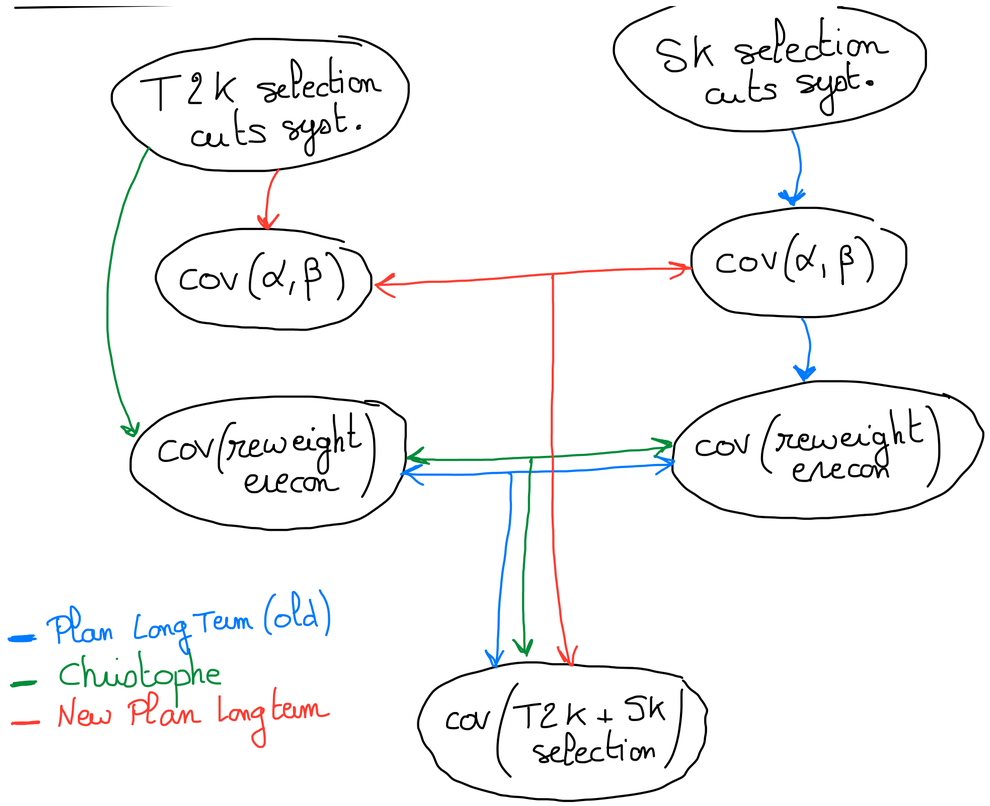
Towards estimating detector systematic uncertainties for T2K_SK Joint analysis

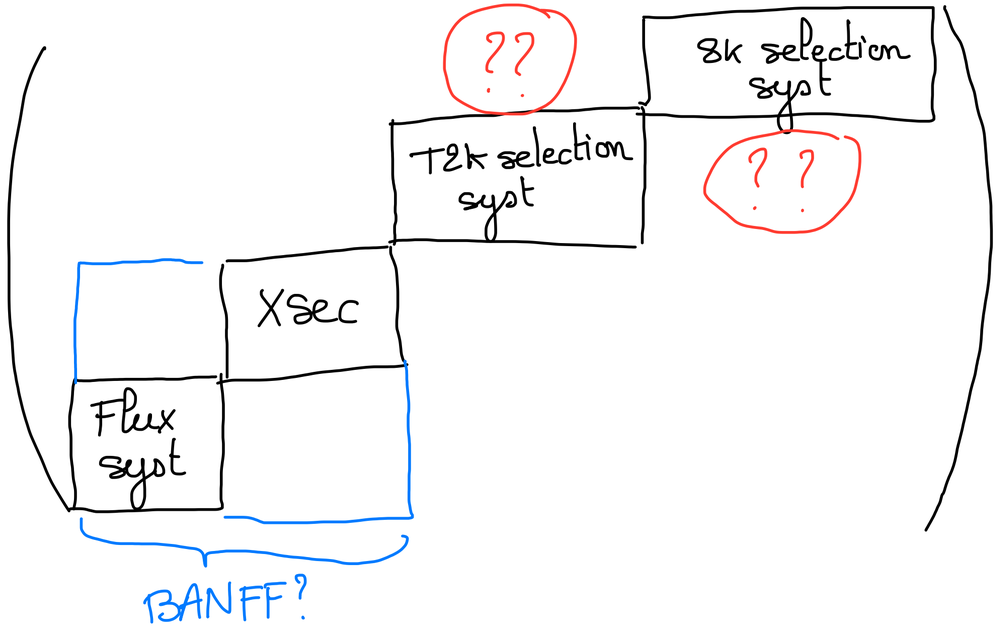
Outline

Context
Method
Finding the cuts
Discovering and implementing MCMC
Performance tools
Results
What's next ?

Lucile Mellet _ neutrino group meeting _ 02/12/2020

Context

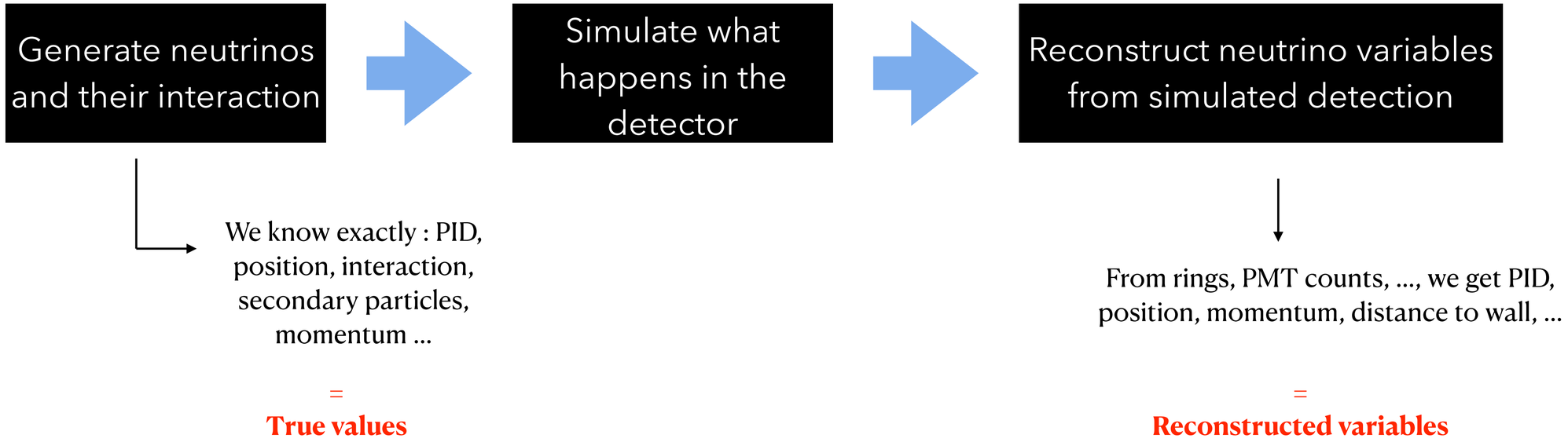




Thank you Mathieu for this beautiful scheme ;)

Method

SK MonteCarlo that we have so far



From this we can build samples by categorizing events (= making a selection based on cuts on variables) :

- (**True topology** samples : categorizing based on **True values)**
- **Reconstructed** samples : categorizing based on **reconstructed variables**

Method

General idea : use MC/ data comparisons

Goal : generate response functions (RF) for each parameter and each sample to give as an input to P-theta software for systematics treatment

RF: distribution of the impact of a parameter on the sample distribution (= weights) as a function of the value of the parameter

—> the parameter values will then be thrown in those distributions and given the correct weight to proceed to marginalization

Procedure to do for : all syst.param on which there is a cut, all reconstructed samples and all true topology (to take into account mis-categorizations due to those parameters) :

1. Take the data distributions (as a funct. of the studied param) and the cuts as fixed

2. Vary the MC distributions with process of smearing (multiplicative factor alpha) and shifting (additive factor beta) —> =Fake data until we have real SK data

3. Retrieve a reasonable (based on Likelihoods between (shifted ans smeared)MC and (fake)data) range of alphas and betas to test

4. Test all those alphas and betas on the general distribution (as a function of momentum or energy) and get the impact —> Build RF

Method : Do steps 2 and 3 with a MonteCarlo algorithm for efficiency purpose and to avoid to assume gaussian distributions for the parameter and choose a 1sigma around best value for instance

Also: check correlations

Method

Keeping it simple for now

No true topologies taken into account

No data -> we vary with alpha/beta the nominal MC

Only on one and then two cuts

Not all samples

Finding the cuts

Need to rebuild samples after shifting and smearing to account for loss/gain of events in sample categories

« guess » them together with Adrien and Lukas

Didn’t get all continuous variables

Able to rebuilt all 8 single-ring (reconstructed) samples with exact same nb of events as « ATMPDEventType » SK variable

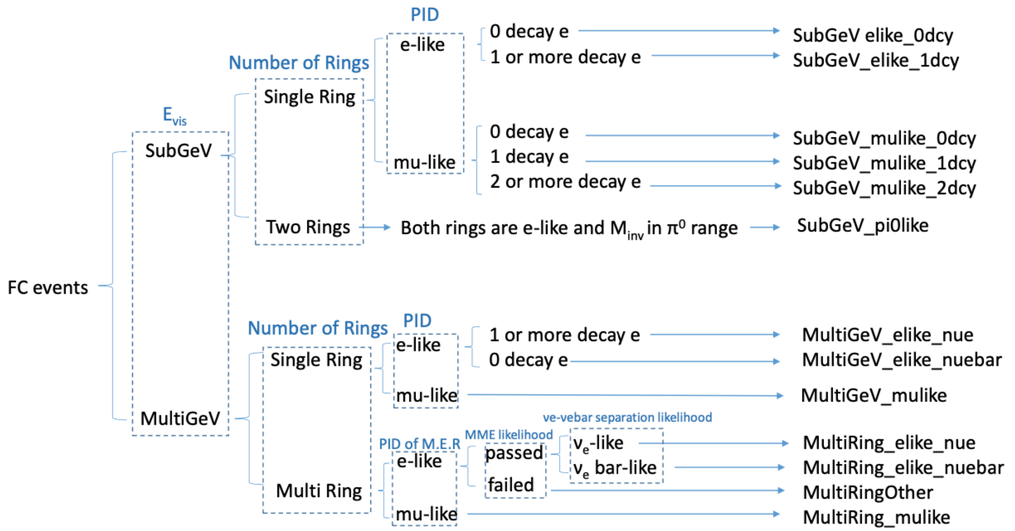


FIGURE 7.1: Summary of event categories for FC samples. The texts in the dashed line boxes shows the result classified by the variables which is on the boxes.

Miao’s thesis

Discovering and implementing MCMC

SubGeV_elike_0dcy	1
SubGeV_elike_1dcy,	2
SubGeV_mulike_0dcy,	4
SubGeV_mulike_1dcy,	5
SubGeV_mulike_2dcy,	6
MultiGeV_elike_nue,	8
MultiGeV_elike_nuebar,	9
MultiGeV_mulike,	10

For each sample:

1. **Cut flow—> fqwall histo**
2. **Loop on MCMC tests/throws (100000)**
    1. **Independent Random picking of A and B in Gaussian priors around A= 1 and B=0**
    2. **Cut flow —> A*fqwall+B histo**
    3. **Poissonian LogLikelihood per bin + Sum**
    4. $proba = min(1, e^{(LLtot - LL[j-1])})$ **—> Metropolis-Hastings**
    5. **if (p <= acc) —> Accept. —> Random**
    6. **Update the prior for next test**
3. **We get a alpha and beta distribution vs Likelihood**

Note : this could allow to study correlations between variables inside a sample but not between samples (should do it with all samples at the same time for this)

Performance tools

MCMC Development: First with gaussians then without doing the cuts

$$\rho_{lag} = \frac{\sum_i^{N-lag}(\theta_i - \bar{\theta})(\theta_{i+lag} - \bar{\theta})}{\sum_i^N (\theta_i - \bar{\theta})^2}$$

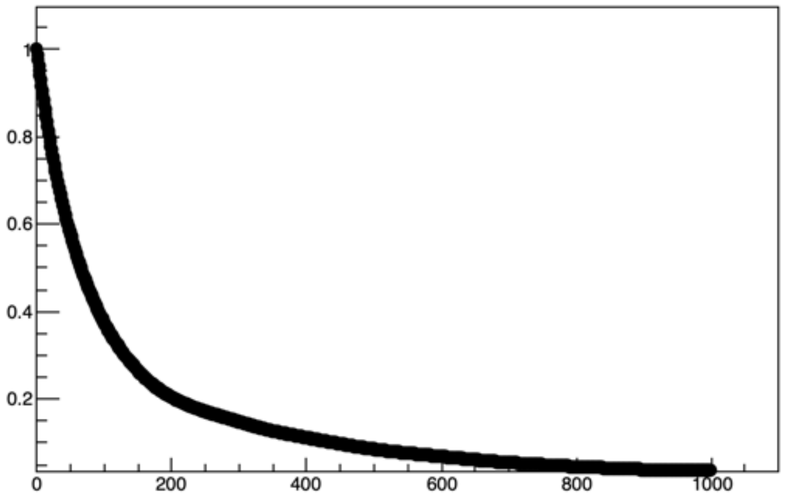
Running time !

Acceptance rate

Autocorrelation ($\rho(lag)$)

Neff: effective number of samples (=sampling of alpha or beta)

$$N_{\text{eff}} = \frac{N}{\sum_{t=-\infty}^{\infty} \rho_t} = \frac{N}{1 + 2\sum_{t=1}^{\infty} \rho_t}$$



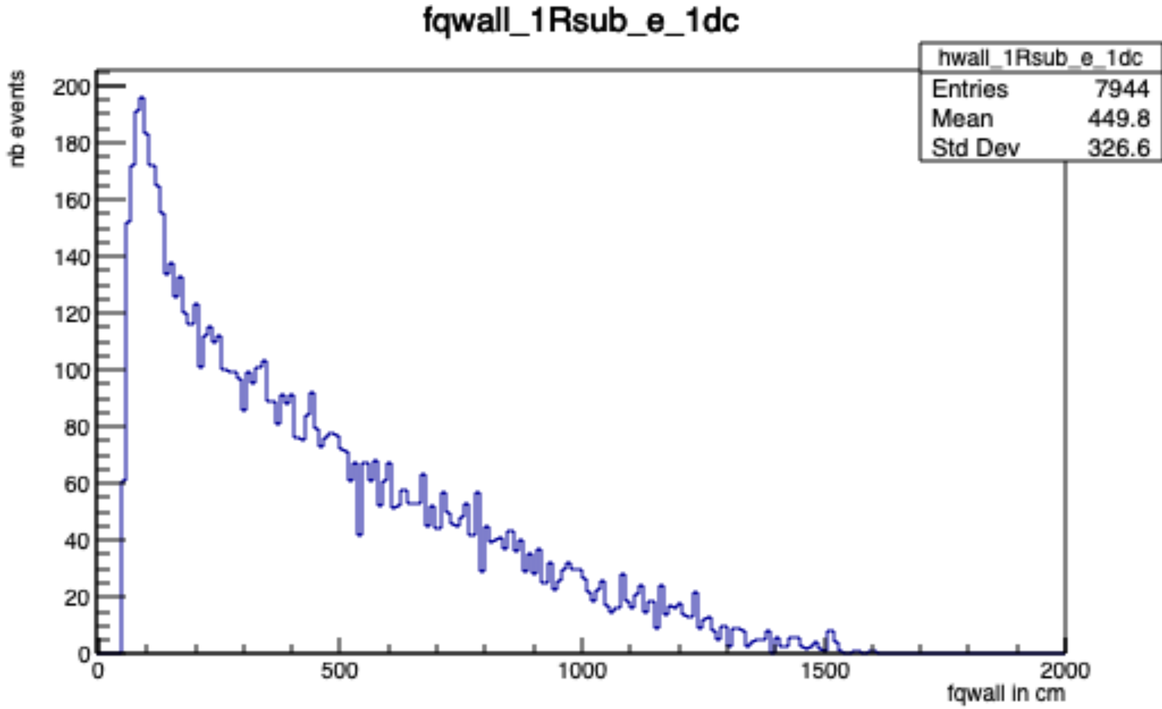
Markov Chain : We expect/need a sample to be maximally correlated with the one just before and just after (+-1) but to be independent from further samples

« Results »

Only on fqwall

(Distance vertex-nearest wall), same cut for all samples
100000 MCMC tests
2h50 (with all graphs drawn and saved—> will be changed soon)

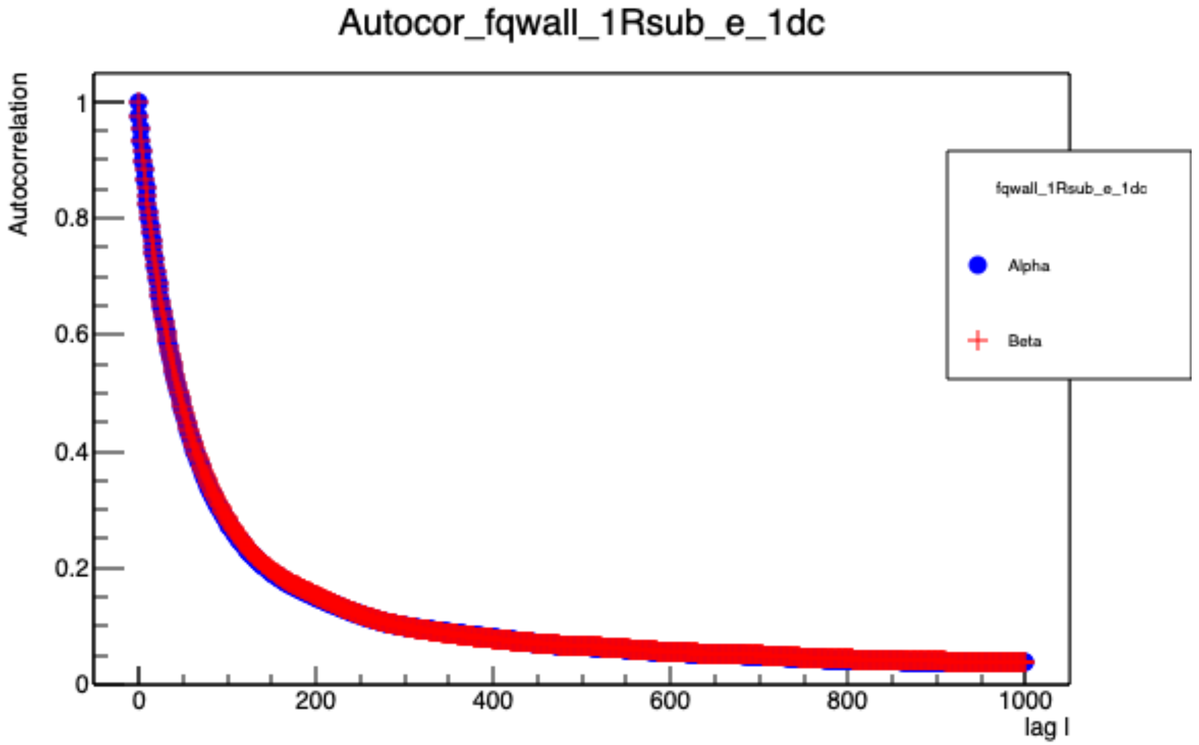
Acceptance : 51.5 - 59% (depending on rec SK samples)
Nb effective samples(alpha or beta): 200 in average —> 0.4% very low -> price of acceptance, maybe not the best optimization … (see later, correlation alpha/beta)

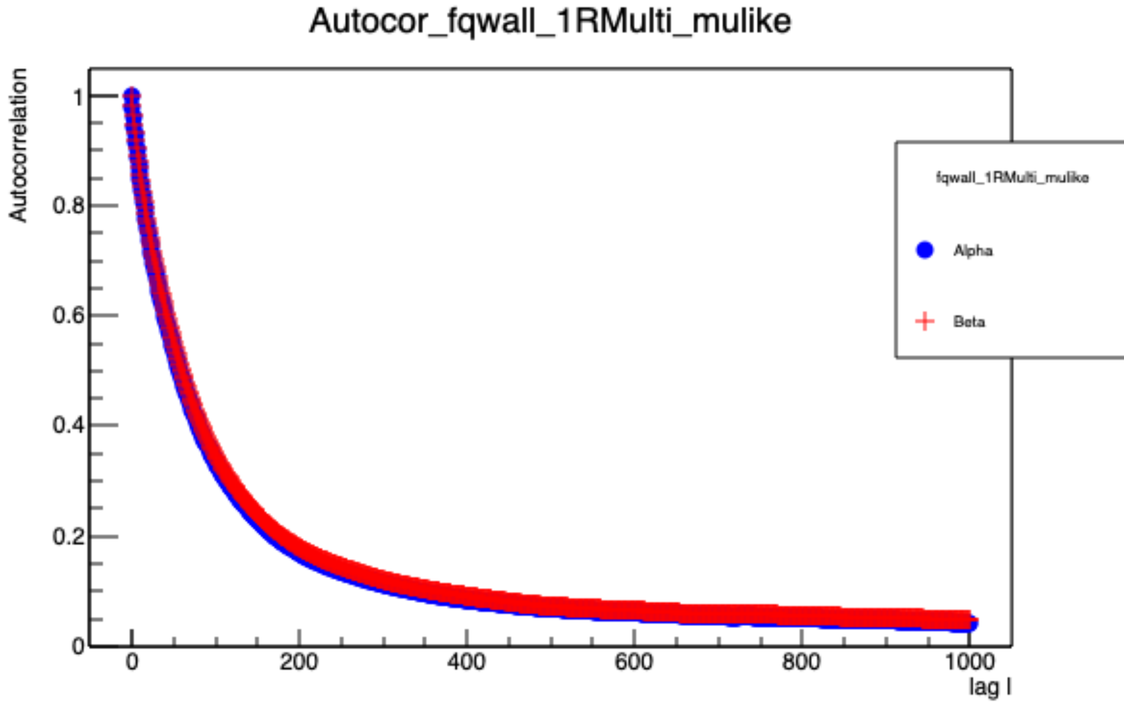


« Results »

Only on fqwall

Autocorrelation functions : behave exactly like expected —> OK

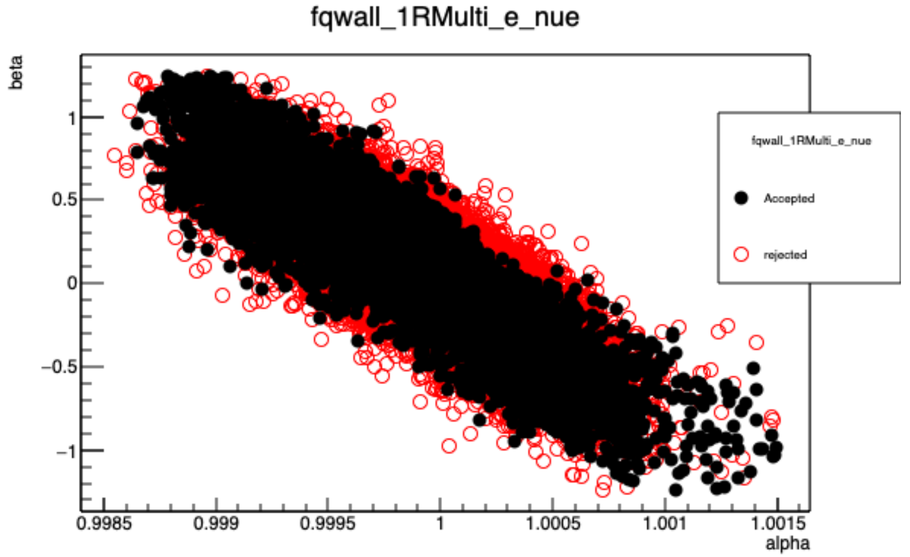


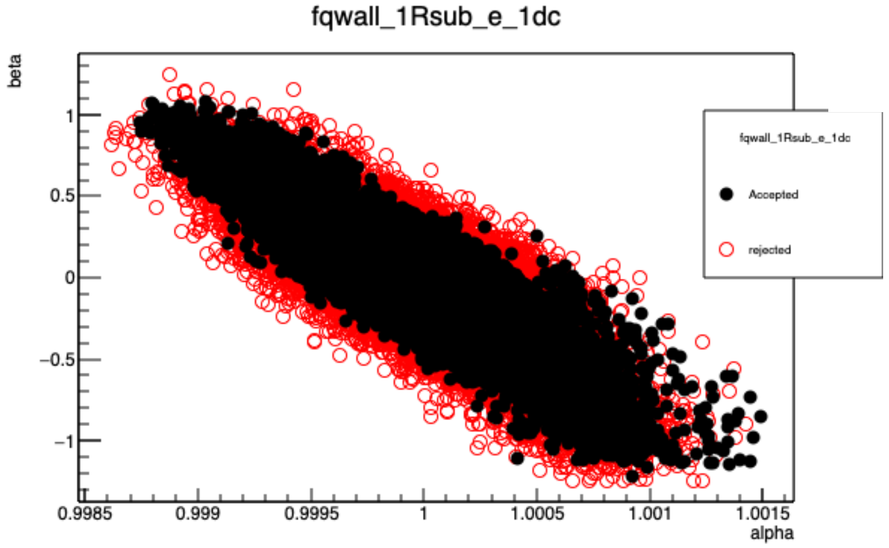


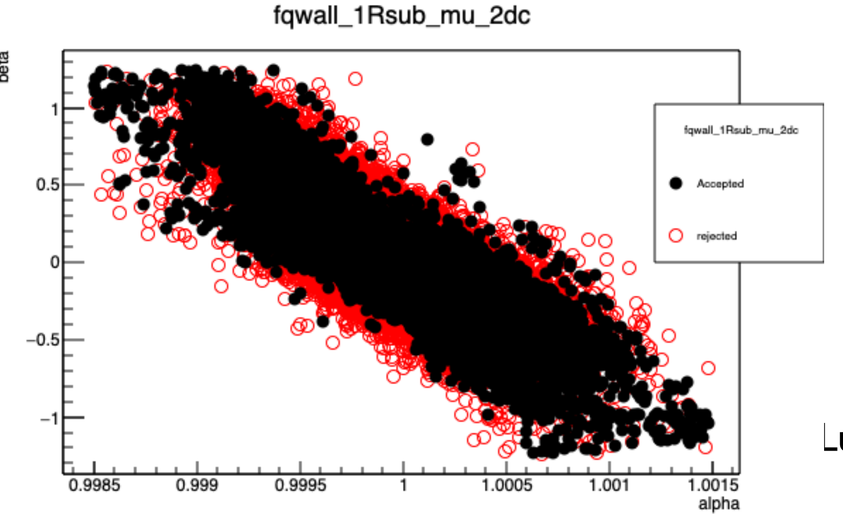
« Results »

Only on fqwall

Beta vs alpha : anticorrelated -> makes sense to get closer to nominal, could be taken into account in prior

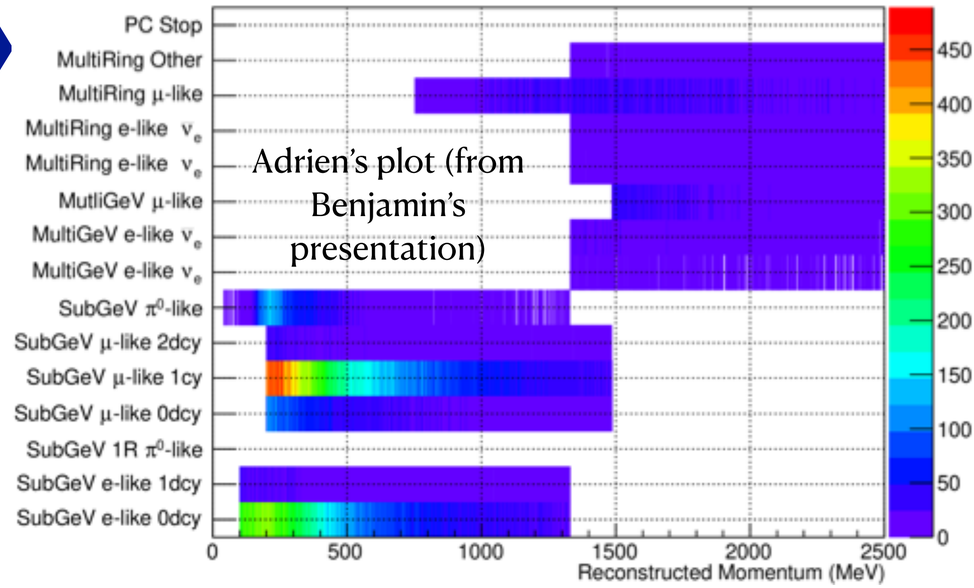






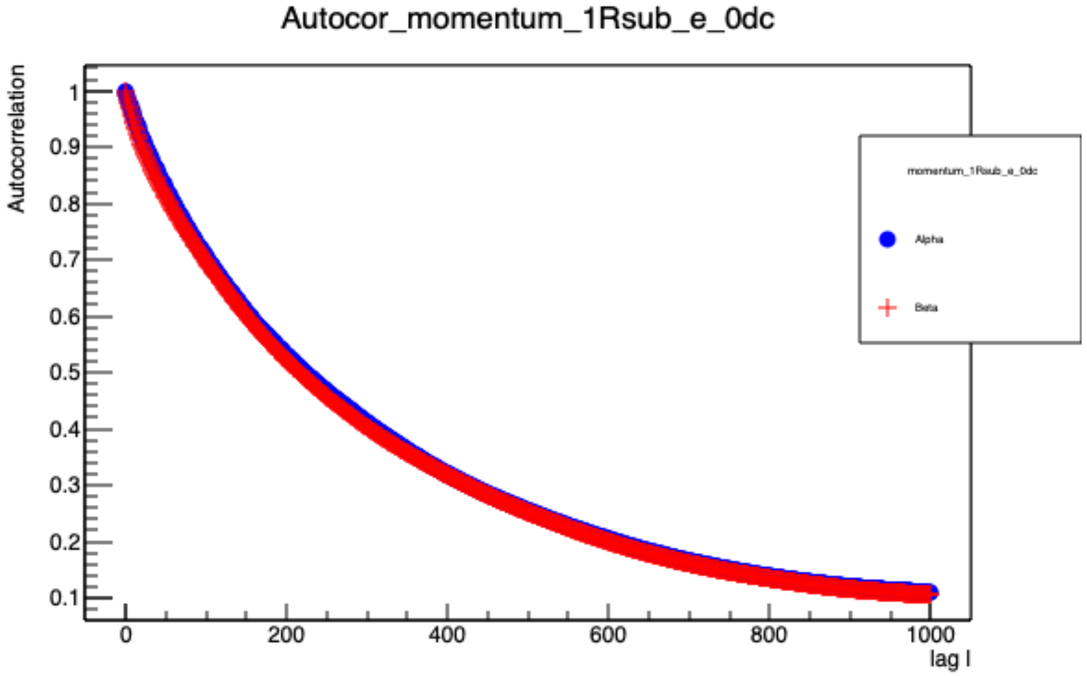
« Results »

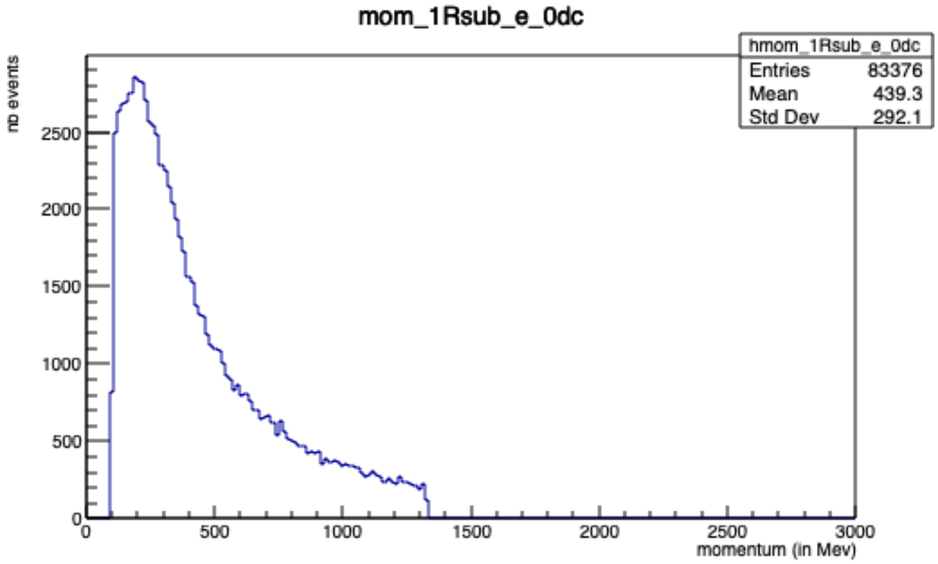
fqwall and momentum



Cut depends on sample
Running time : +10mn
Acceptance : - 10%
Neff : Wall->170
-> 0.4%
Mom ->0.16%

Autocorrelation : OK

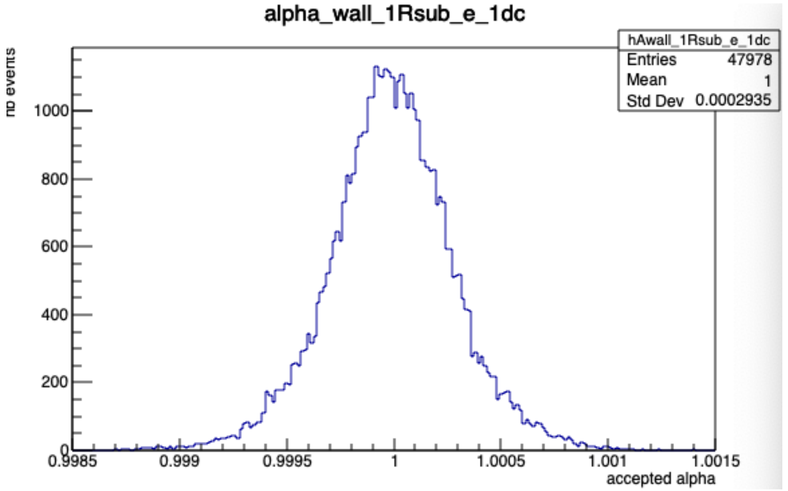


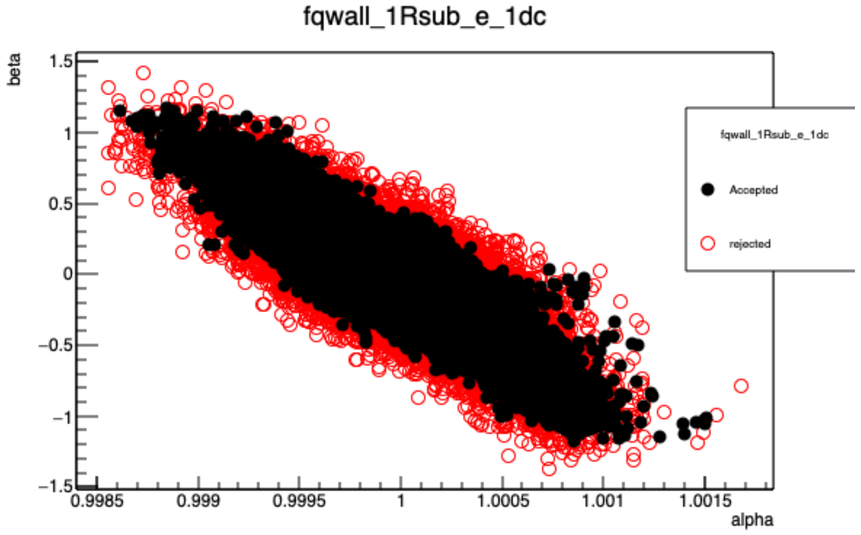


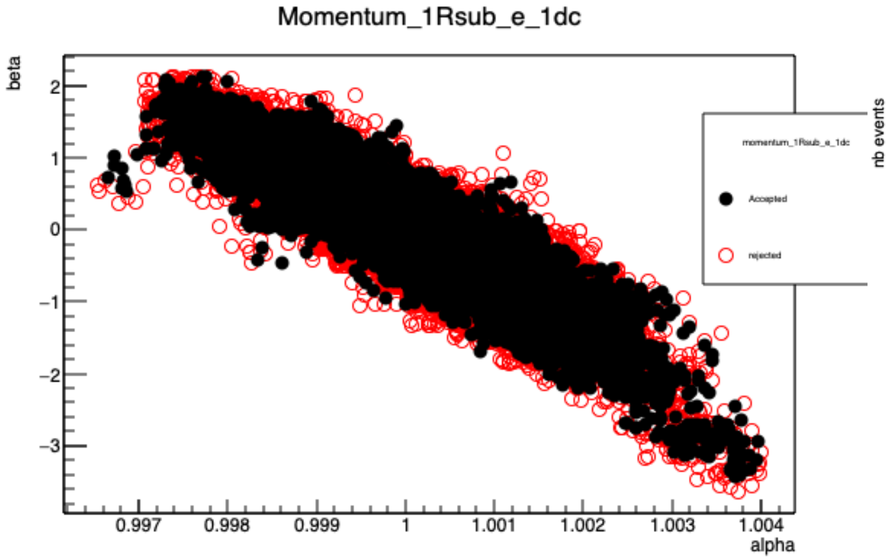
« Results »

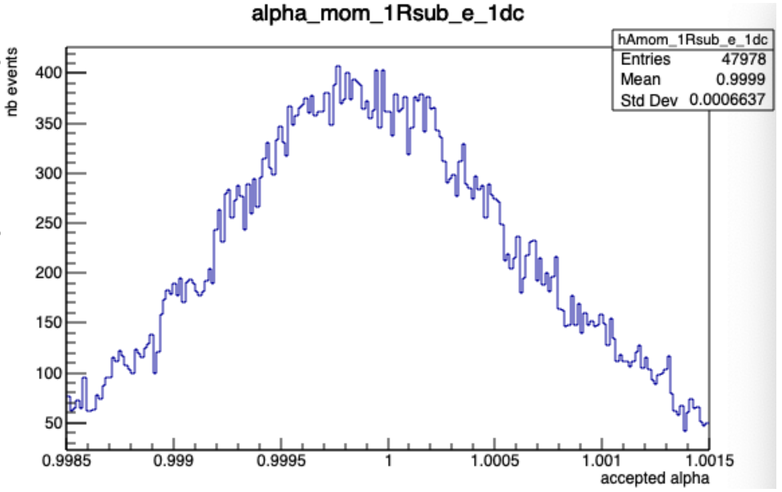
fqwall and momentum

Momentum —> larger range for accepted alphas and betas







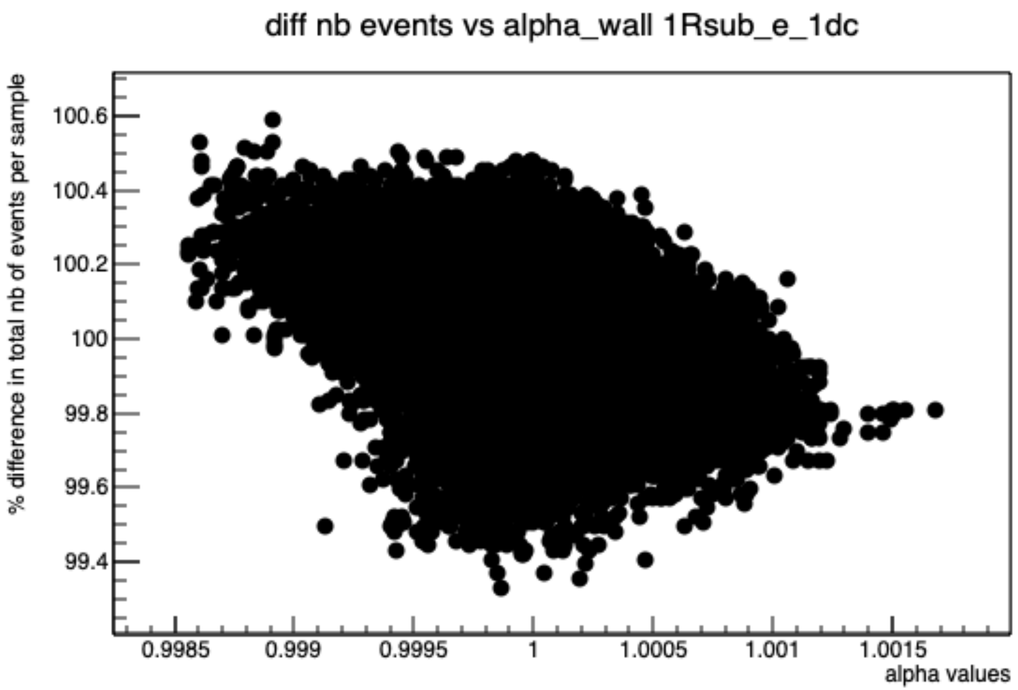


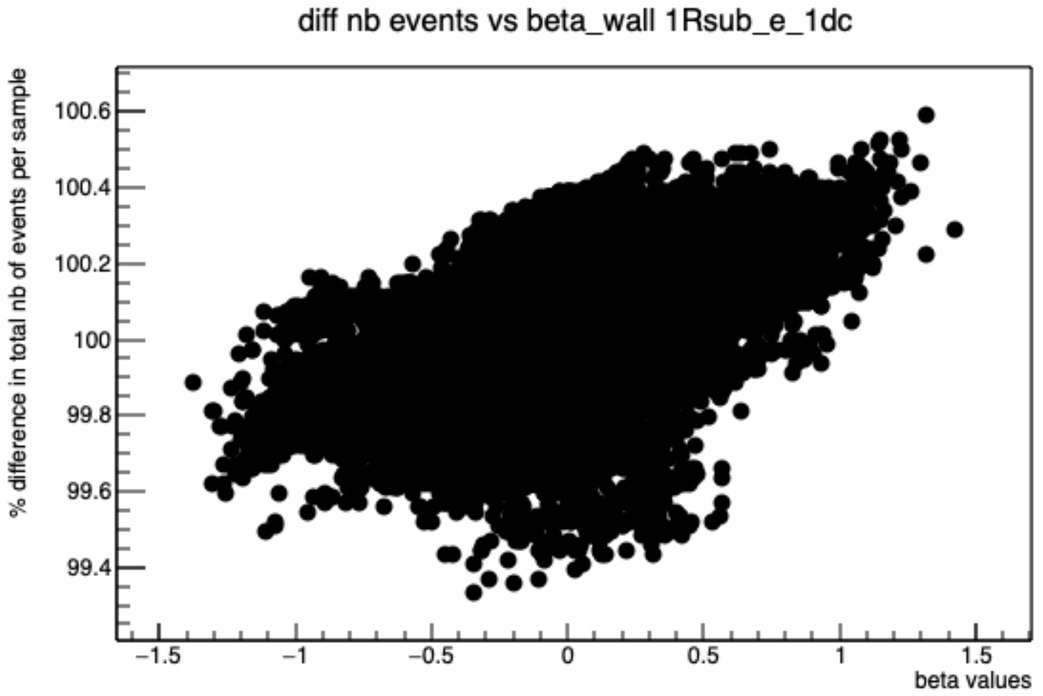
« Results »

fqwall and momentum

Center around 100% and nominal value as expected
Other points at nominal—> other parameters non nominal
+- between 0.25% and 0.8% of events maximum (total per sample)

Change in total nb of events (as a function of alpha/beta wall only)

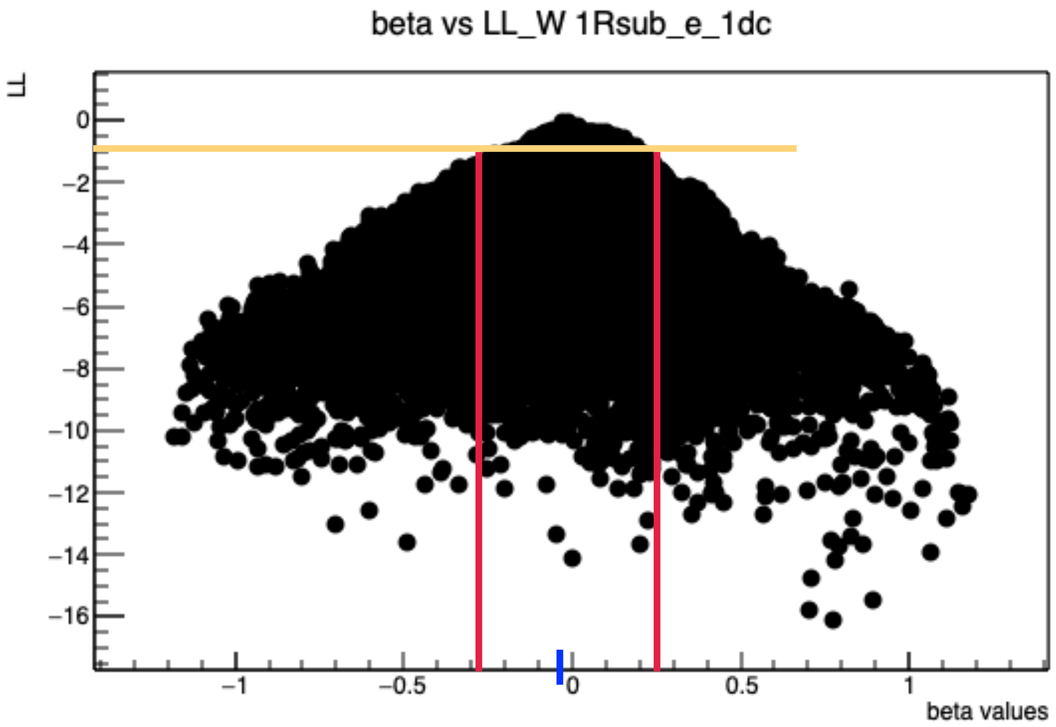


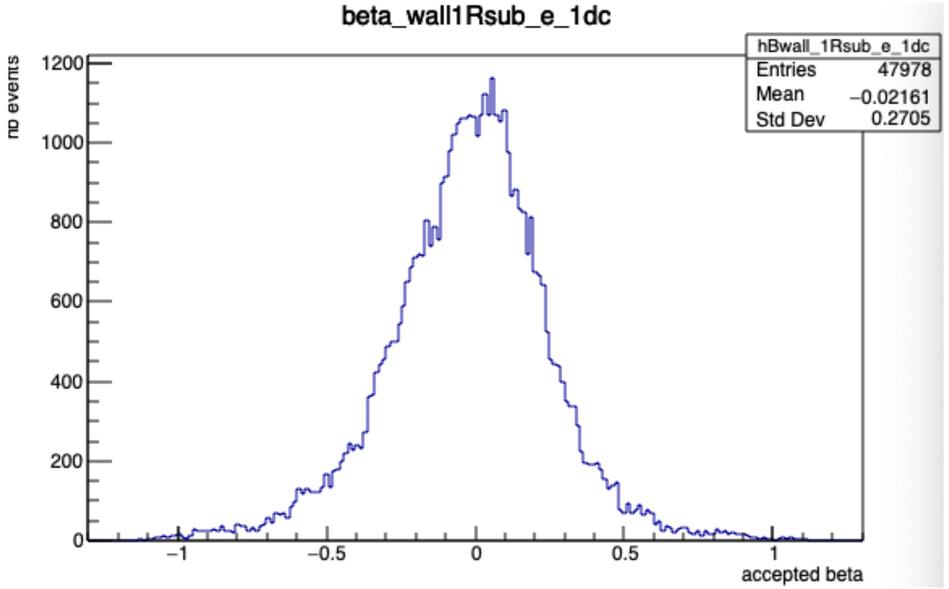


« Results »

fqwall and momentum

Likelihoods vs shifting and smearing parameters —> retrieve +- 1sigma range



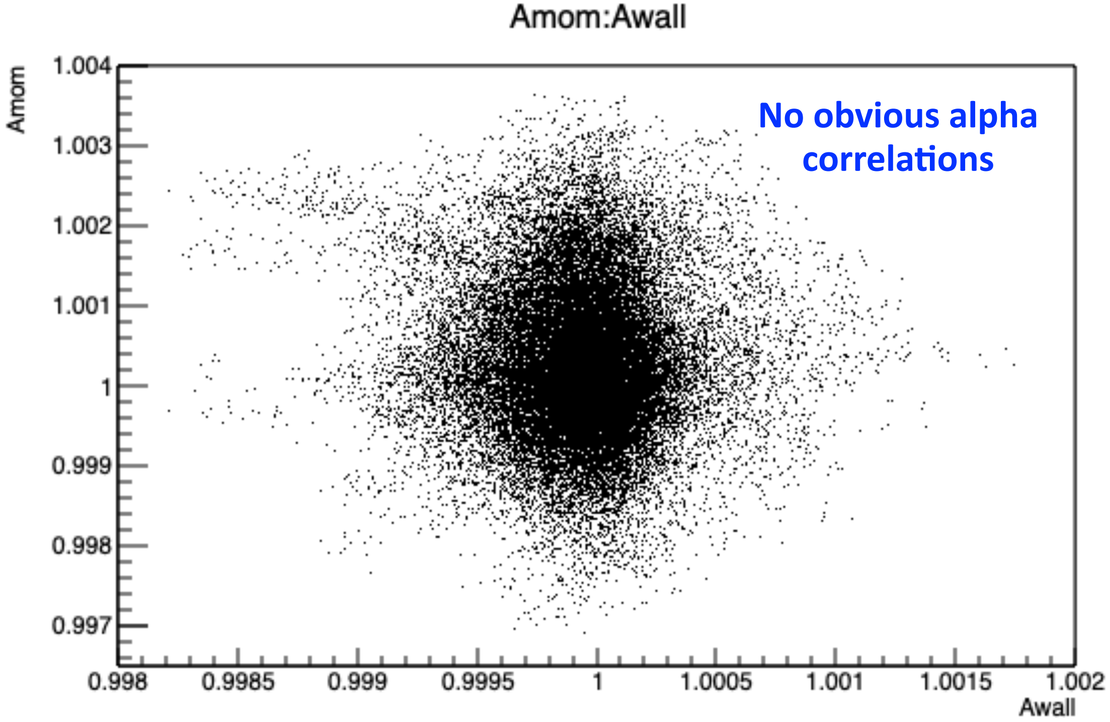


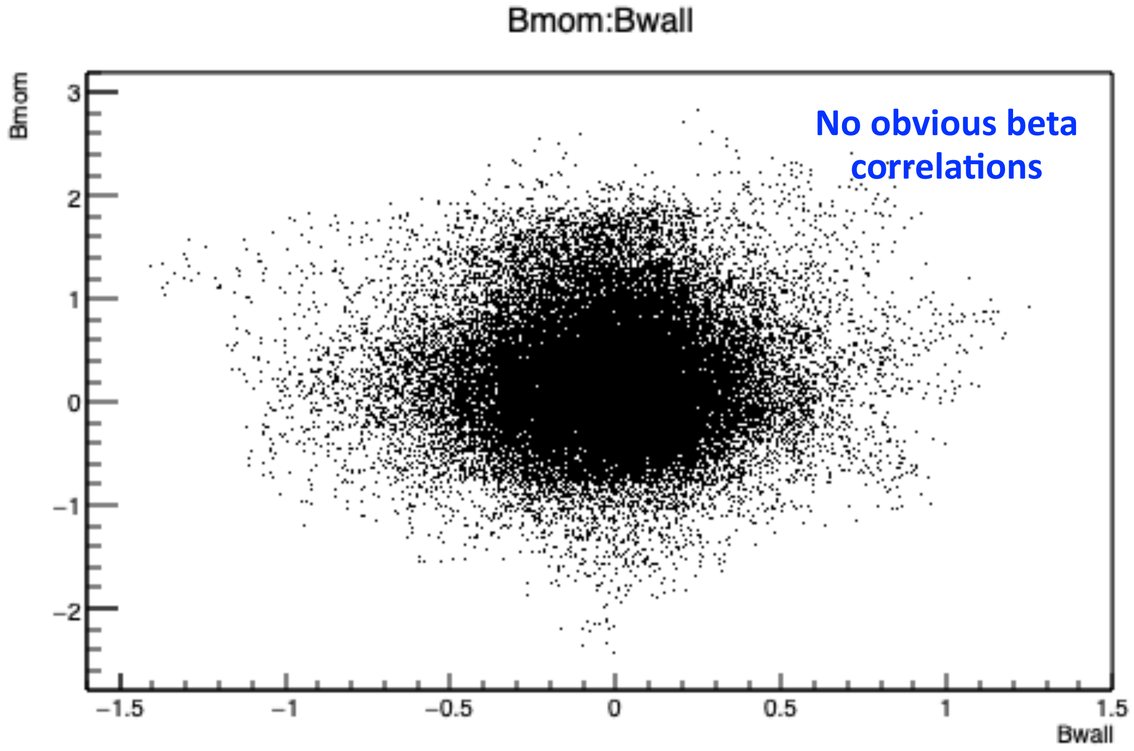
« Results »

fqwall and momentum

1Re 1dc

Correlations between cuts inside a sample ?





What’s next

—> I now have a running MCMC that I can adapt for the next steps

- **Do it with T2K cuts**
- **Take alpha/beta correlations into account**
- **Do a version with all samples to study correlations**