Red Round 1 questions
Black Round 1 answers

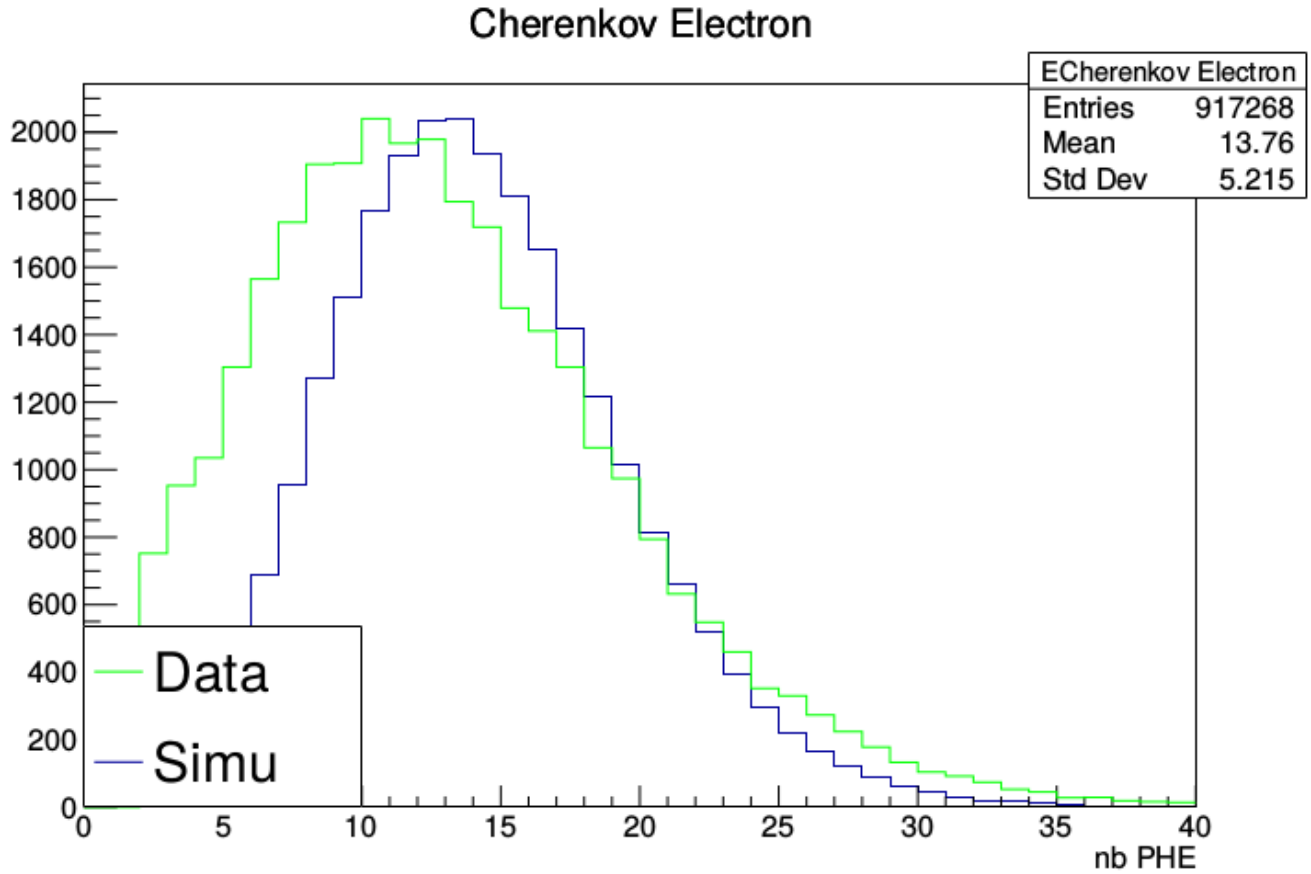
Blue Round 2 questions
Green Round 2 answers

Editing change
The plots are now regrouped within their own sections for more clarity. This add the effect to shift the line numbers.

Valery's comments

Page 16
It will be interesting to present the HTCC Nphe yield for the identified pions and electrons for data and MC events. The expected yield is presented in Fig. 1.
A 2D plot could indeed further improve the lepton PID at high momentum. However, we believe that the neural network implemented for this analysis already provides a good PID.

It was clear from the very beginning that HTCC response for pions and electrons/positrons is very different. We may clear see it from the MC plots below. The MC simulation is not perfect, it gives twice more photoelectrons than data but what is important the difference between pions and leptons response. The current HTCC MC is working much better. I am absolutely sure that adding HTCC information to the AI estimator will significantly improve the pion's rejection. It is very strange for me that it was not implemented in the code.



HTCC simulations are now almost matching the data in term of nb of photoelectrons (see above), however as explained below, the neural network provides very good performance (as evaluated now fully on data). Is it worth redoing this part of the analysis again to gain few percent of efficiency ?

Silvia's comment : as written below, the method implemented here provides a very good pion rejection, improving the background/signal ratio from 50% to 5%. Given the low statistics of this measurement, and its being a « pioneering » and very much awaited first-time measurement, we think that publishing our results in a relatively fast manner is more important than adding minor fine-tuning improvements that, given the low statistics, will for sure not modify the physics conclusions of the paper. Please keep in mind that when this analysis was done, the data/MC agreement for HTCC was not satisfactory enough to allow us to implement the HTCC nphe distribution in the NN analysis, and from the figure that Pierre shows above doesn't appear fully satisfactory even now. Going back now to the start of the whole PID procedure, including new variables, and having to redo the NN training, would be extremely time consuming (especially given the little time that Pierre can now devote to this project, having moved to a non-CLAS12 postdoc), and bring no substantial changes to the conclusions. We are aware that we are not providing a positron ID for CLAS12 that could be optimal and general for any type of future analysis, but it is good enough for this particular analysis, given that the statistical uncertainties are very big and dominate over the systematic ones.

However, we believe that the neural network implemented for this analysis already provides a good PID.

How did you estimate qualitatively that your method provides a good PID?

The discussion of section 2,3,6 shows that the ratio Background/ratio is reduced from 50% to merely 5% applying our method. We believe that it can be qualified a "better" pid using this method.

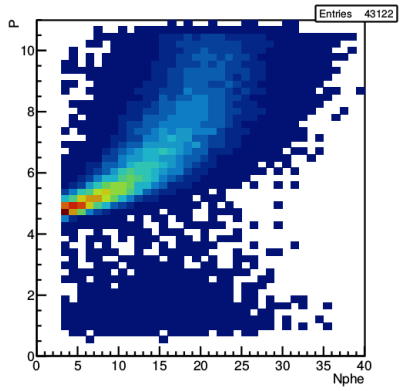
Furthermore please refer to the study described below where the neural network is tested against data, especially outbending electrons.

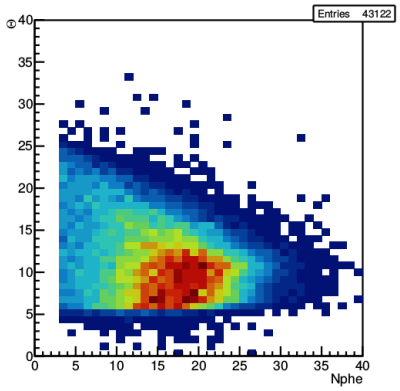
Page 17
Please provide HTCC response to the simulated pions: Nphe as a function of θ and φ as in Fig. 2.5

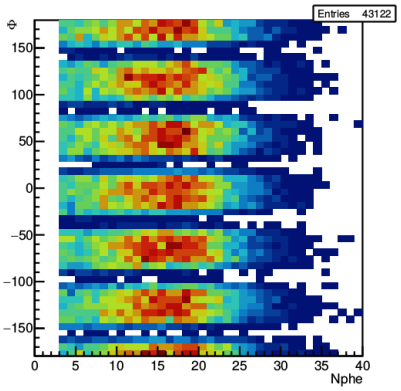
The requested figures are provided here below, however HTCC response is very poorly reproduced by simulations.

The important conclusion from these plots is the significant difference between pions and leptons HTCC response as I pointed out in the remark to the previous question. The training sample may be taken from the electrons. Outbending data are better but Inbending is still OK I think.

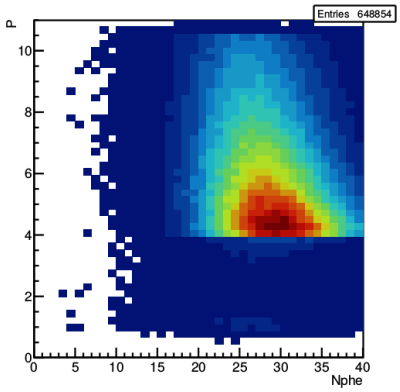
For MC-generated pions which are mis-identified as positrons:

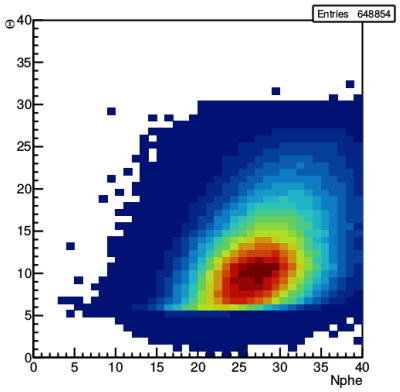


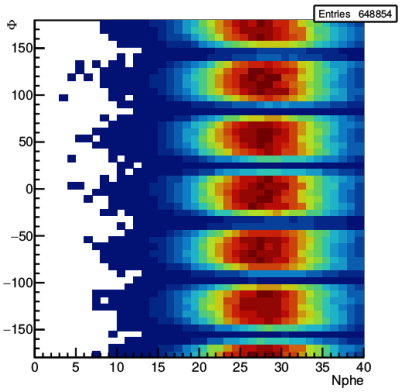




For true positrons:







Page 19: Multivariate analysis
MVA requires clean sample for the signal and background. The problem with positron ID appears only for particles with momentum > 5 GeV. However, the variables that are used in the analysis (SF, shower profile) have very weak dependence on the particle momentum. It means that we can test the proposed PID method for particles with P < 5 GeV or even apply MVA for low momentum particles. What is wrong with such approach?
I agree that such an approach could be used for smaller momenta, but further studies would have to be done. In particular the impact on training the MLP with events with small momenta should be checked. This goes beyond the scope of this analysis.

As I understand the positron efficiency and background are based in the MC study. Outbending electrons may help to estimate the positron efficiency what is very important.

See section 2,3,10. The efficiency obtained using outbending electrons is more than 98%

I believe that the goal of this analysis is to provide as good PID as possible. Why this study is beyond the scope of this analysis? How do you decide what is in and what is beyond the scope of the CLAS12 analysis note?

The neural network was applied all the way to low momenta for both leptons. See Section 3,12 "**Lepton neural network PID: further checks**"

Fig 2.21, 2.23, 2.26
Could you please show the distributions starting from lower momentum (2-3 GeV?)?
The neural network approach is only used for positrons with P>4GeV. In the lower momentum region, I do not apply any further cut than standard EB ones.

I know that you used neural network for positrons with P>4 GeV. However, you can show us how this method is working for full momentum range. I don't suggest you to change your analysis or physics results I just want to see your PID where we know that electron is electron and positron is positron.

See the previous point above discussed now in the systematic section 3,12

Fig 2.20 and 2.25
The MC efficiency (2.20) is close to 100%. Does 2.25 really present the signal efficiency?
What value of cut was used for the event's selection?
Figure 2.20 is done for simulated positrons while 2.25 was done using the same neural network but for data positrons. In this case the positrons can be both real ones and mis-identified ones. The first drop is due to the bad positrons being removed by the MLP.

In the analysis, the cut on the output of the MLP is 0.5. This cut is varied and the resulting variation is included in the systematics (see section 3.13)

The MC efficiency is 100% but what is the efficiency for real data?

See outbending electron results. We found that the efficiency is higher than 98%

Page 29 section 2.3
Can we apply positron and electron PID for all momenta? What will be the result?
In principle yes, however in the note the neural network was validated only for positrons above 4 GeV. One would need to re-run simulations below 4 GeV and validate the process again. In the case of electrons, we tried the neural network developed for positrons on simulated and data electron (coming from TCS events) (see Fig 2,30). In this case the signal efficiencies for both sets agrees, showing that the contamination from pi- is minimal.

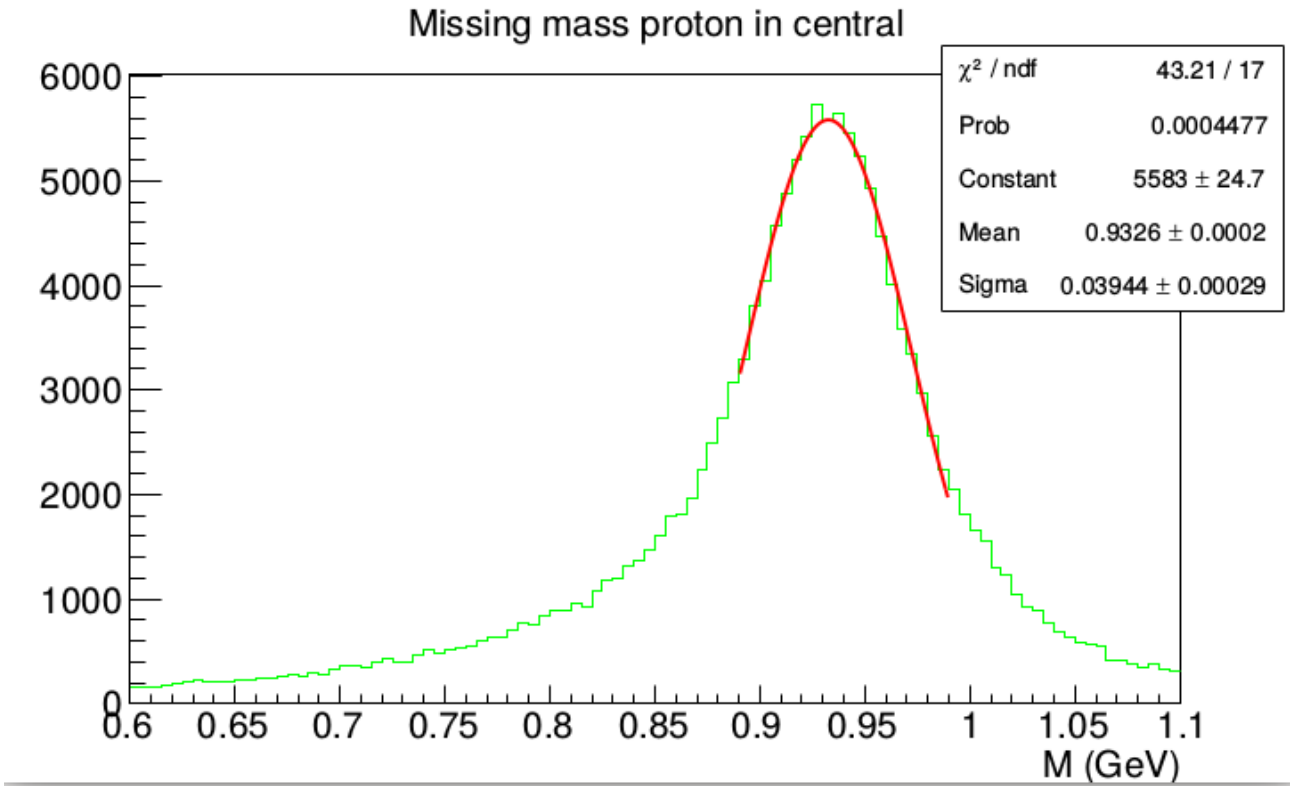
Can we apply your lepton PID without new validation and simulation? Just take it as is and run the analysis. That is was my question.

See above. This was done and no change larger than any systematics already studied is seen.

Page 34 Section 2.4.2
The data-driven correction is based on the e, $\pi+$ and $\pi-$ momenta measured in the FD.
Did you apply corrections to the momenta of these particles? Missing mass fit may help to understand the accuracy of the FD momenta measurements.
No correction is applied to pion momenta. The « detected photon » correction shown in Section 2.5.2 is applied to the electron.

I still want to see the fitted missing mass distribution. The e, $\pi+$ and $\pi-$ momenta corrections will be needed in case the missing mass is shifted from the nominal position.

Below is the fitted missing mass of the proton. The fitted mass is 0,932 GeV. The 6 MeV shift corresponds to a 0,15*sigma shift which seems reasonable



Page 52 Fig 3.6
Is there a possibility to compare cross sections?
The aim of these plots is mostly to compare the shape of the spectra to ensure the simulation agrees with the data, in order to obtain good acceptance later in the analysis. We have not done a comparison of experimental and MC cross sections.

I think that it is very important to show that the experimental cross section is close to what we expect. Fig. 3.5 and 3.6 are not comparison of experimental and MC cross sections. The simulated spectrum was normalized to have equal integral. I am asking the comparison of cross sections.

Silvia's answer: I don't think it is necessary to show that the absolute value of the simulated cross section of the data matches the experimental one. This analysis doesn't aim to measure absolute cross sections: we are going to publish ratios of cross sections. The absolute normalization of the RGA data has not yet been attempted by any analysis, and it should be carried out first on a known channel, in order to make sure it is correct. This is a long project that goes way beyond the scope of this paper. The MC simulation in this analysis is used to compute acceptance corrections. Acceptance is a ratio, so the overall normalization doesn't count, what matters is the slope of the kinematic distributions. We show that the slopes are in good agreement. We also provide a systematic check on the acceptance using different MC slopes (with/without CS weighs). Trying to extract absolute cross sections from these data would be a full new PhD thesis.

Line 948
Misprint: event.
Could not find the mistake…can you please clarify?

Now it is line 958
Final states were identified using the **even** builder particle IDs

This is now corrected

Page 75 Line 1181
Where is it coming from the range 0.5±0.1 in neural network systematic error study? How do you choose the value 0.1?
According to Figure 2.25 the signal efficiency is fairly flat in the 0,4 to 0,6 cut region. We intend to stay in this region. Furthermore, according to Figure 2,19 the background rejection is also fairly constant in this region

There has to be the reasonable explanation when you are doing the systematic error study. You can take 0.5+/-0.01 for example. It is also in the flat region. I still don't understand why you use 0.5+/-0.1 for this study.

We now use 0,5+/-0,3. These values corresponds to the region where the signal efficiency is large while the background reduction stays high. This is explained in greater details in the note in chapter 3,12 and in Section 2,3,10

Fig 4.1 to 4.12
Can we compare these measurements with the theoretical model predictions?
How do you calculate the average values for $< t >$, $< M >$, $< E\gamma >$? It becomes important when you are trying to compare your data with the theory predictions. If your average is just weighted with the cross section (and/or acceptance) value it may give you incorrect result. This is especially important when the cross section changes significantly inside the kinematic bin. I believe it is your case. …

This point is delta with in a new section in Section 4,2

I like your new Section 4.2.

"Also, we have shown that the standard deviation between the mean kinematic points (Method 1) and the average yield points (Method 2) is in our case close to the standard deviation of both M and E"

How do you determine the standard deviation between the mean kinematic points (Method 1) and the average yield points (Method 2)?
Did you use Fig. 3.6 for the standard deviation of both M and E?

The method to obtained this value is now explained in this section.

Page 105 Appendix A
The background/signal ratio is completely based on the MC. There is no systematic error study connected with this method.
The quoted 5 % value for the background/signal ratio is not used directly in the final results. The systematic check connected with this is the one assessing how the results vary when the positron ID cut is varied.

In this case you have to answer to my question Page 75 Line 1181.

See commentaries discussed above. This is now checked and discussed in the note

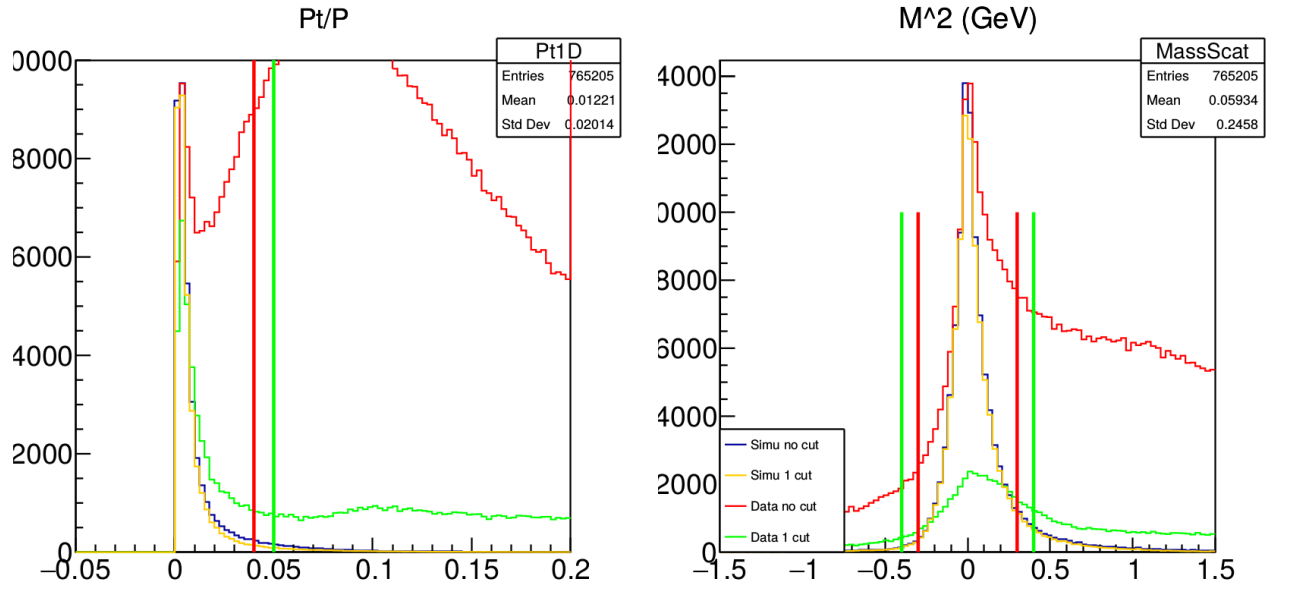
Timelike Compton Scattering data analysis – Review
Aram Movsisyan
Section 3.1

It would be useful to specify which of the observables are implemented in MC generators.
This would help to better understand the further sections.
The generator attributes to each event a weight obtained as described in Section 3.1. Each time we compute MC observables in the following (i.e. the BH red points in the results plots), the analysis code is ran on simulated events, with the weight corresponding to each event.

My point here was just to mention in the text of the section which observables are expected to vanish for given MC generator and which observables are expected to have non-zero signal. This will help reader to better understand sections of systematic studies. (Like do you expect non-zero AFB from Grape generator, or do you expect non-zero BSA from TCSGen with TCS weights, …….).

I have added few words to clarify this points in section 3,1,2

Fig.3.1 does not give any clue on the numerical values of the cuts. In addition, 2D plots are not showing the consistency of resolutions in data and MC. Could you please provide 1D plots of missing mass and transverse momentum (comparison of data/MC) to show that the resolutions are properly reproduced in MC.
The data/MC comparison shows compatible resolutions for pt/p while the width of the MM peak is not as well reproduced. This is likely due to the fact that our protons are mainly in the CD and that the simulation is not yet well tuned to reproduce the CLAS12 performances. The MC was nonetheless necessary to guide the choice of the cuts. The values were picked in order to minimize the loss of good events on the MC, while at the same time cut out the background from non-photoproduction events on the data. To illustrate this statement see next picture where simulation and data 1D exclusivity cuts are shown, along the values of the cuts used.



Perfect. Why not to add a sentence in the text describing how exactly the values were picked.

I have added this plot/comments to the note in Section 3,2,2

Line 1170-1171. Not clear, how the systematic uncertainty is calculated.
The uncertainty is the difference between the observable extracted directly from MC, and its value calculated after all the steps of the simulation (GEMC, COATJAVA). As I had two different MC samples coming from two different generators with different characteristics, I decided to test this procedure for both generators. The chosen uncertainty is the « worst » values of the difference in any of the two generated samples. Note that the difference between observables extracted from the two MC samples is always small. This method is redundant but provides a thorough cross check.

Ok. Choosing the "worst" of two values is an option. The other possibility would be to take the RMS of two values with respect to zero: \sqrt([err1*err1+err2*err2]/2).

We have decided to keep the worst option.

Lines 1177 – 1184. Is the reference to figure 3.31b correct? What is the effect of background merging on the results?
This was tested in the manuscrpit of my thesis. The variation is minimal. However, BG merging should be used and included in the final results. Thus all the results shown in this note include it.
Is the variation smaller or larger (and how much) compared with the systematic error due to proton efficiency?

A new Subsubsection in Section 3,12 now shows the plots from the thesis and they are discussed wrt to your comments.

Why do you assume that background merging and proton detection efficiency in CD should be comparable?
As all the observables we extract in this analysis are ratios, one could expect the efficiency contributions in the numerator and denominator to cancel out.
Unless you combine data with different efficiencies, but this discussion is beyond TCS analysis where errors are dominated by statistical uncertainties.
If this is the case then adding the proton efficiency should not have large effect on the extracted values. We wanted to verify this statement by producing this systematic uncertainty.
I just want to make sure that the systematic uncertainty due to background merging is smaller then the size of correction itself (see the comments above)

See the point above and its discussion in the note.

Lines 1196 – 1202. How the systematic uncertainty is applied. Are there mis-identified protons in MC for \xi^2 beyond 3 \sigma (For TCS process)? If not, then why you apply 3\sigma cut in simulation? If yes, then you need an estimate for proton mis-identification.
The systematic uncertainty is defined as the difference of the measured observable with or without a chi2 cut on the proton. The acceptance is also recalculated with the correct Monte Carlo cuts.
Yes, this 3 sigma cuts removes the tails in MC. The parameters of the gaussian are extracted both in MC and data to account for the difference in width of the distributions.
You already have systematic uncertainties due to positron ID, proton detection efficiency and exclusivity cuts. So applying one more systematic error due to proton \xi^2 is perhaps overdoing. But its up to you, as long as you are dominated by the statistical errors.
As for now we'll keep this set of systematics

Figure 3.32: Why the AFB value for "DATA" in the first -t bin is different in plots (a,b,c,d,e,f)?
I compare yellow from (a), blue from (b) and blue from (e)
I have mistaken some old plots with recents ones when copying them to the tex folder. I have fixed this and put only the latest plots.
Thanks for the notice !

Timelike Compton Scattering data analysis – Review
Francesco Bossù
Subsection 2.3.5
- It is not clear to me that this can be done so easily: could you provide comparisons data-MC of the input variables to show that the shapes are compatible?
The comparison is shown in Figures 2.27 and 2.28 of the updated note. One can see that the distributions are compatible, providing a further cross check.

Thank you for the additional plots.
The sentence “the distributions are compatible” is rather vague. In addition the plots show differences in the mean values between data and MC and tails that are not really the same.
Could you quantify this compatibility between the distributions?

We decided to validate the procedure by comparing directly the Signal Efficiency VS background rejection. This is shown in Figure 2,33. Both ROC curves for simulated positrons and for data (outbending electron/missing neutron events) are cery close. Furthermore the Signal efficiency estimated using outbending electons shows no dependance with the kinematic variables. This make us believe, that while the Input variables distributions are slighty different, the performance on data are well reproduced by simulations and can thus be used for acceptance calculations.

o) the calorimeter signals for the pi+ you have it from the reaction cited above
As shown in Figure 2.3 there is some background below the neutron peak. So the mis-identified pions cannot be identified at an event-by-event level.

The sPlot technique could be explored: it allows you to extract the weights for the signal and the background on an event-by-event basis that can be used to obtain the distributions necessary to train the MVA classifier.

This is a possibility, but I believe this has to be done with care. Some events can also be given negative weights by the sPlot technique, in this case a MLP can’t deal with them (but they can be ignored during training according to the ROOT MVA documentation).
It is a nice idea but rather time consuming at this stage of the analysis.

Figure 2.30. The top right plot shows a second population below the fitted curve. What does this population come from?
These are the events around 27° which, due to resolution, are reconstructed above 27°.

So, which correction function are you using for this population?

The corrections for proton above 27° is used. A slightly cleaner way of handling these could be developed. Nevertheless, these are few events and the error on the correction should not induce large effect on the final observables.

Also, since the magnitude of this effect depends on the beam position, this study should be done ideally run by run. Or at least, since in the run period that you analyzed, the beam spot moved (i.e.there are at least to CCDB tables for the beam postion), this study and this correction should be done separately for the two sub-periods (or at least a cross check that this correction is independent on the run period)

ccdb shows the following two beam shifts for the run period used in this analysis:

Run range: 4760-5277; based on 5038; x, y (cm) 0.0031 -0.041
Run range: 5278 5419; based on 5306; x, y (cm) 0.035 -0.10

There is therefore only a 300 microns variation between the two parts of the run. We don't believe that our corrections can reach such a level of precision.

Actually, since the FD and the CD are not aligned and since the FD is not as precise as the CD for the vertex reconstruction, I don't think one can assume a complete correlation between the measurements by the FD with the vertex in the CD.
Then, I agree that given the precision, the corrections will not change dramatically.

Thus, we shall keep the current procedure, though larger statistic measurement should deal with this point more carefully

If equation 2.25 is actually used, then I do not understand how the following figures (like 2.42) are bounded to 1.
Here « missing » means ep→ e (p) pi- pi+ events where the ***missing proton*** can either be detected or remain undetected. In other word it is the total number of events where a proton is expected to be detected.
I see. I think that the text is confusing and the definition of "missing proton" could mislead the reader. I suggest that you rephrase this sentence. Maybe "expected" instead of "missing".
And in the text you should specify that "missing" means that you did not consider if it is reconstructed or not, i.e. the total number of protons.

I have added a sentence with the definition of missing in section 2,8. This hopefully clarify the section.

==========

Additional comments:

Section 1.1

I think there is a mismatch between the \phi symbol used in the text and the \varphi used in the schematic of Figure 1.2. It will help any future reader to normalize the convention.

This is now changed

Line 161. "…, the quark loop in the TCS…". The term "loop" suggests some (N)NLO pQCD diagrams. Is this the accepted way to describe it? I would suggest to use "line" instead of "loop", since the diagrams are at tree level.

I have adapted the phrasing

Line 202. J/\Psi → J/\psi
This is changed

Section 3.12, Systematic uncertainties.

*) Acceptance model. Line 1243. Can you explain the choice of associating ±Delta_acc/2 as uncertainty? Are you assuming a particular probability density function of this uncertainty?
If not, I do not see the point of using the factor ½: either you use an asymmetric uncertainty, or you decide to use Delta_acc as symmetric around the central value, as you don't know the true distribution of this uncertainty.

Everywhere in the note, I use symmetric distribution with total width equal to Delta_acc (ie Delta_acc/2 symmetric around 0).

Assuming gaussian distribution, the estimator of the \sigma yield to \sigma_bar=1/2(X_1-X_2). This is actually in this case the biased version of the estimator. In the unbiased case it yields \sigma_bar=1/sqrt(2)(X_1-X_2).

We have used the biased estimator here.

*) Exclusivity cuts. These kind of variations used to estimate some systematic uncertainties are dangerous.
It is not written, so I assume that the acceptance x efficiency is not recomputed with two set of cuts.

The AxE is recalculated accordingly. I have made the note clear on this point.

If this is the case, (i.e. same AxE as for the standard cuts), then the variation seen in Figure 3.32 (e) is due to the slightly different data sample. And I am not sure it actually makes sense to consider it as a systematic uncertainty. I would suggest to recompute the AxE for these cuts.

As said above this is not the case.

Instead, if the AxE was in fact recomputed and applied for the different cuts, then the variation shown in figure 3.32 (e) is worrisome: there is a ~30% variation on the A_FB, if I read correctly Figure 3.32(e). This would suggest that there is a background contamination that is not under control.

These large variations are mostly visible in the case where a very small fraction of events are used to calculated the observable (eg AFB). In the case of BSA and R ratio this systematic is less prevalent. We believe these variations comes from the low statistic of the analysis. The smoothing technique implemented for the systematics helps lower this effect.

And the statement "However, the observed systematic shift always remains within the statistical error bars" is not correct: these populations comes from the same data set, therefore are completely statistically correlated and their compatibility cannot be judged on the base of the statistical uncertainty.

Important Note
A new subsection 2,3,11 deals with the choice of the cut for the neural network.