a super-high statistics asymmetry distribution – for example by summing over all regions of phase space where Sigma is positive or negative.
- But there are pitfalls!

Pitfall number 1

- Observe the angular distribution of hits in CLAS:



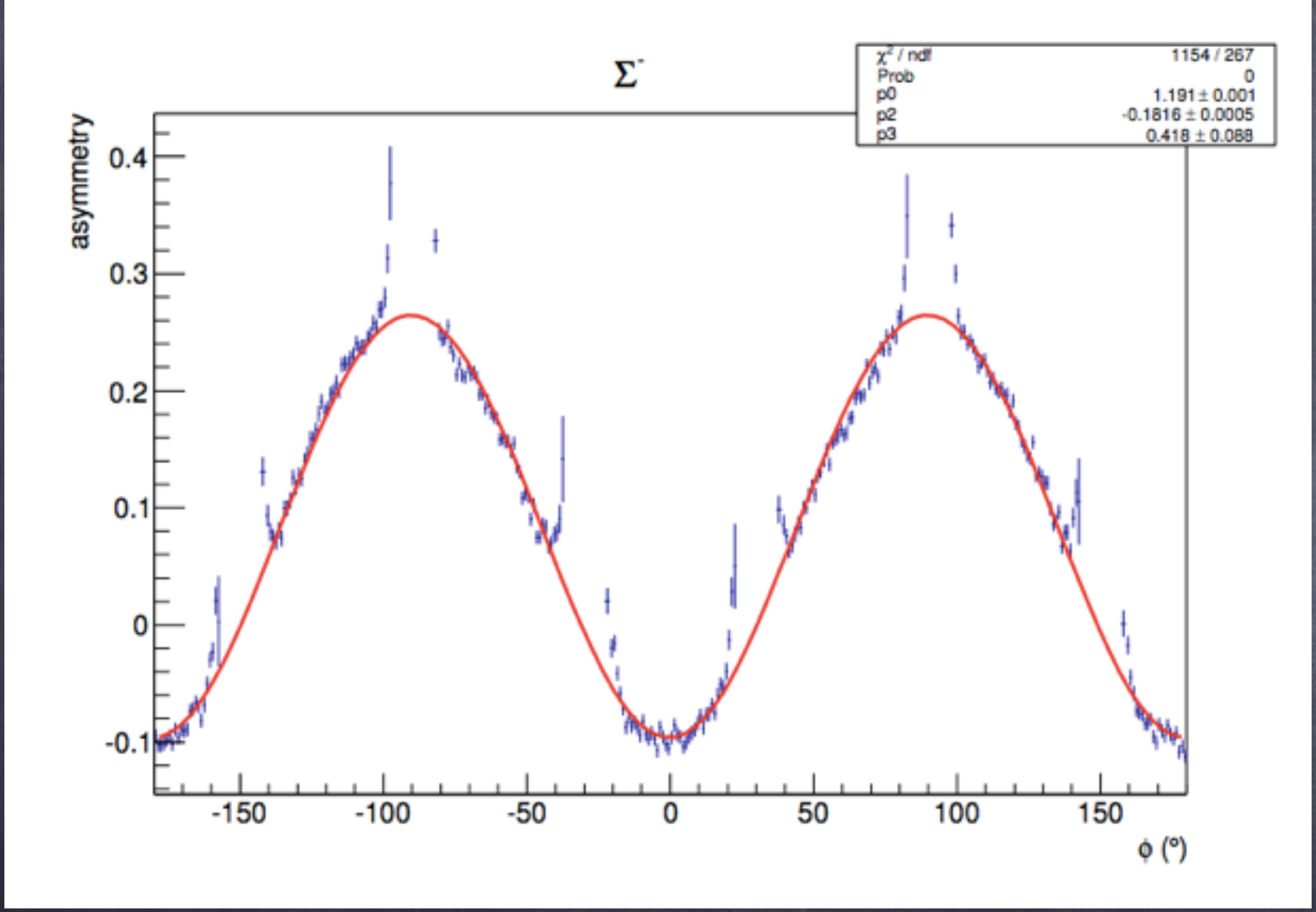
CM frame, all the data in the setting 4.2/1.3 GeV. Pink lines show fiducial cuts ultimately applied for the phi_0 study.

- For phi-bins close to the torus coils, the very forward / backward data is missing.

- This means the normalisation constant will be different in bins at the torus coils from bins in the centres of the sectors (because cos-theta has a W correlation, so the contributing datasets will be different).

Pitfall number 1 continued

- ... and what you get is something that looks like this:

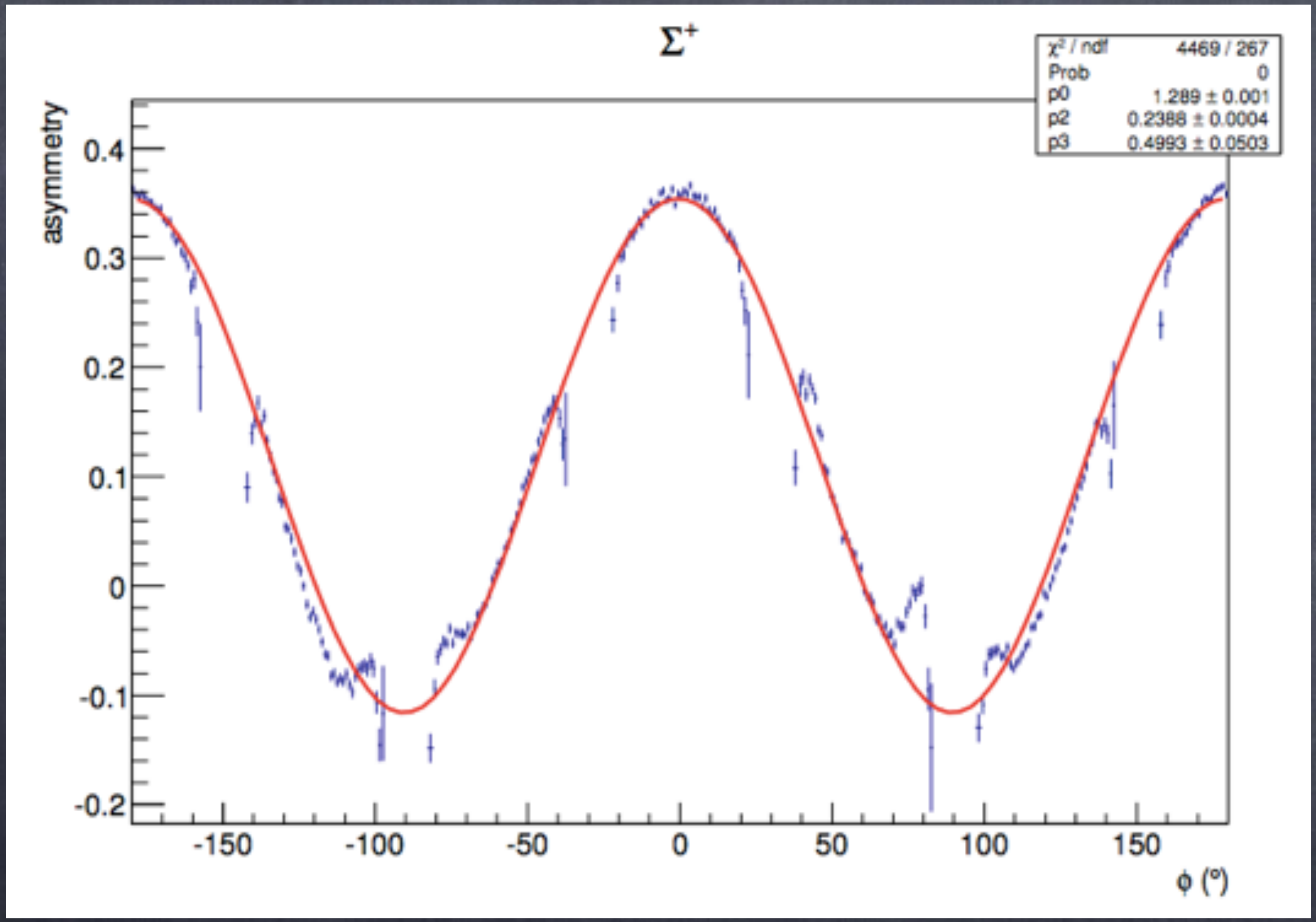


Asymmetry from the whole dataset, selecting only those regions where Sigma is negative

- Observe the whiskers at the torus coils – those bins have a different normalisation constant from the bulk of the phi-bins.

Pitfall number 2

- Additionally, remember that Sigma varies dramatically with theta! Which means the phi-bins close to the coils also have a different average Sigma from those in the middle of the sectors:



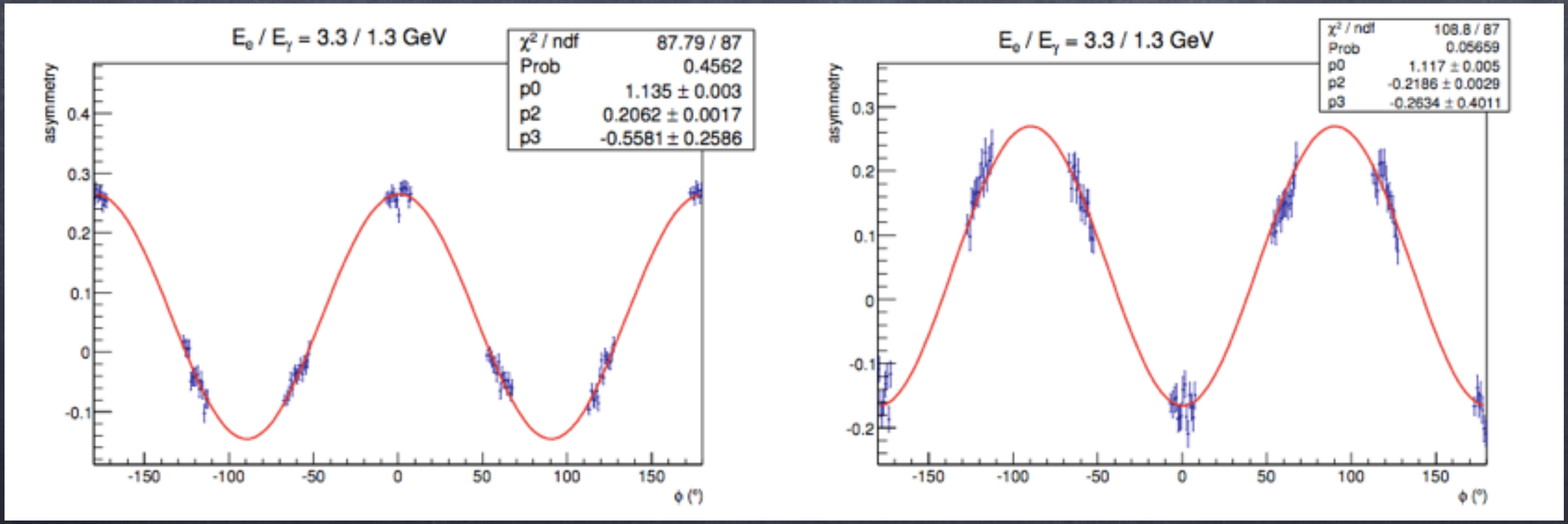
Asymmetry from the whole dataset, selecting only those regions where Sigma is positive

- In addition to the whiskers, there is also severe distortion close to the coils — since Sigma determines the amplitude of the sine modulation.

A solution

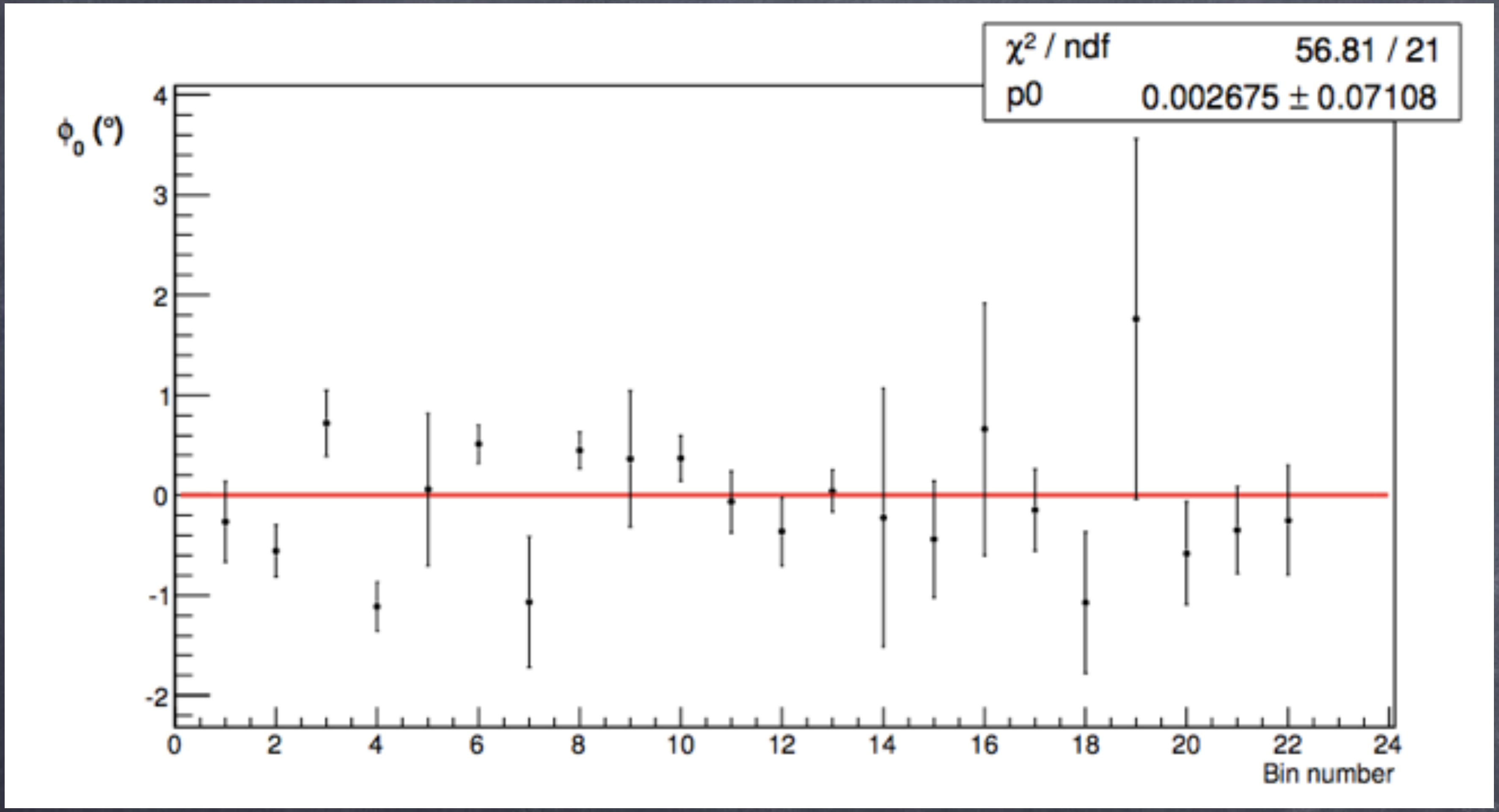
- Throw the heck out all phi which don't have contributions across the whole available range of theta (+/-20 deg, I think).

- A shame to lose much data, but can keep the very high stats and rely on a perfect fit in the middle of sectors:



A solution

- This was done for each setting separately and an average phi_0 offset determined:



Phi_0 determined from fits to positive and negative asymmetries from all settings.

- Conclusion: no axes rotation to speak of. Phi_0 = 0. Ken did an amazing job with that goniometer.

Can we use the “bad”, tripping-beam data?

How much of the data was trippy?

- During the run, we had a lot of trips. Here is how many of the final, good events for this reaction are labelled "trippy", per setting:

E_e/E_γ GeV	% of events removed by trip cuts from the final data
3.3/1.3	8.8
3.9/1.3	11.1
4.2/1.3	11.9
4.1/1.5	8.5
4.5/1.5	13.0
4.1/1.7	8.9
4.7/1.7	13.5
5.1/1.9	18.0
5.1/2.1	17.5
5.2/2.1	17.1
5.2/2.3	18.8

- May be greedy, but seems a shame to lose 10 - 20% of data for no good reason...

The problem with trips and how to compare:

- In principle, “bad beam” shouldn’t affect asymmetries.
- In practice, polarisation tables were produced on data with trip cuts applied. So the average polarisation might not be the same for “trippy” and “non-trippy” data.
- Check by comparing Sigma from data with trip-cuts applied and from the trippy events that those cuts remove (after applying bin-width / stats corrections).
- Do this for each setting as polarisation tables were produced setting-by-setting.
- Compare by plotting a distribution of ratios (Sigma from bad beam / Sigma from good beam) and fitting it with a straight line.
- Ideally, want to compare individual E-counters. Can’t do that as due to Fermi motion, each E-counter corresponds to a wide range of W. Not enough stats to bin in E-counter, W and cos theta, so just bin in 20 MeV (W), 0.1 in cos-theta to avoid problems with wildly changing Sigma.
- Throw away data-points which have an average N_{phi} per phi-bin that’s lower than the min Sigma-dependent boundary, or a PSigma < |0.1|, or a bad fit probability.

Criteria to keep / discard trippy events:

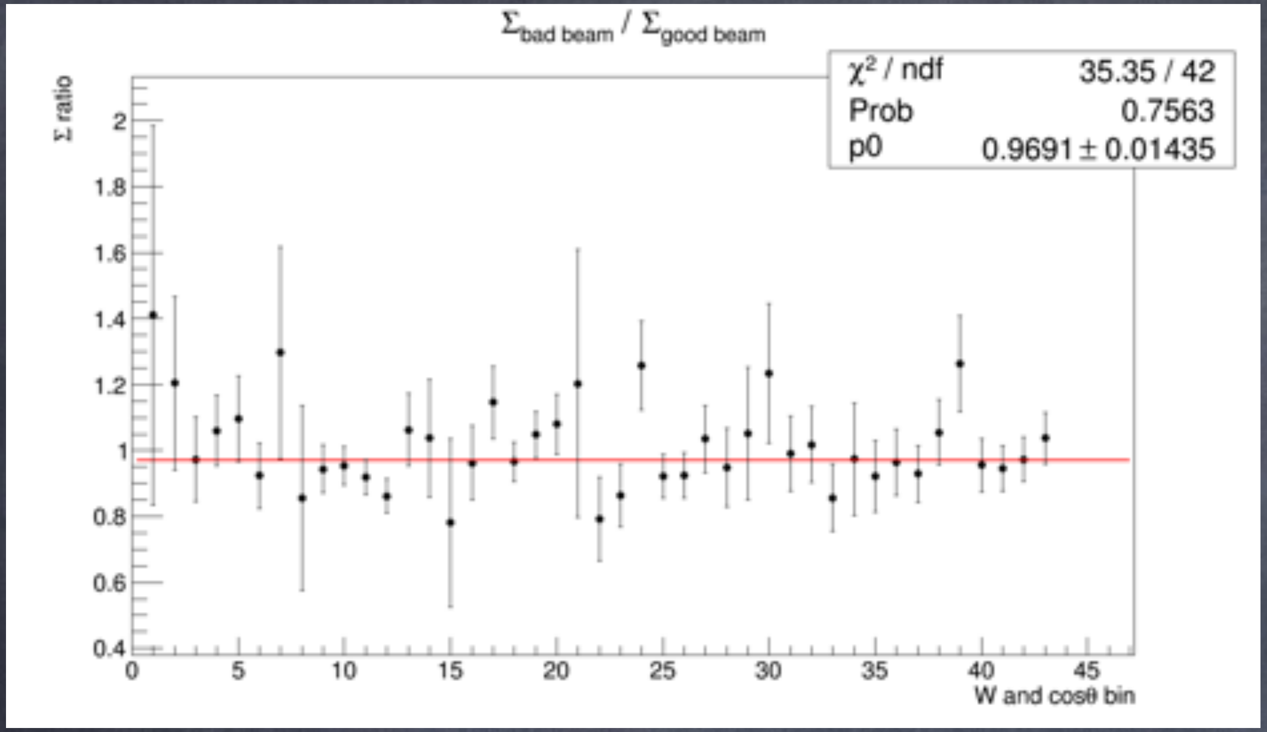
- Decide to keep the "bad beam" data for settings where the straight-line fit produces an average Sigma ratio of 1 within 3 sigma.
- What about making a pull distribution (or Welch's t-test) for each setting? Tried, but statistics too low to be able to meaningfully fit with a Gaussian.

Thoughts/comments...?

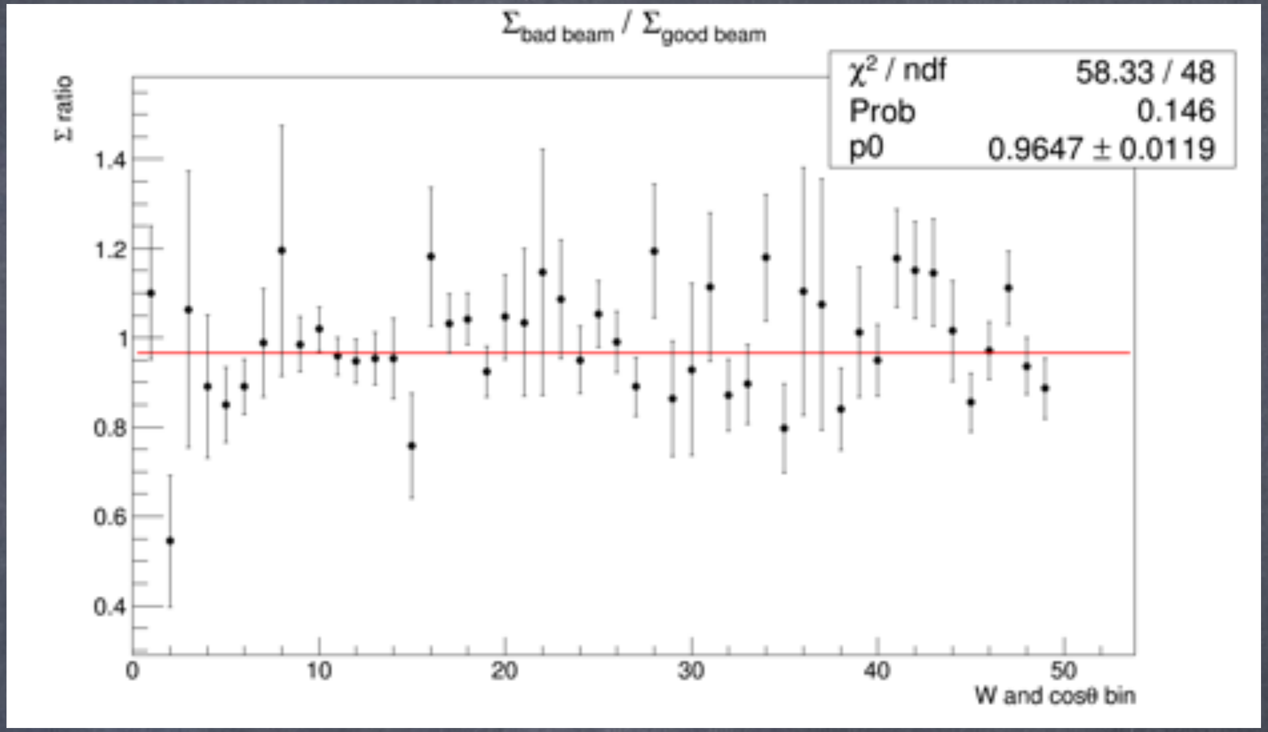
- See results on next page...

"Bad beam" / "good beam" comparison: I

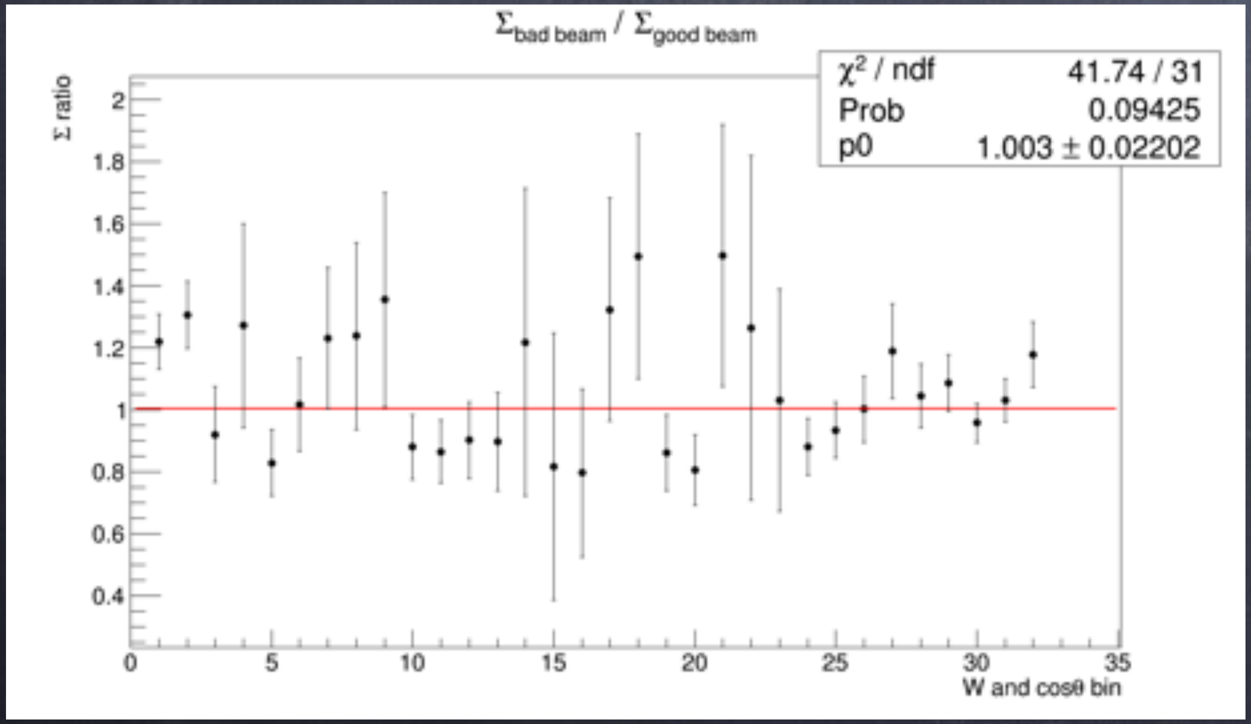
3.3/1.3 GeV



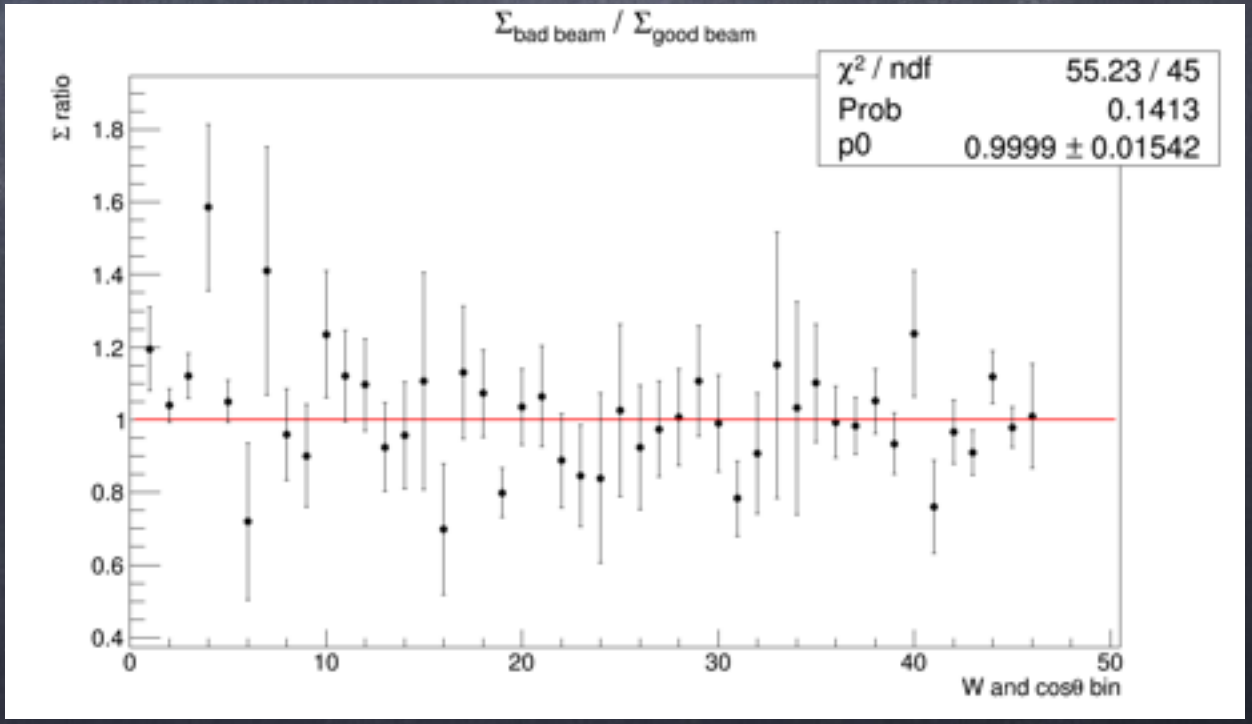
4.2/1.3 GeV



4.1/1.5 GeV

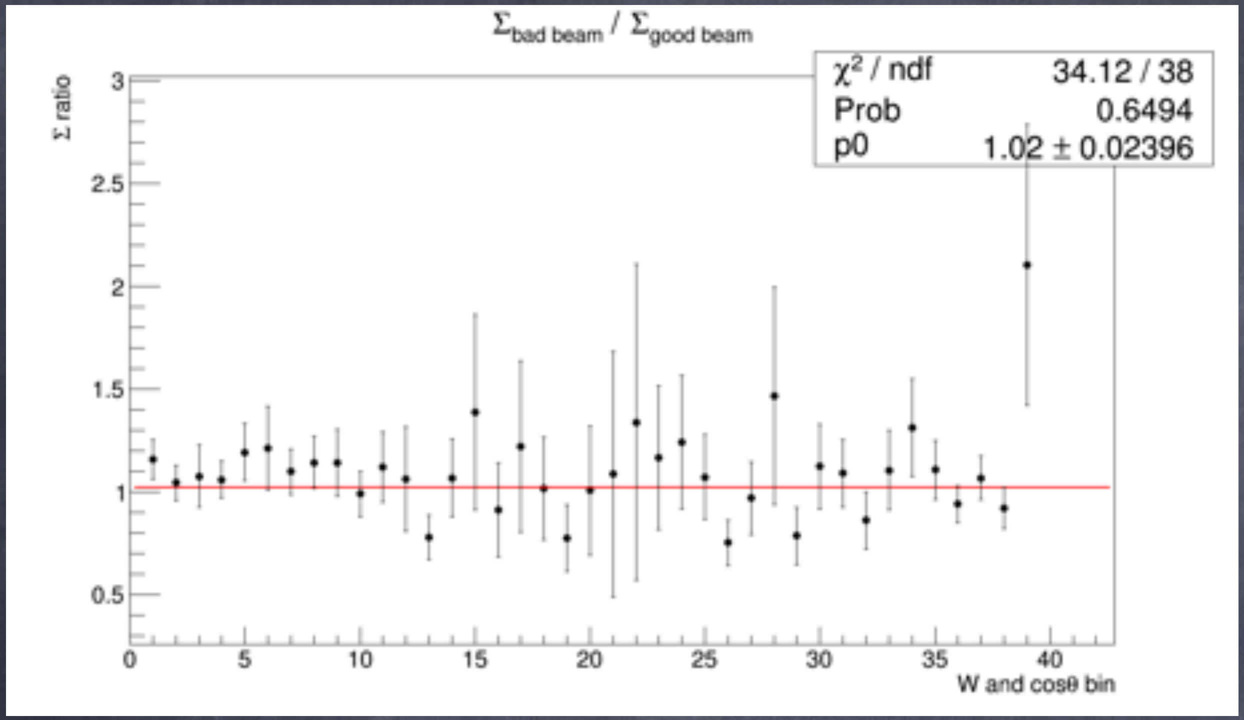


4.5/1.5 GeV

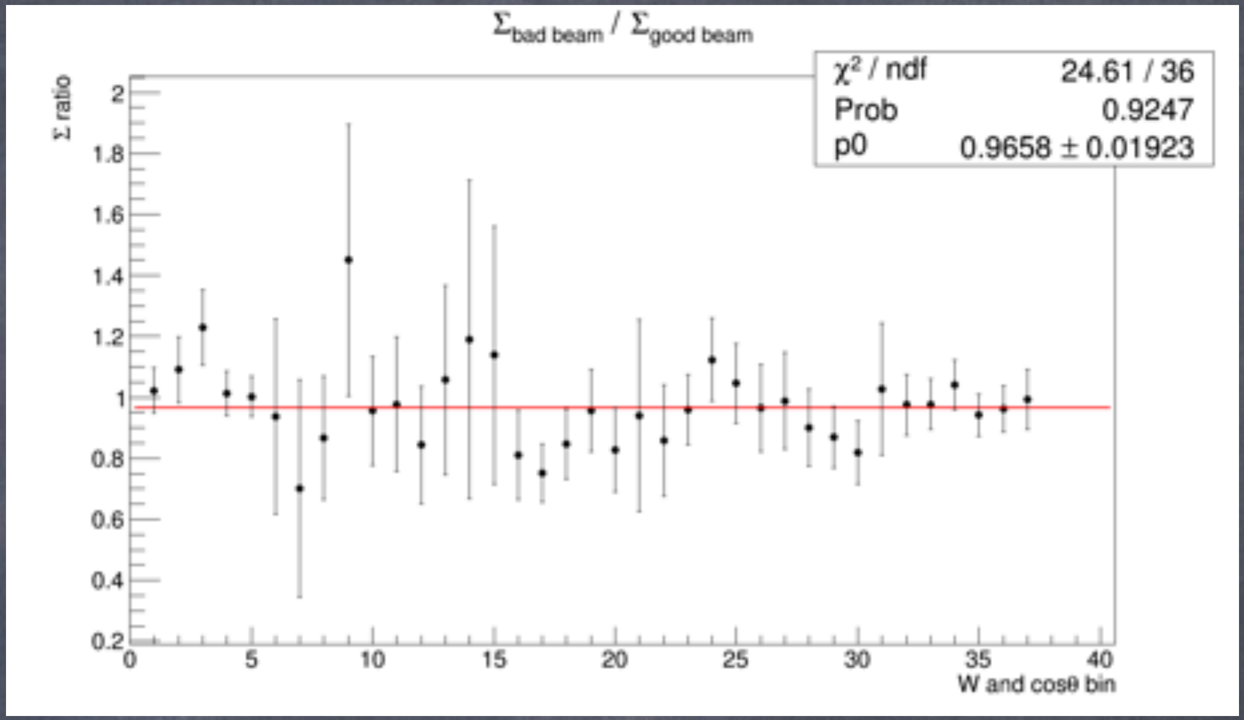


“Bad beam” / “good beam” comparison: II

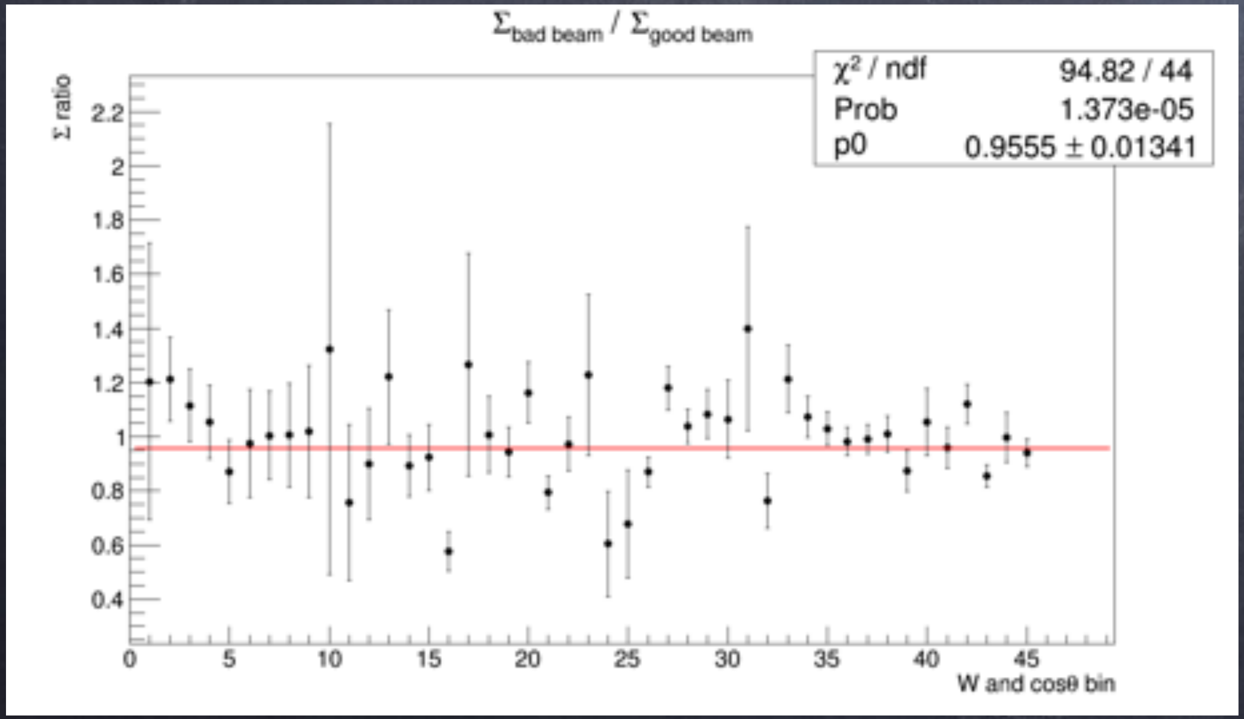
4.1/1.7 GeV



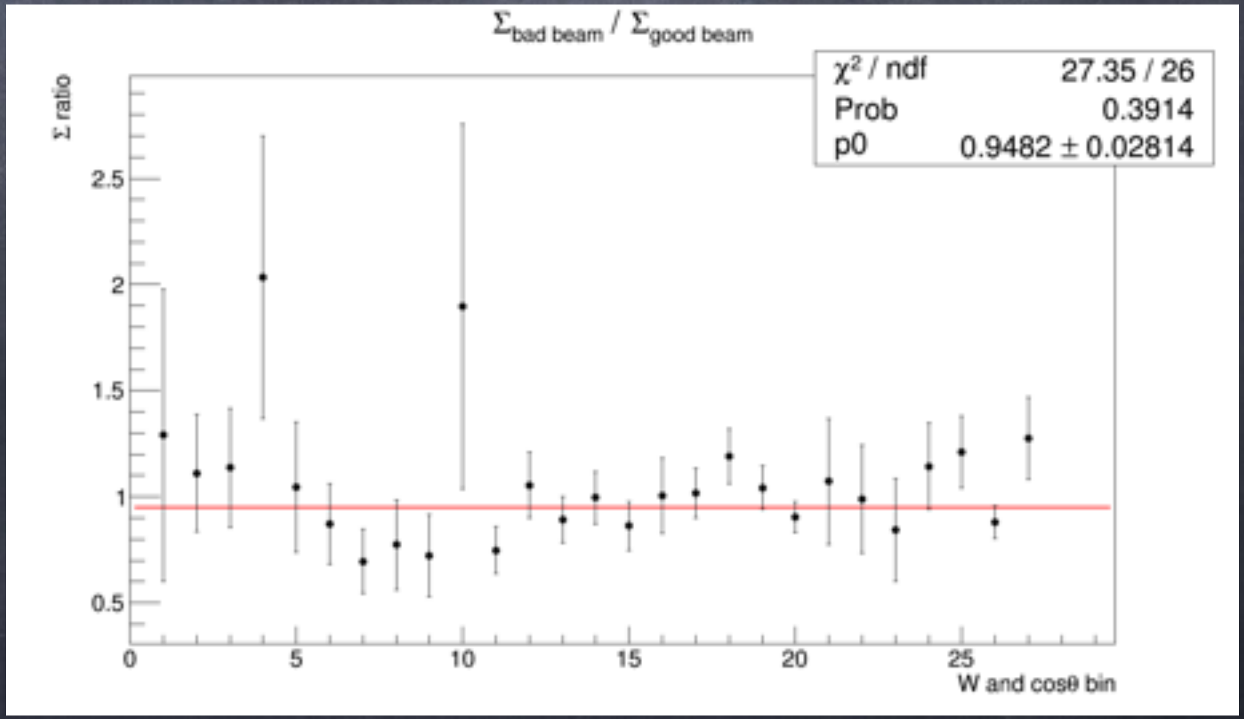
4.7/1.7 GeV



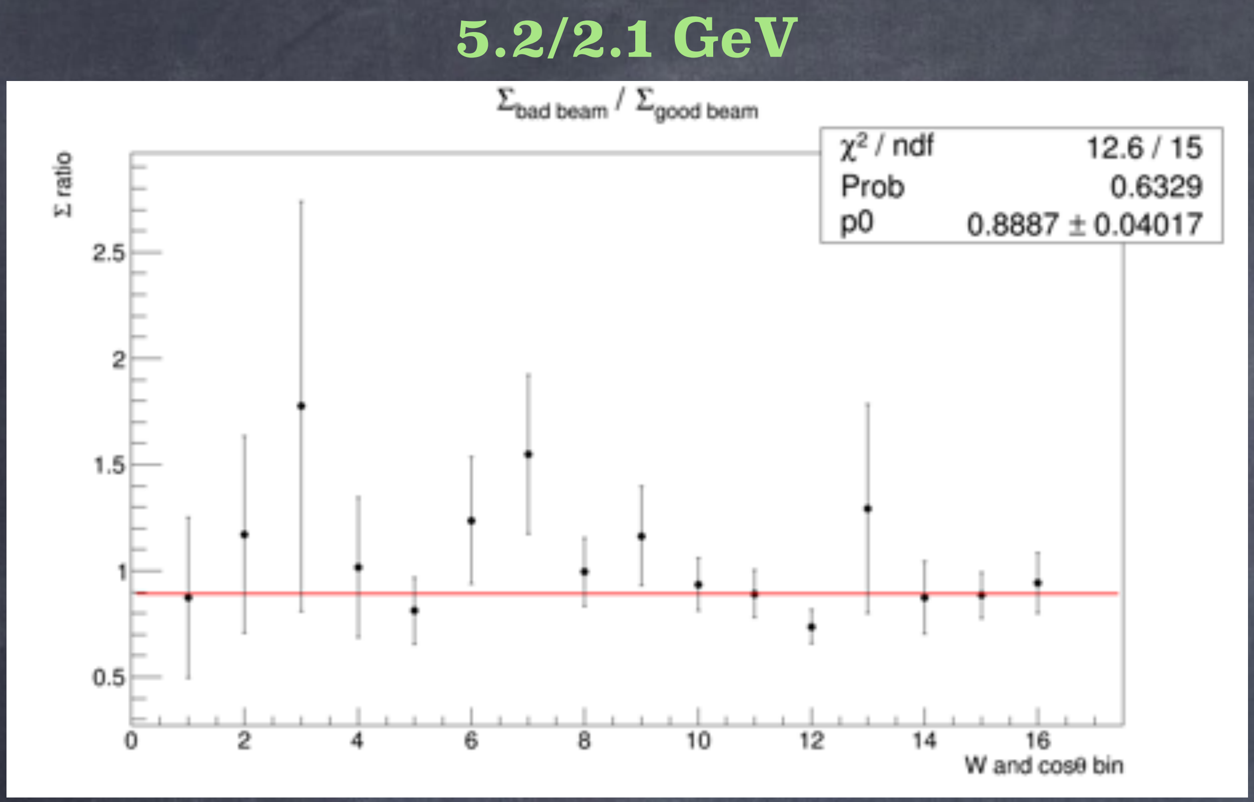
5.1/1.9 GeV

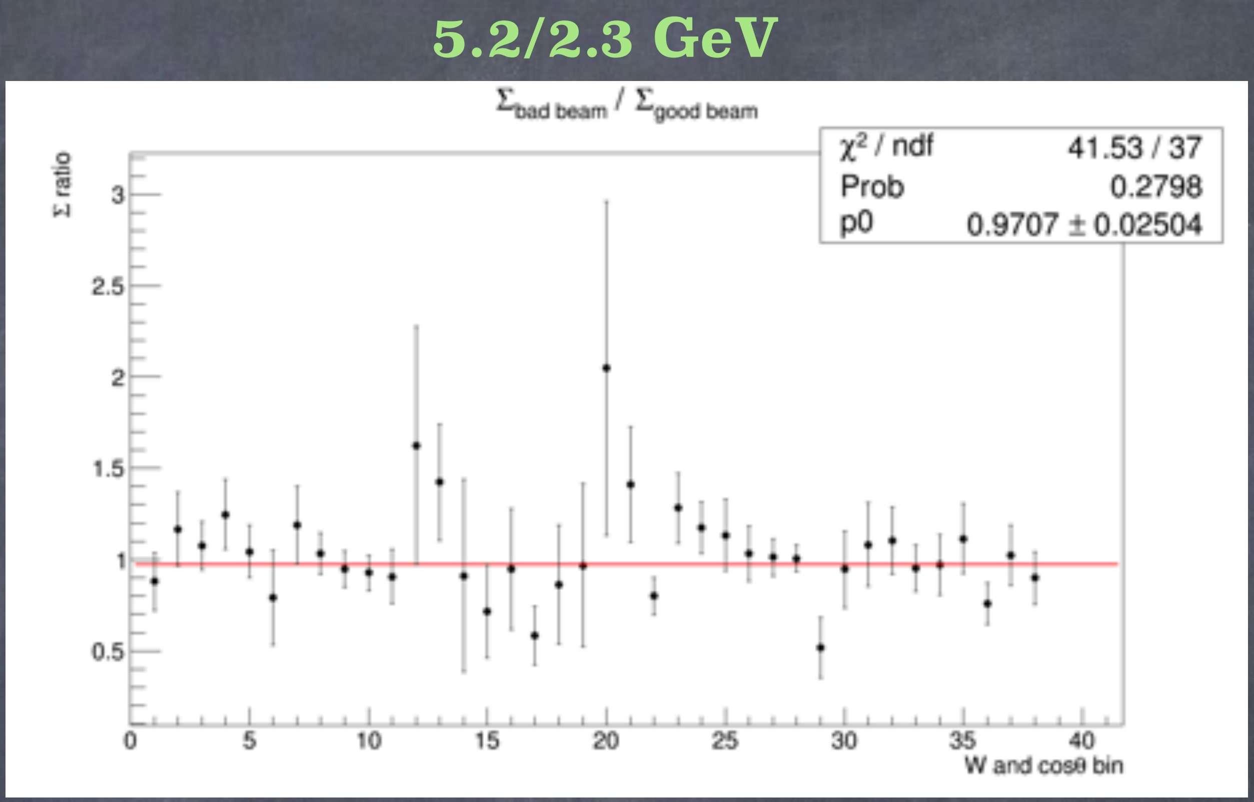


5.1/2.1 GeV



“Bad beam” / “good beam” comparison: III

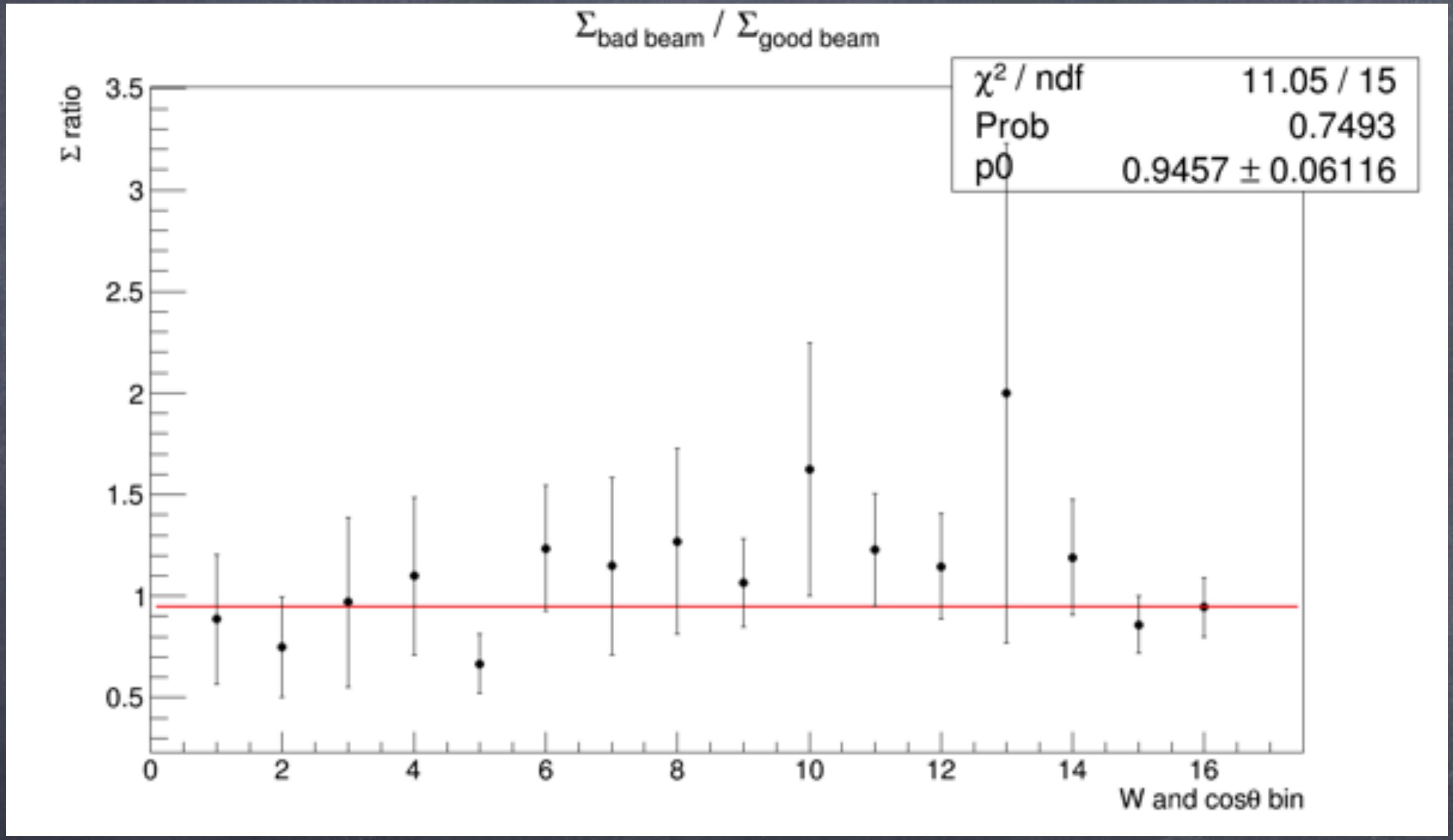




- A bit of a black art… would conclusions change if there were more data-points for some of the settings? Impossible to tell…
- Ultimately, will compare final results with trippy data included / excluded. Effect is likely to be well within the error…
- A final sanity check — see next slide.

Sanity check

- What if we take data we're reasonably confident in having the right polarisation (eg: 5.1/1.9 GeV setting after trip cuts), split it artificially into 9/10 and1/10 (by separating out every 10th good event which makes it past the cuts) and compare the sigmas extracted from the 9/10th and 1/10th of the data?



- This would pass the criteria for "good agreement".

Systematic error in polarisation

Systematics due to P

- These have the biggest effect on the value of Sigma.

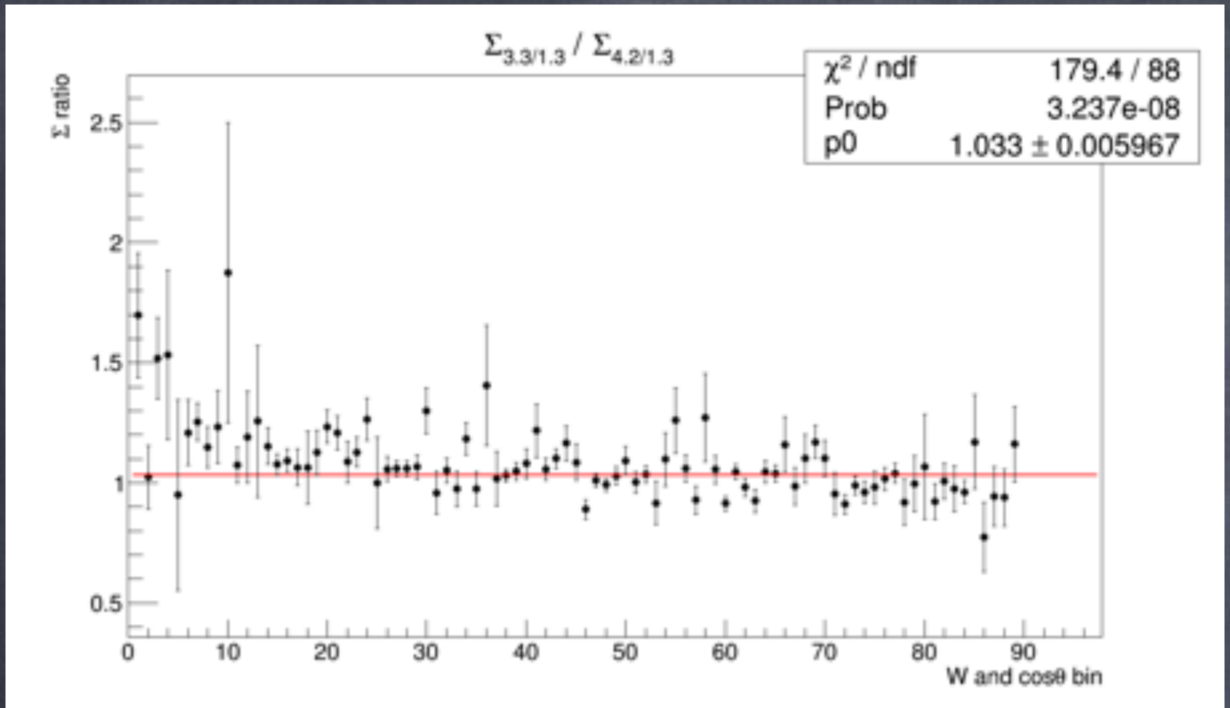
- *If we had a stationary target*, we could compare results from individual E-counters which came from settings with the same Ee but different nominal coherent edges. We could then be confident that average Sigma is the same (as E-counter would define W) and with enough statistics angular distributions would also be the same. Any difference in the result would then be due to the systematic error on polarisation.

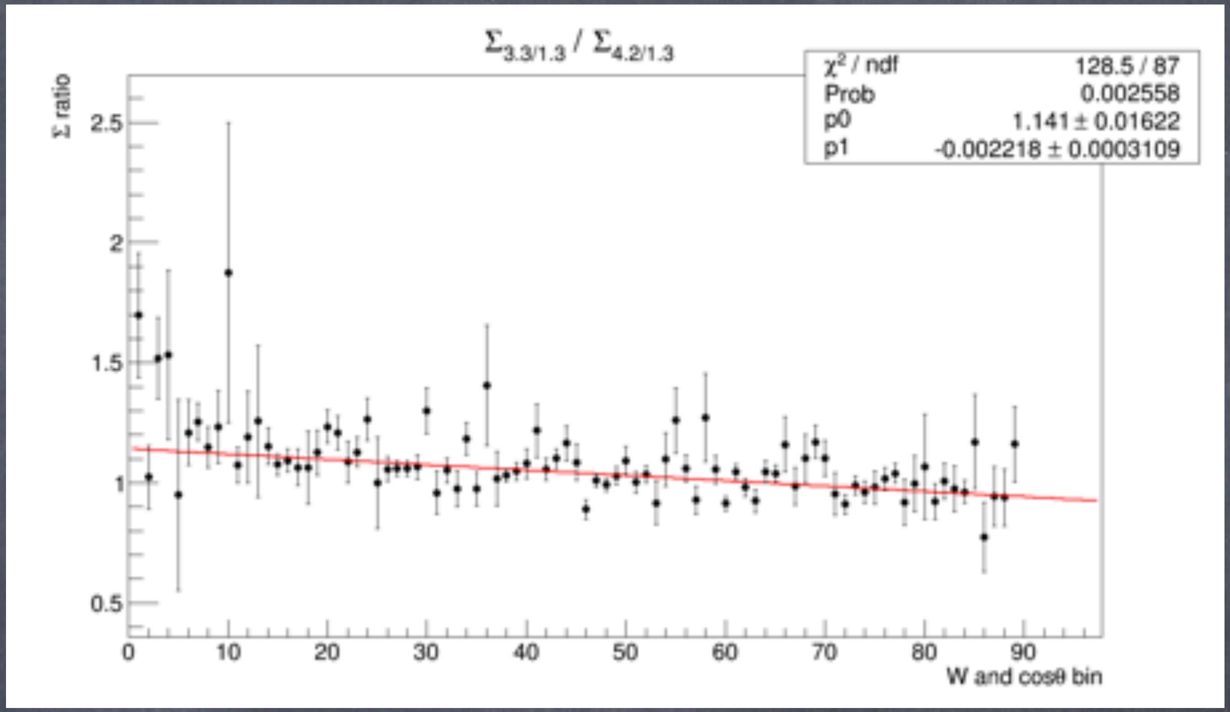
- No such luck with our target – Fermi motion is a bitch. Each E-counter has a whole range of data in W. So there's nothing to be gained by looking at individual ones, given the limited statistics per counter.

- A more promising tack is to bin as finely as I can in W and cos theta and compare the results from different settings. This does not provide a counter-by counter systematic, but an "average" one.

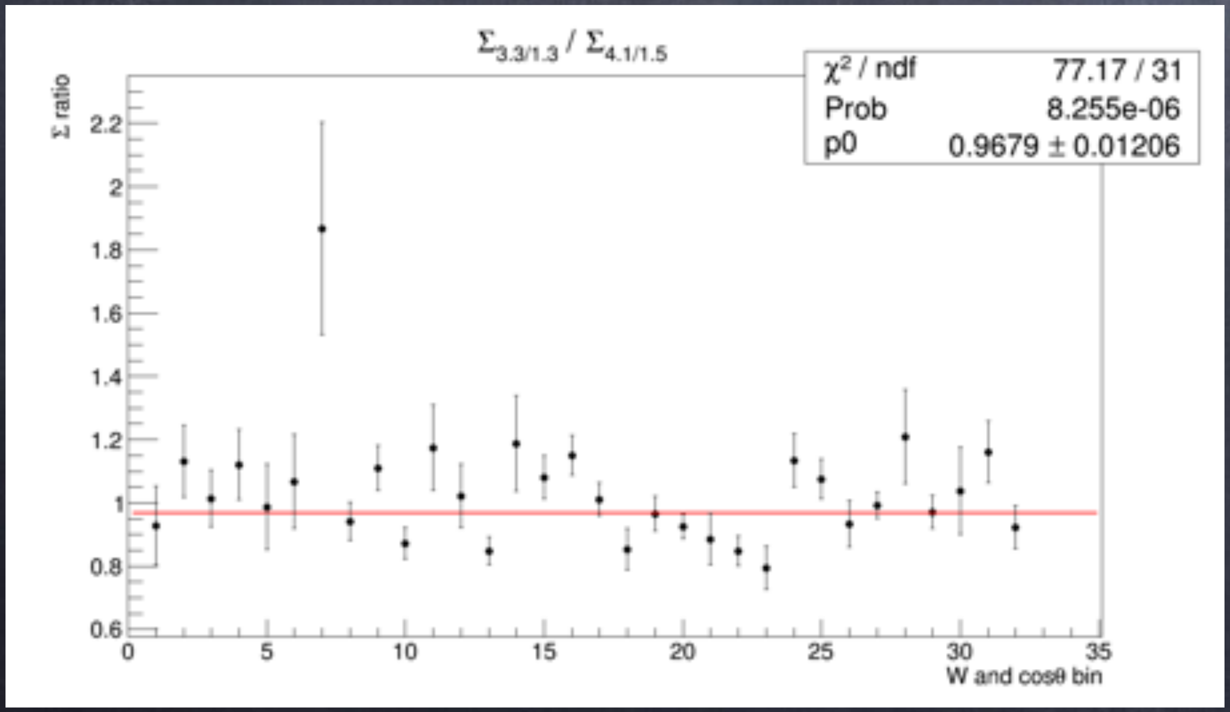
- Following slides: ratios of Sigma, binned 20 MeV in W, 0.1 in cos theta.

(3.3/1.3 GeV) / (4.2/1.3 GeV)

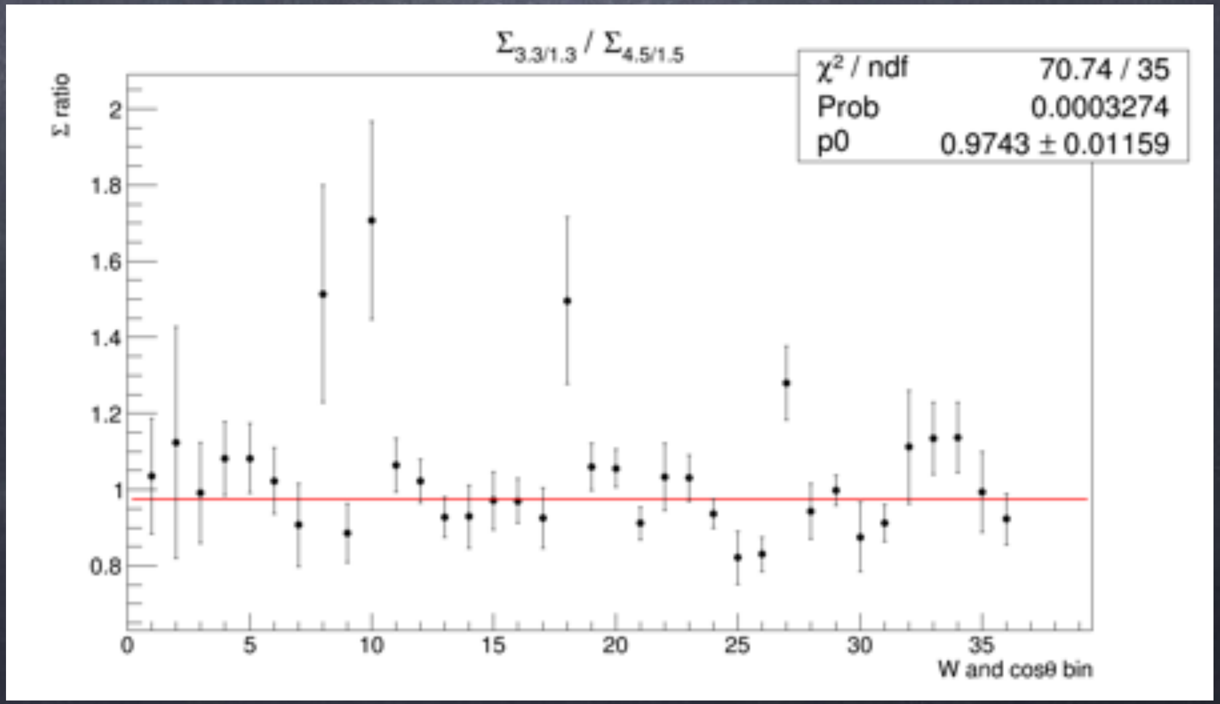




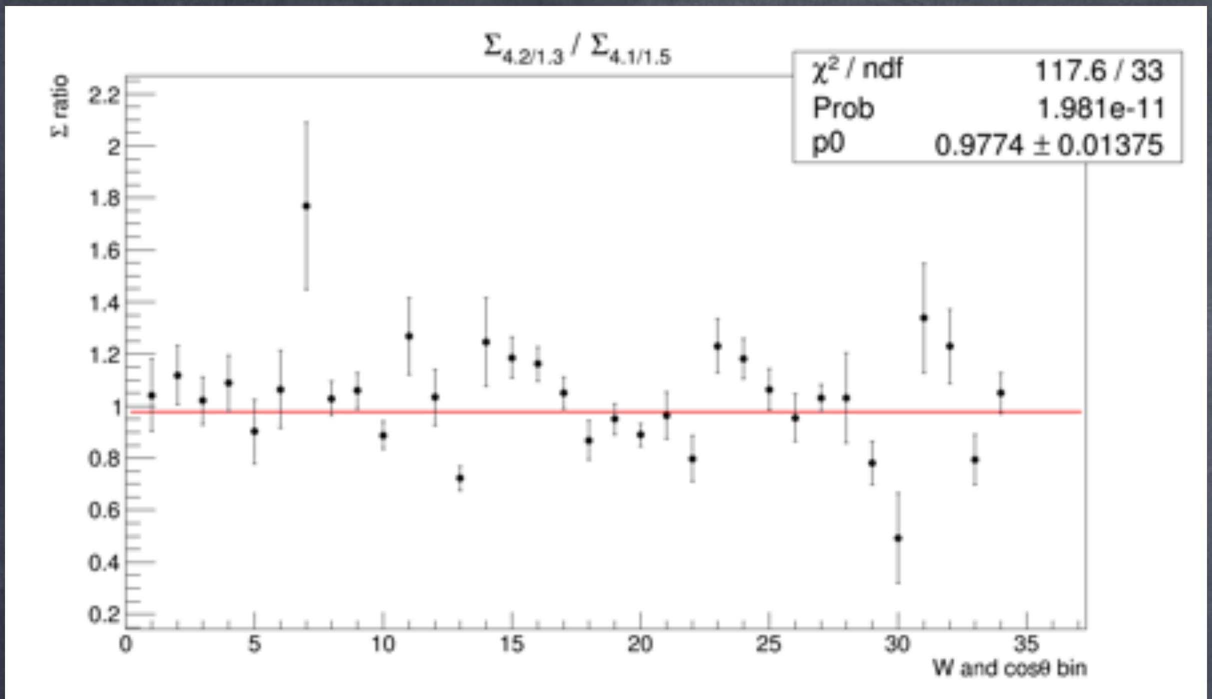
(3.3/1.3 GeV) / (4.1/1.5 GeV)



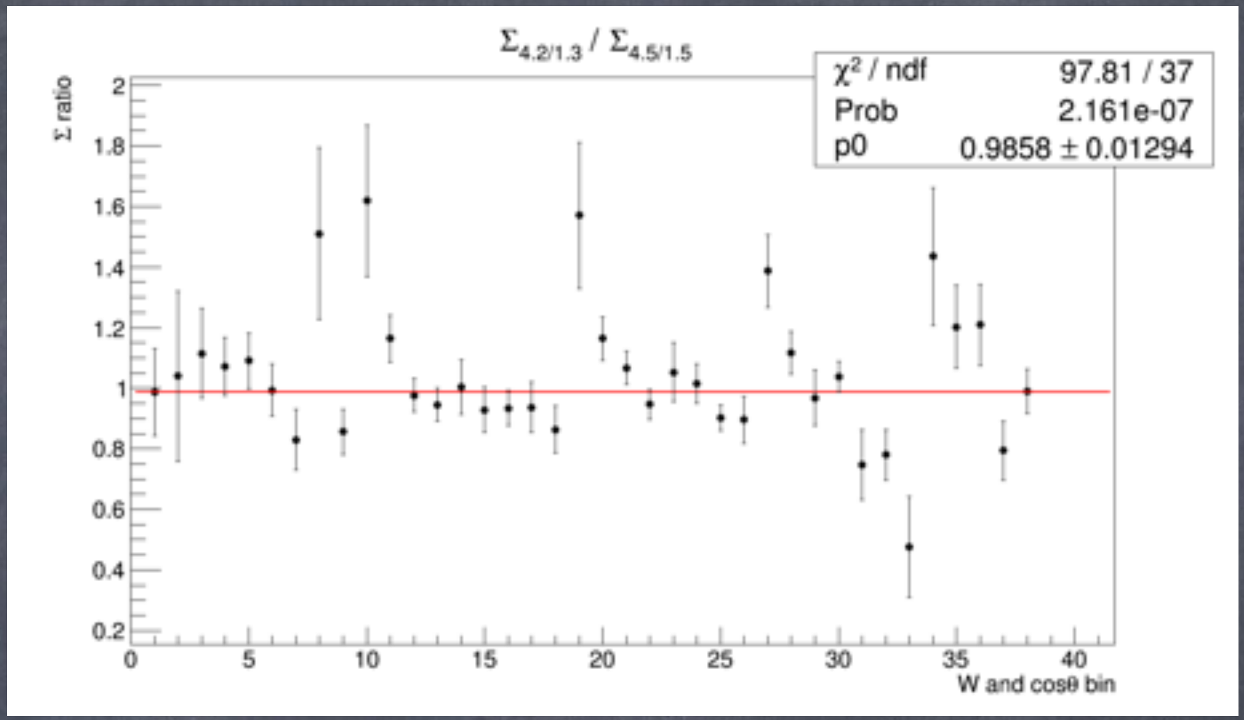
(3.3/1.3 GeV) / (4.5/1.5 GeV)



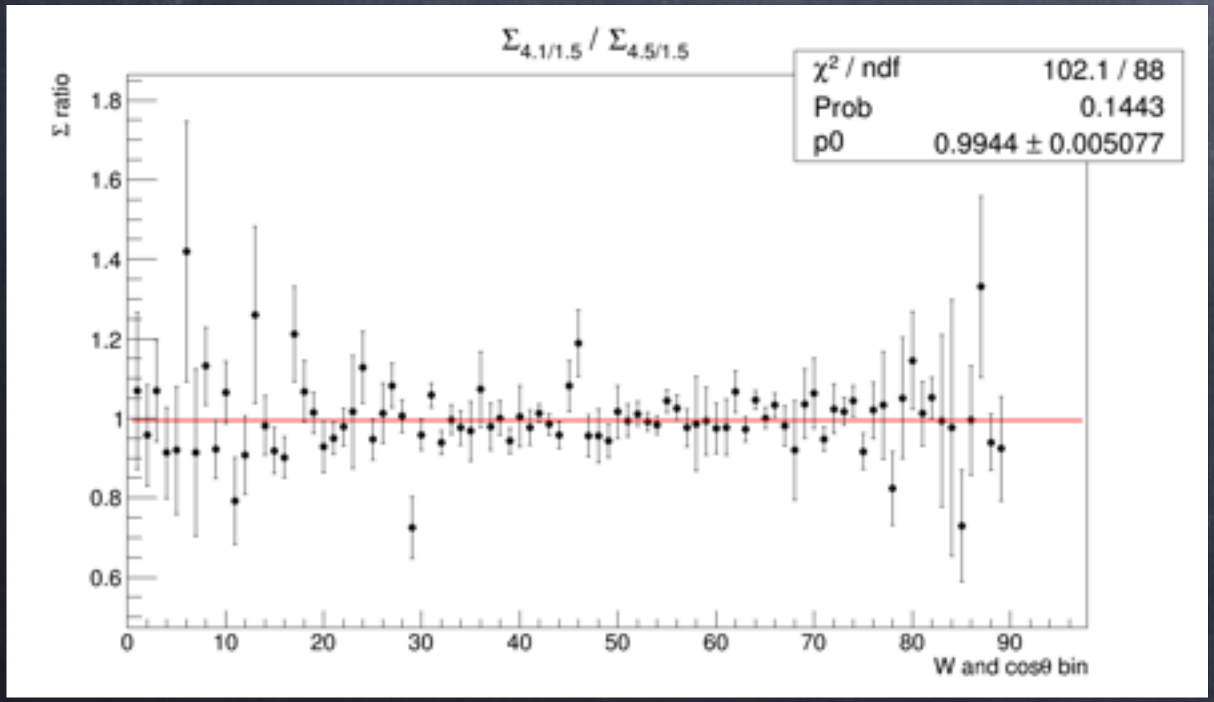
(4.2/1.3 GeV) / (4.1/1.5 GeV)



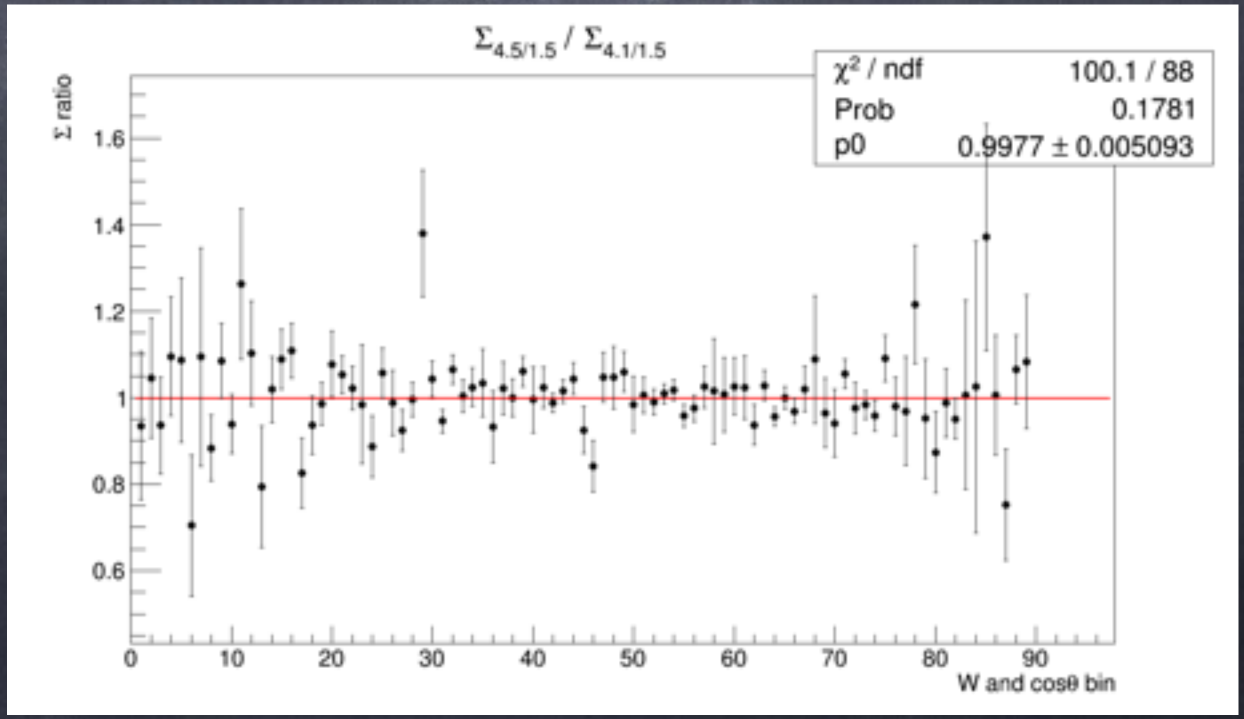
(4.2/1.3 GeV) / (4.5/1.5 GeV)



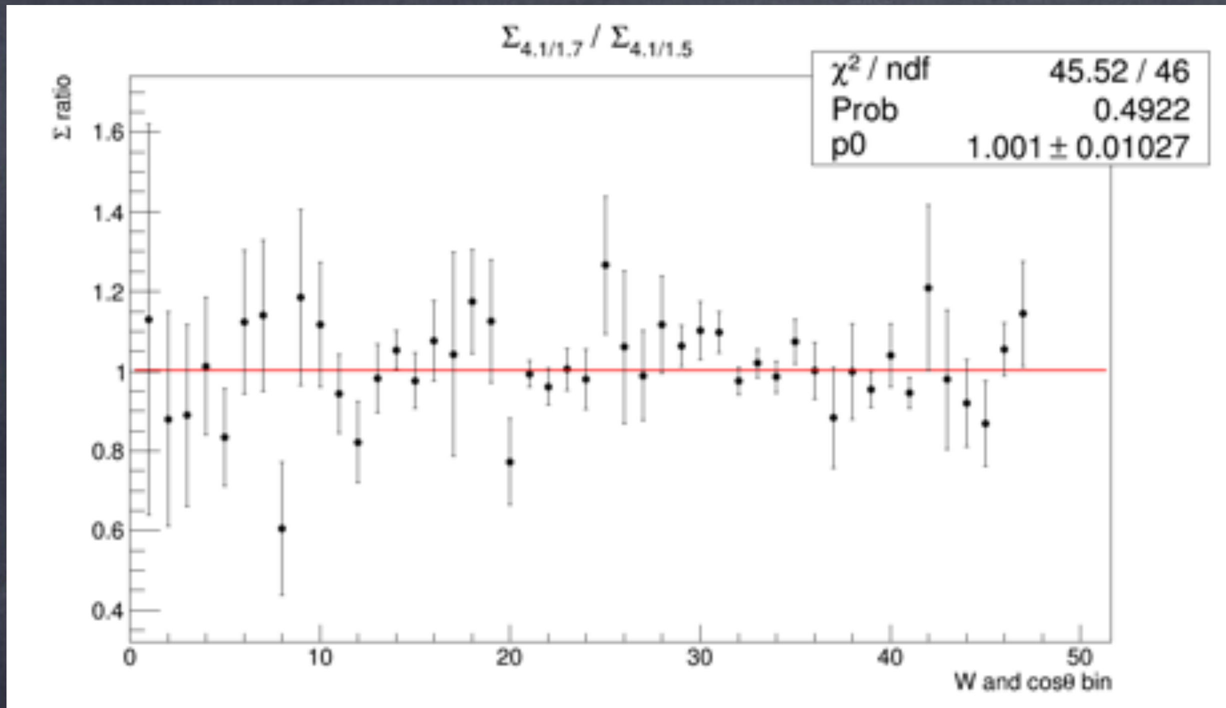
(4.1/1.5 GeV) / (4.5/1.5 GeV)



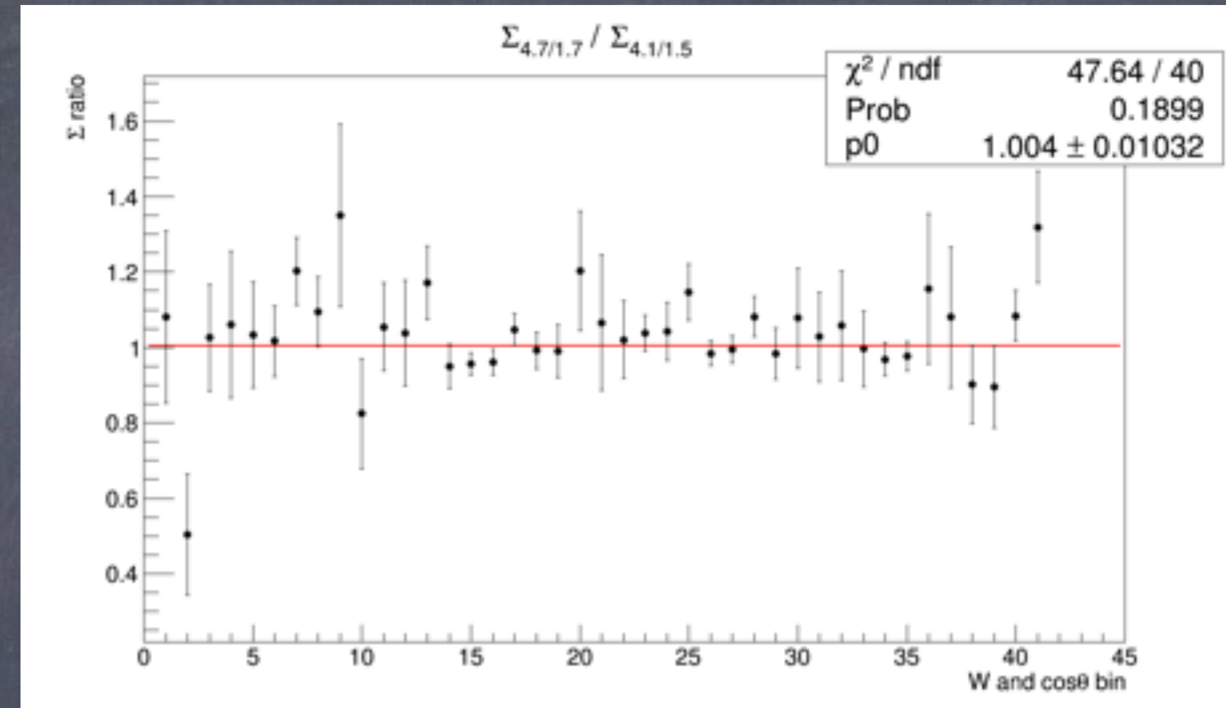
(4.5/1.5 GeV) / (4.1/1.5 GeV)



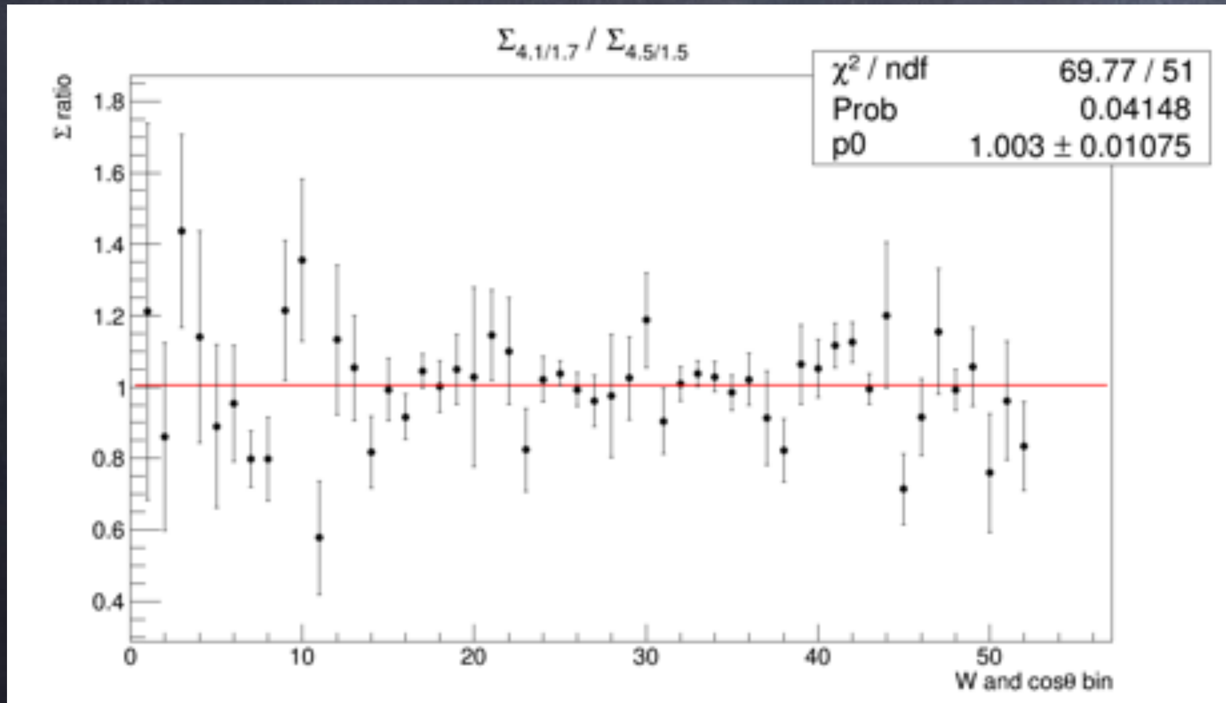
(4.1/1.7 GeV) / (4.1/1.5 GeV)



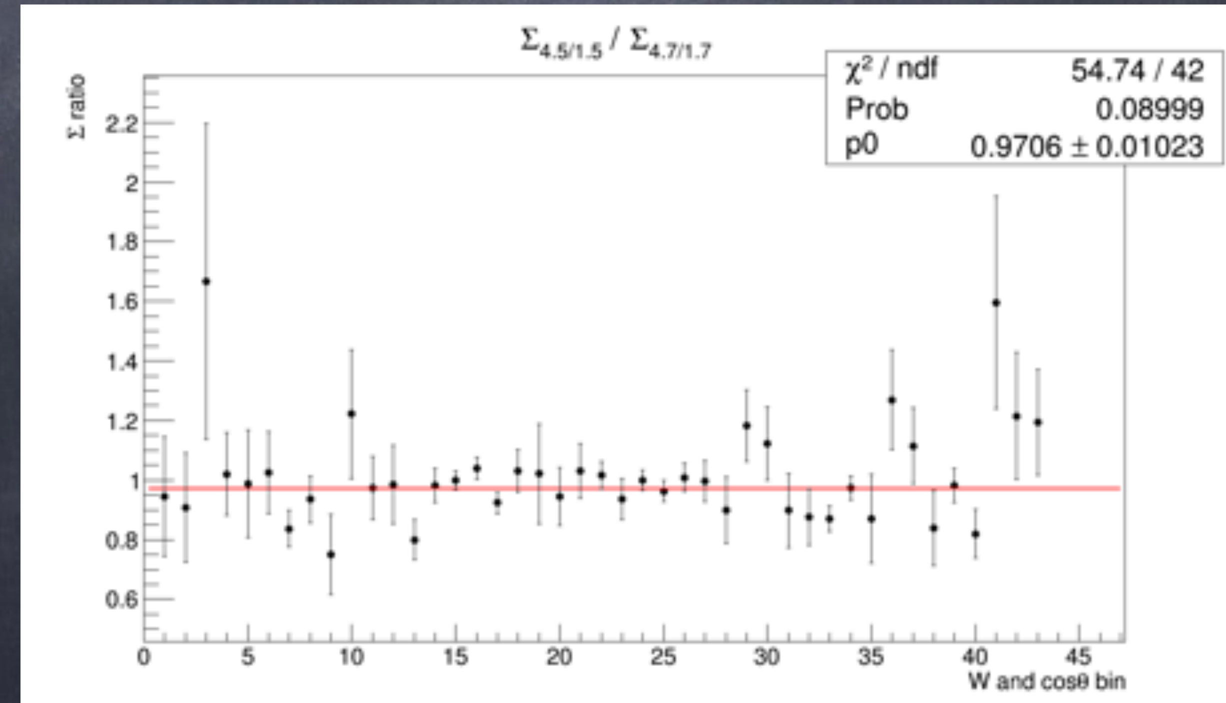
(4.7/1.7 GeV) / (4.1/1.5 GeV)



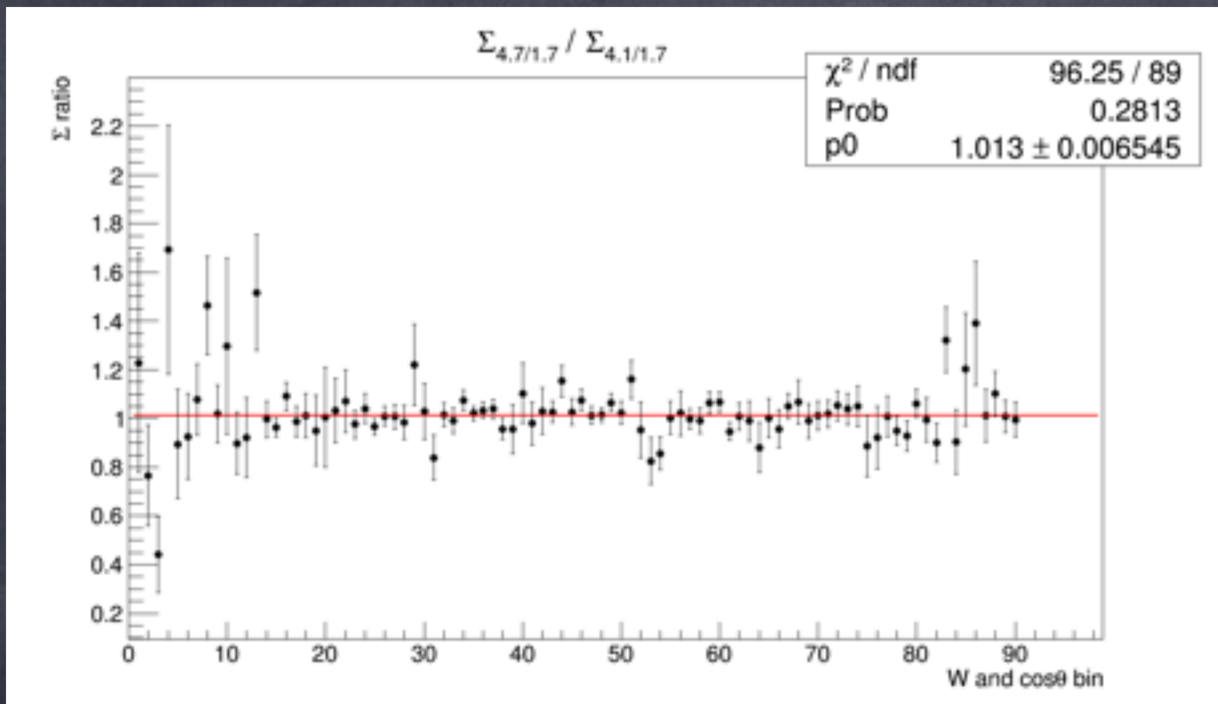
(4.1/1.7 GeV) / (4.5/1.5 GeV)



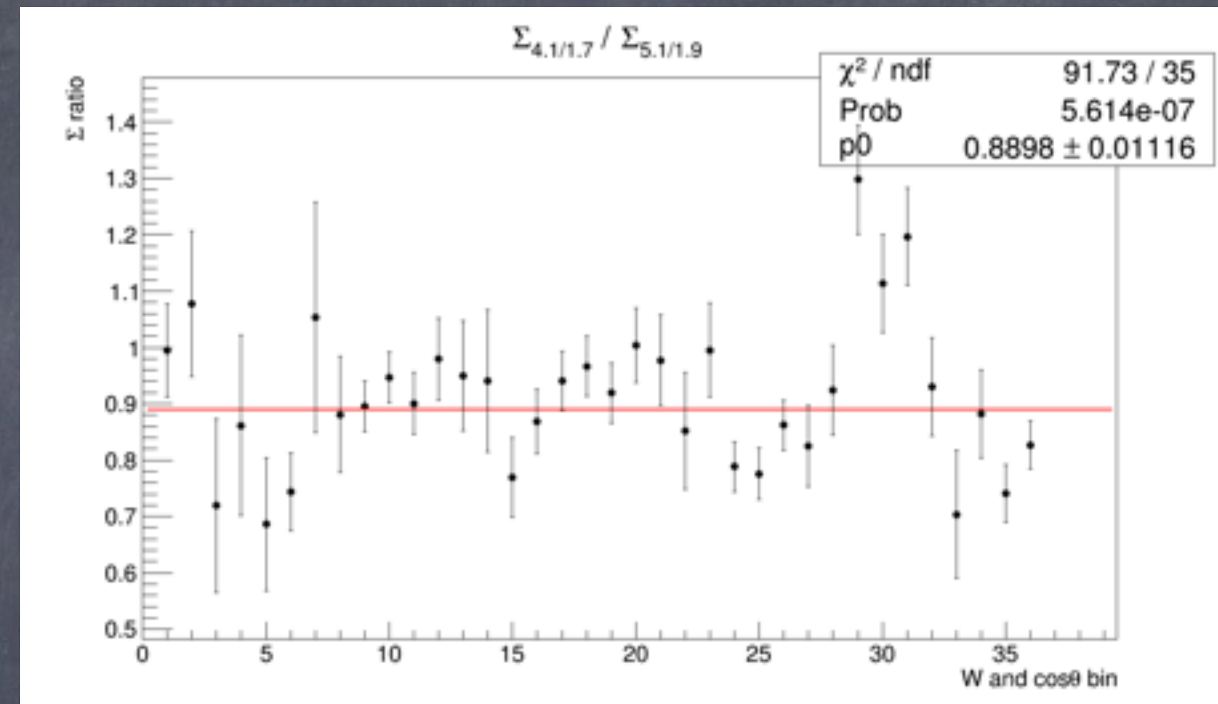
(4.5/1.5 GeV) / (4.7/1.7 GeV)



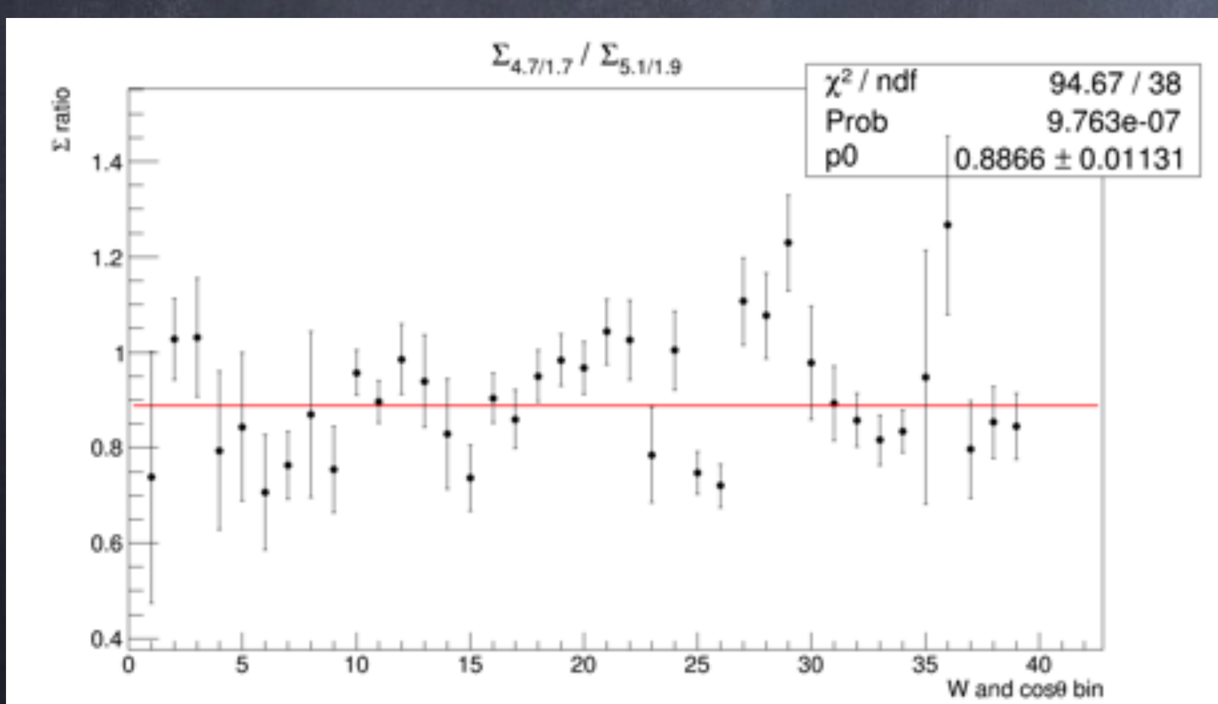
(4.7/1.7 GeV) / (4.1/1.7 GeV)



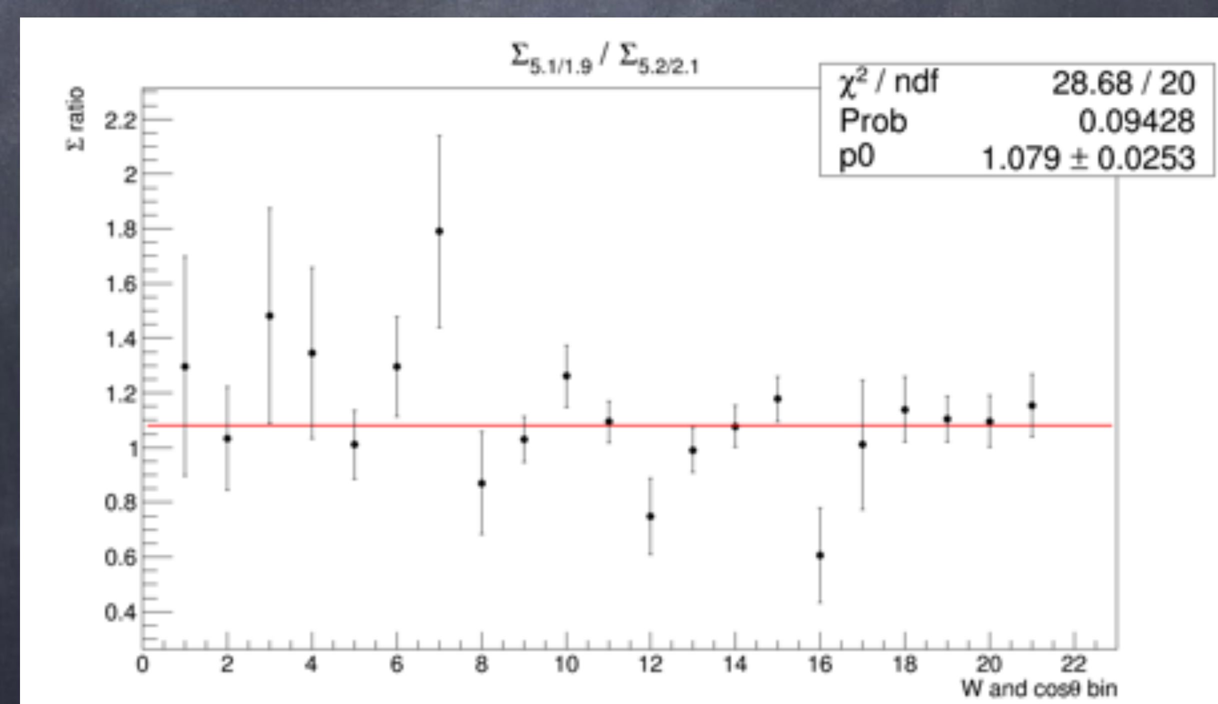
(4.1/1.7 GeV) / (5.1/1.9 GeV)



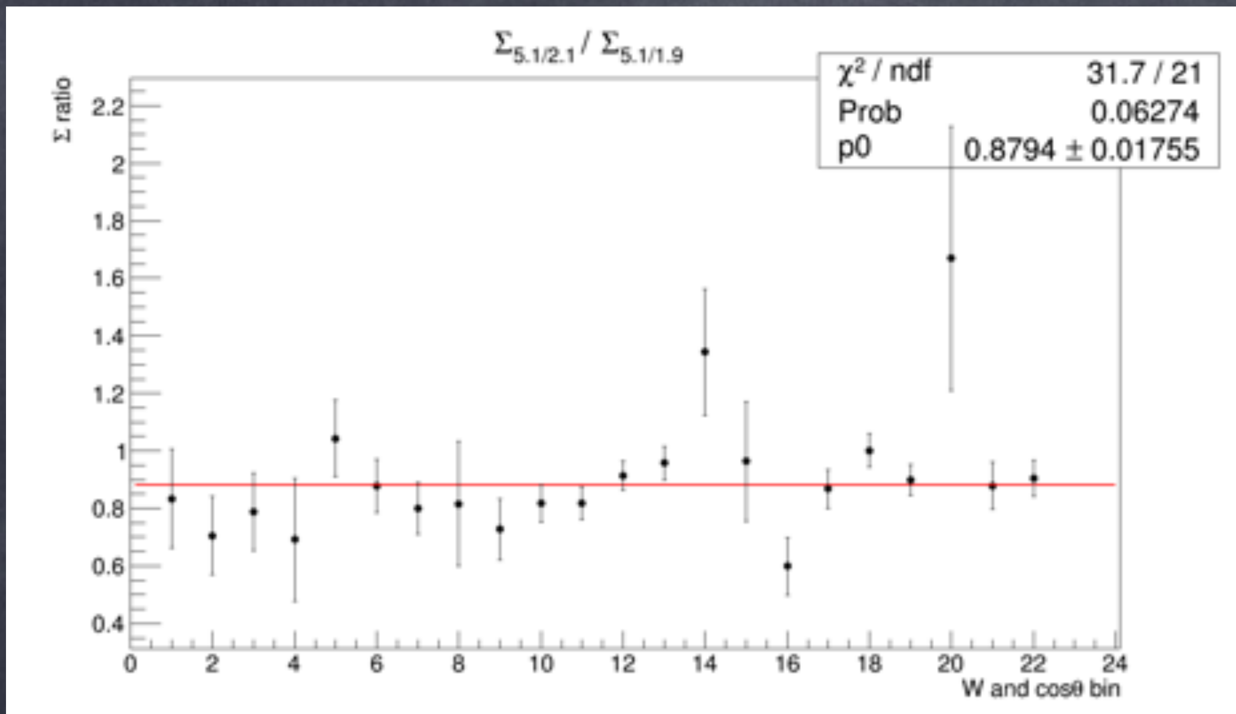
(4.7/1.7 GeV) / (5.1/1.9 GeV)



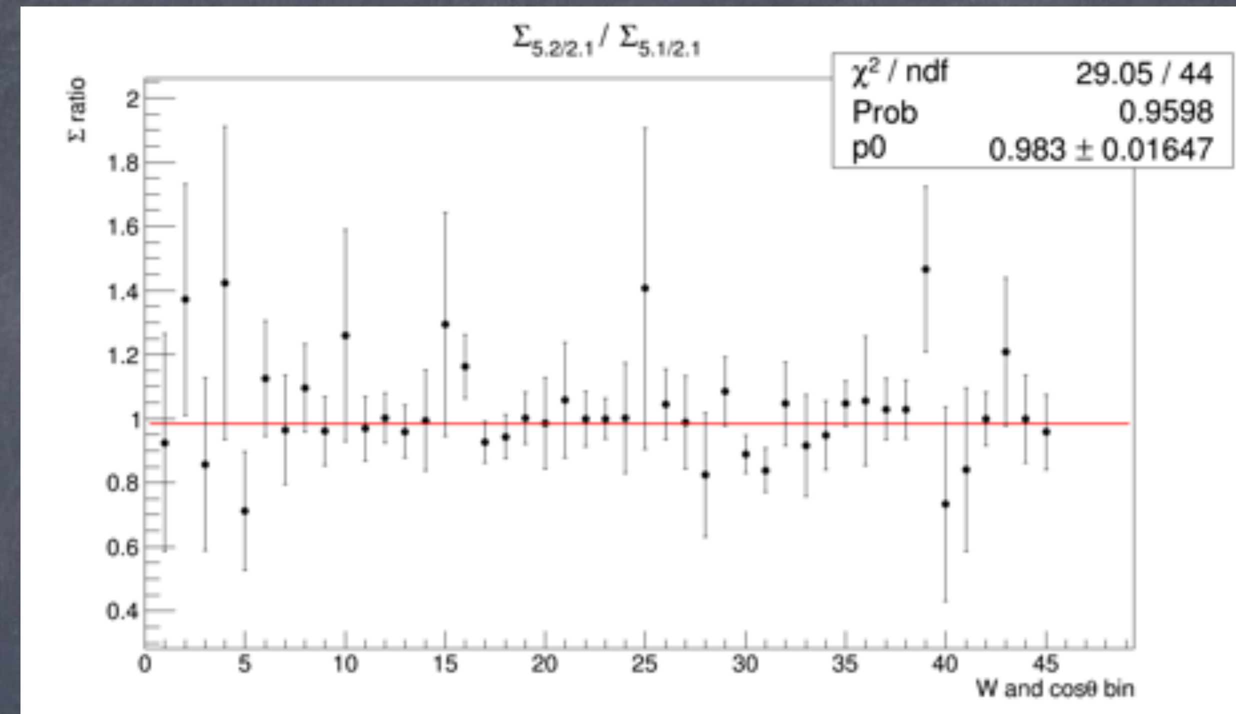
(5.1/1.9 GeV) / (5.2/2.1 GeV)



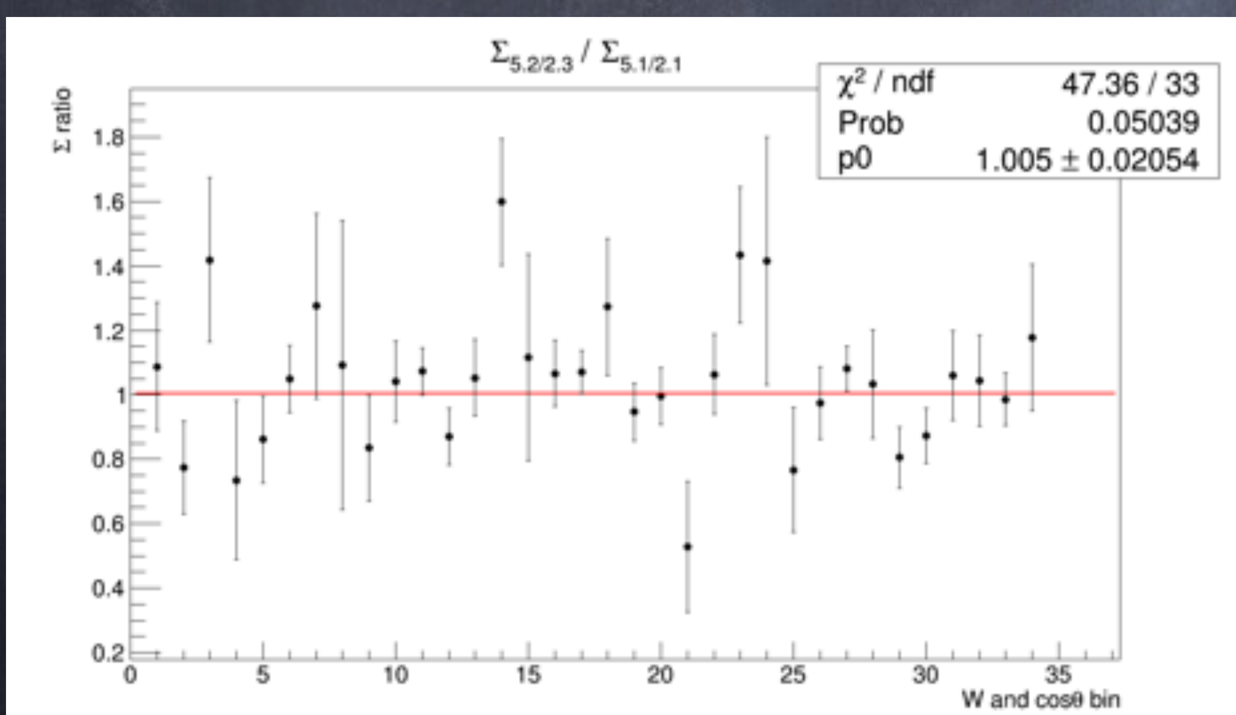
(5.1/2.1 GeV) / (5.1/1.9 GeV)



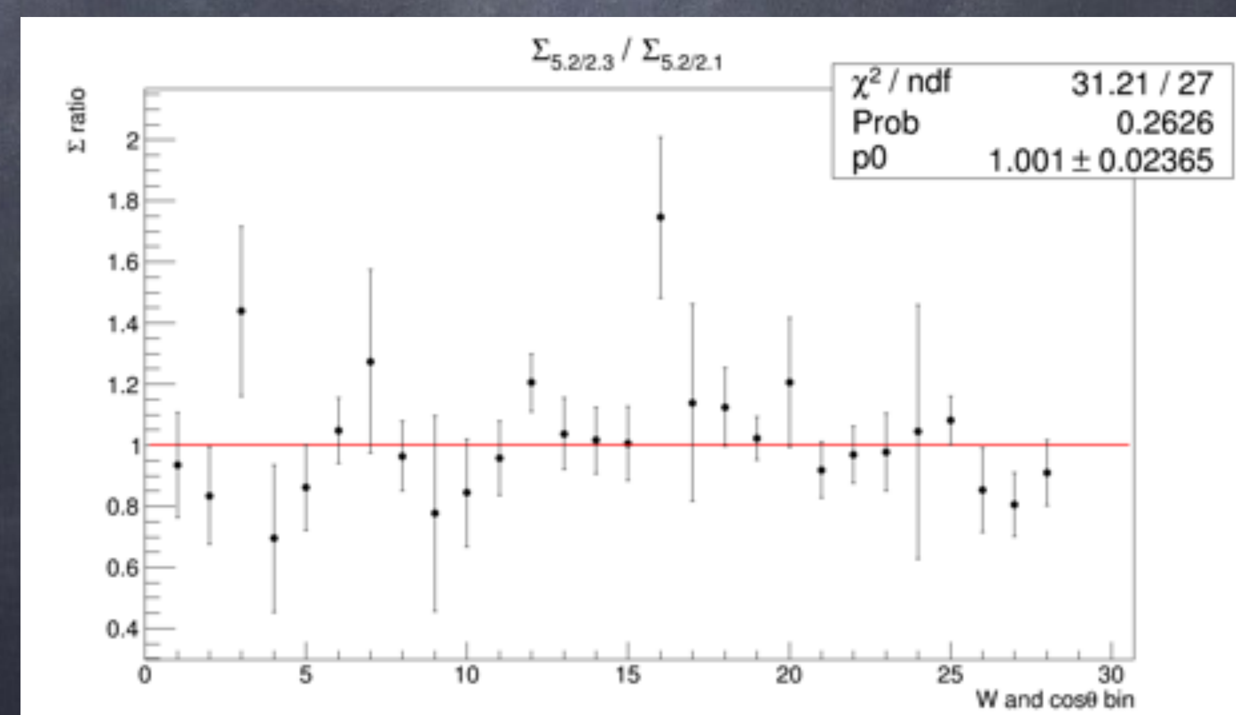
(5.2/2.1 GeV) / (5.1/2.1 GeV)



(5.2/2.3 GeV) / (5.1/2.1 GeV)



(5.2/2.3 GeV) / (5.2/2.1 GeV)



Systematics due to P

- For some setting combinations, get a perfect straight-line fit. Can use the result of the fit to estimate an average systematic error on P. Ranges from 0.1% - 12%.

- For some setting combinations the fit fails – it seems the systematic error on P in those cases depends on the energy.

- Perhaps ratios are not the best thing to check, as they would not be symmetric around 1. Differences in Sigma could be a better way of checking whether the results from different settings agree, but would be useless as extracting a systematic on P...

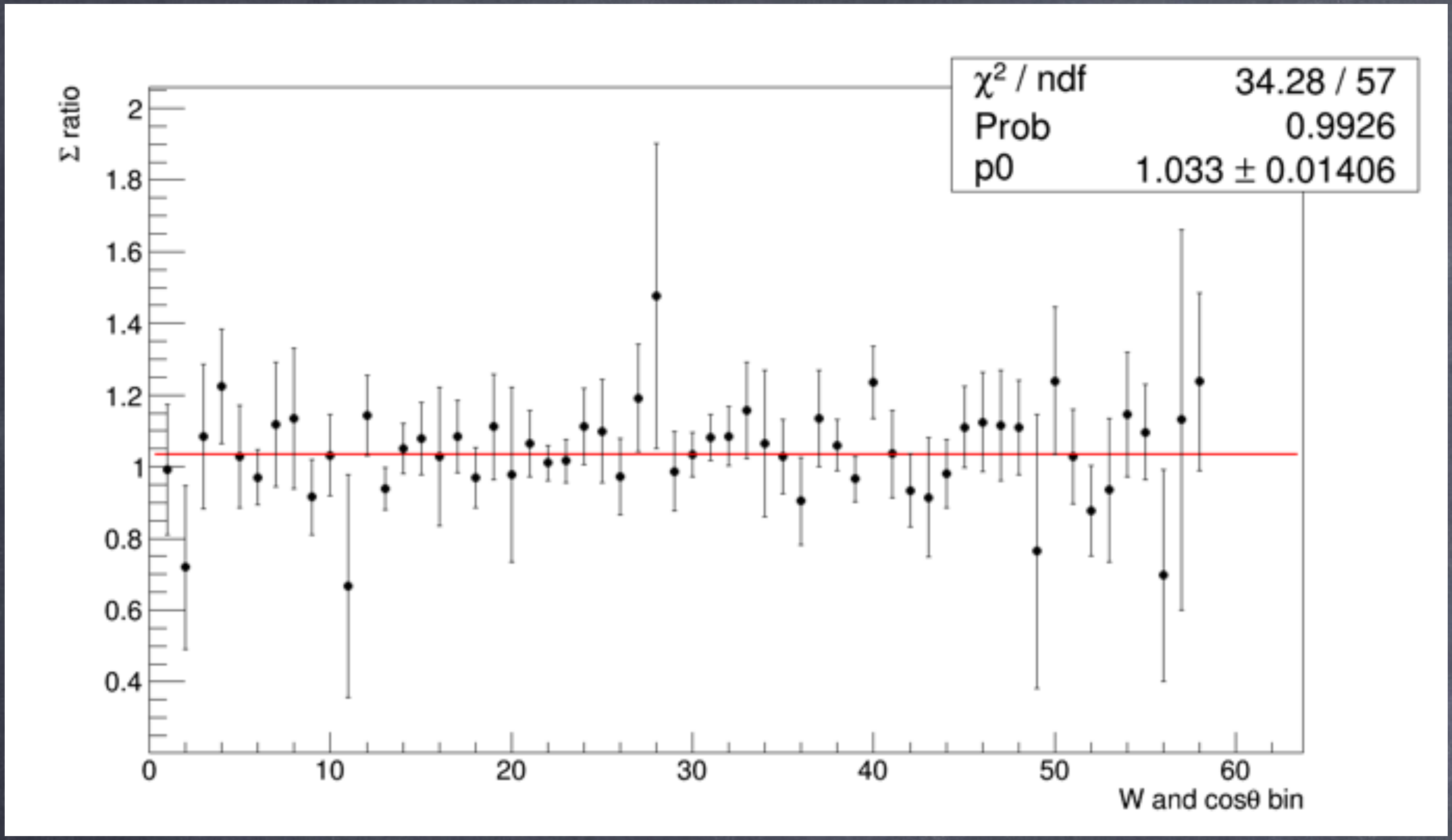
- Plan is to fit to all the data points from the graphs together and take the systematic P error as the average from the fit + 3 sigma.

Keep the 3.9/1.3 GeV setting or discard it?

What about the 3.9/1.3 GeV setting?

- A dodgy one. 5.5 times more PERP events than PARA, so a Pol Table was never produced for the PARA events.
- Could either include the PERP (~3.2M events) with the other settings or ditch it.
- Decide by comparing Sigma (binned in 20 MeV (W), 0.1 (cos theta) from:
- A), half of PARA events from 3.3/1.3 and 4.3/1.3 (all the odd events, say) and all the PERP data from the same settings.
- B), the other half of PARA events from 3.3/1.3 and 4.3/1.3 (all the even events) and the PERP data from 3.9/1.3.
- See overleaf...

Comparison of 3.9/1.3 GeV and the 3.3/1.3, 4.2/1.3 GeV settings



Ratio of Sigma obtained from set A / Sigma from set B (see previous slide for details)

- Consistent within 3 sigma...
- Conclude that the data-sets can be combined.

Conclusions for the data

- Trippy data included, 3.9/1.3 setting included.

- The variable binning method was used for phi, where a maximum number of bins was sought as long as each bin had at least 10 events and was at least 2 deg wide. The rule was motivated by not binning below the phi resolution and keeping the statistics-related systematics to a minimum.

- Cases where the number of bins was < 13 were discarded. Six sectors, six coils, phi = 0 is in the middle of a sector: 13 bins min.

- Cases where the total number of events was smaller than an asymmetry-dependent boundary (800 - 1000) value were discarded.

Results

The fit function

$$\frac{\sigma_\perp - \sigma_\parallel}{\sigma_\perp + \sigma_\parallel} = \frac{(N_R - 1) + \frac{N_R P_R + 1}{P_R + 1} 2 \bar{P} \Sigma \cos(2(\phi + \phi_0))}{(N_R + 1) + \frac{N_R P_R - 1}{P_R + 1} 2 \bar{P} \Sigma \cos(2(\phi + \phi_0))}$$

- Phi_0 set to 0 in the fit, on the basis of my study.
- Pr fixed from Pol Table calculations by N. Zachariou.
- Nr allowed to vary. Typically returns 0.8 - 1.2. For some settings it returned 2.
- Sigma extracted from fit, TMinuit2.
- Tried the Log Likelihood option, but does not work – zero events bins mess up the fit royally, so have to rely on chi squared.

Can we justify adding all settings together?

- No a priori knowledge about which settings have the more accurate P, only an average value of the systematic.
- So sum across the settings to improve the statistical error on the data-points.
- Final results over-leaf... next week