

Timelike Compton Scattering data analysis

First draft

Pierre Chatagnon, Silvia Niccolai, Stepan Stepanyan, the ee analysis group
for the CLAS collaboration

Abstract

Generalized Parton Distributions (GPDs) describe correlations between the longitudinal momentum and the transverse position of the partons inside the nucleon. GPDs have been studied in many experiments worldwide mainly using Deeply Virtual Compton Scattering. This note reports the measurement of the time-reversal conjugate process of DVCS, Timelike Compton Scattering (TCS). TCS on the proton is the photoproduction of a virtual timelike photon, which then decays into a lepton pair. Experimental studies of DVCS and TCS are complementary. Indeed, beam and target spin asymmetries for DVCS give direct access to the imaginary part of combinations of Compton Form Factors (CFFs, which are related to GPDs), whereas the angular asymmetries of the decay lepton pairs in TCS allow to access primarily the real parts of CFFs. The experimental measurement of the TCS angular asymmetry provides new information on the real part of GPDs, which is less constrained by existing DVCS data than the imaginary part. The upgraded CEBAF accelerator and the recently constructed CLAS12 detector of Jefferson Lab provide the ideal setting to perform a TCS experiment. CLAS12 took data with a 10.6 GeV electron beam impinging on a liquid-hydrogen target in the spring and fall of 2018. This note assesses the analysis of CLAS12 data and present results for TCS.

Contents

1 Physics motivations	7
1.1 Phenomenology of TCS	7
1.1.1 The CFFs of TCS	8
1.1.2 TCS cross section	9
1.2 TCS observables	10
1.2.1 R ratio	11
1.2.2 Forward-Backward asymmetry	11
1.2.3 Photon polarization asymmetry	11
2 Analysis strategy, particle identification and momentum corrections	13
2.1 Analysis strategy for TCS	13
2.2 Particle identification	14
2.3 Positron identification	15
2.3.1 Evidence of π^+ contamination	15
2.3.2 1D and 2D cuts from the simulations	17
2.3.3 Multivariate analysis approaches	18
2.3.4 Training, testing and comparison of MVA classifiers on simulations	20
2.3.5 Test and comparison of MVA classifiers on data	21
2.3.6 Remaining background estimation	25
2.3.7 Systematic checks on simulations	27
2.3.8 Effect on data	28
2.3.9 What about electron PID?	29
2.3.10 Crosscheck using electrons in the outbending torus magnetic field configuration	31
2.3.11 Choice of the cut on the output of the neural network	32
2.4 Proton momentum corrections	38
2.4.1 Monte-Carlo corrections	38
2.4.2 Data-driven momentum corrections	39
2.5 Lepton momentum corrections	42
2.5.1 Monte-Carlo corrections	42
2.5.2 Detected radiated photon correction	42
2.6 Fiducial cuts	47
2.7 Background merging	49
2.8 Proton efficiency correction	49
2.8.1 Efficiency correction in the central detector	49
2.8.2 Efficiency correction in the forward detector	50
3 Simulations and extraction of the TCS observables	55
3.1 TCS event generator and simulations	55
3.1.1 GRAPE	55
3.1.2 TCSGen	55
3.2 Event selection	56
3.2.1 Final state selection	56
3.2.2 Exclusivity cuts	56

63	3.3	Phase space of interest and Simulations/Data comparison	60
64	3.4	Acceptance estimation	62
65	3.5	Background estimations	68
66	3.5.1	Electro-production of a lepton pair $ep \rightarrow pe^+(e^-)e_{scattered}$	68
67	3.5.2	Pion contamination	73
68	3.5.3	J/ψ contamination	73
69	3.6	Experimental cross-section ratio	74
70	3.7	Phenomenological study of the TCS Forward-Backward asymmetry	75
71	3.7.1	Early considerations and comparison with other models	75
72	3.7.2	TCS A_{FB} kinematic dependencies	77
73	3.7.3	TCS A_{FB} model dependencies	78
74	3.8	Experimental Forward-Backward asymmetry	80
75	3.9	Experimental beam-spin asymmetry	81
76	3.10	Photon polarization transfer	82
77	3.11	Binning of the data	84
78	3.12	Systematic errors estimation	85
79	4	Results and comparison with model predictions	93
80	4.1	Complete CLAS12 results for the TCS observables	93
81	4.1.1	R' ratio	93
82	4.1.2	A_{FB}	93
83	4.1.3	BSA	94
84	4.2	Note on the theoretical comparison	94
85	4.3	Tabulated results	110
86	4.4	Comparison with CLAS results	113
87	4.5	Comparison Data/Models and physical interpretations	114
88	Appendices	121	
89	A Derivation of the background/signal ratio	123	
90	B Generator checks: comparison between <i>GRAPE</i> and <i>TCSGen</i>	125	
91	C Final state particle kinematics	129	
92	Bibliography	132	

⁹³ Introduction

⁹⁴ This note describes the Timelike Compton Scattering analysis using the first CLAS12 data. It is
⁹⁵ divided in four chapters. In the first chapter, the motivations for the measurement of the TCS process
⁹⁶ are briefly presented. In the following chapters the analysis procedures are detailed. The final-state
⁹⁷ particles are identified using the CLAS12 EB described in the RGA analysis note. An enhanced positron
⁹⁸ identification algorithm, especially developed for the TCS measurement, is presented in the second
⁹⁹ chapter of this note. The second step of the analysis consists in correcting the data. These corrections
¹⁰⁰ aim at matching the measured momenta of the detected particles with their actual momenta at the
¹⁰¹ vertex. This step is also presented in Chapter 2 of this note. Once the final state particles are well
¹⁰² identified and their momenta are corrected, exclusivity cuts are applied in order to make sure that each
¹⁰³ event is a TCS event (i.e. from the $\gamma p \rightarrow p' e^+ e^-$ reaction). The observables are then computed with
¹⁰⁴ the kinematic variables of these good events. This step is performed in Chapter 3. Finally in Chapter
¹⁰⁵ 4 the results of the full analysis of the CLAS12 data are displayed and discussed against theoretical
¹⁰⁶ predictions.

¹⁰⁷ This note includes sections and subsections which can also be found in the RG-A common analysis
¹⁰⁸ note. These sections have been developed during this analysis but are of common interest to all
¹⁰⁹ analysers. In order to keep consistency and clarity, these sections are clearly identified with the
¹¹⁰ following markers:

¹¹¹ *Start of common analysis section*

¹¹³ ...

¹¹⁴ *End of common analysis section*

¹¹⁶ Data set

¹¹⁷ **Analysis :** Fall 2018 - Inbending - Skim 1 (TCS/Jpsi)

¹¹⁸ Run list : 5032 5036 5038 5039 5040 5041 5043 5045 5046 5047 5051 5052 5053 5116 5117 5119 5120
¹¹⁹ 5124 5125 5126 5127 5128 5129 5130 5137 5138 5139 5153 5158 5159 5160 5162 5163 5164 5165 5166
¹²⁰ 5167 5168 5169 5169 5180 5181 5182 5183 5189 5190 5191 5193 5194 5195 5196 5197 5198 5199 5200 5201
¹²¹ 5202 5203 5204 5205 5206 5208 5211 5212 5215 5216 5219 5220 5221 5222 5223 5225 5229 5230 5231
¹²² 5232 5233 5234 5235 5237 5238 5239 5247 5248 5249 5250 5252 5253 5257 5258 5259 5261 5262 5300
¹²³ 5301 5302 5303 5304 5305 5306 5307 5310 5311 5315 5316 5317 5318 5319 5320 5323 5324 5325 5333
¹²⁴ 5334 5335 5336 5339 5340 5341 5342 5343 5344 5345 5346 5347 5349 5351 5354 5355 5356 5357 5358
¹²⁵ 5359 5360 5361 5362 5366 5367 5368 5369 5370 5371 5372 5373 5374 5375 5376 5377 5378 5379 5380
¹²⁶ 5381 5382 5383 5386 5390 5391 5392 5393 5394 5398 5399 5400 5401 5402 5403 5404 5406 5407 5414
¹²⁷ 5415 5416 5417 5418 5419

¹²⁸ **Positron PID :** Fall 2018 - Inbending - Skim 1 and Skim 4 (inclusive)

¹²⁹ Simulation settings

¹³⁰ GEMC 4.3.2 (no bg merging) / 4.4.0 (with bg merging)

¹³¹ CoatJava 6.5.8 (no bg merging) / 6.5.9 (with bg merging)

₁₃₂ **Chapter 1**

₁₃₃ **Physics motivations**

₁₃₄ Generalized Partons Distributions (GPD) have been experimentally studied mainly through Deeply
₁₃₅ Virtual Compton Scattering (DVCS) polarization observables. Such observables are mainly sensitive to
₁₃₆ the imaginary part of the Compton Form Factors (CFF), which are quantities directly related to GPDs.
₁₃₇ To access the real part of CFFs, DVCS doubly polarized beam-target asymmetries or unpolarized cross
₁₃₈ sections are needed. It is also possible to access the real part of CFFs through the time-reversal
₁₃₉ symmetric process of the DVCS: the Timelike Compton Scattering (TCS)[1, 2, 3].

₁₄₀ Measuring TCS observables also provides a test for the universality of GPDs. Indeed photon-
₁₄₁ polarization dependent cross section of TCS is sensitive to the imaginary part of CFFs. Comparing
₁₄₂ the results obtained in TCS and DVCS will help proving that GPDs are universal functions and are
₁₄₃ not only related to DVCS. In this chapter, the motivations to measure TCS are presented.

₁₄₄ **1.1 Phenomenology of TCS**

₁₄₅ TCS is the time-reversal symmetric process of DVCS. The reaction of interest is $\gamma p \rightarrow \gamma^* p'$, where
₁₄₆ the incoming photon is real ($Q^2 = 0$) and the outgoing photon is virtual. The virtual photon decays
₁₄₇ in a lepton pair which can be detected. The full reaction is therefore $\gamma p \rightarrow p' l^+ l^-$. Note that in the
₁₄₈ following we refer to the $\gamma p \rightarrow p' l^+ l^-$ reaction directly as the *TCS reaction*. Contrary to DVCS, where
₁₄₉ the large spacelike virtuality of the incoming photon gives a hard scale which ensures factorization, the
₁₅₀ TCS hard scale is given by the timelike virtuality Q'^2 of the outgoing photon. For large Q'^2 such that
₁₅₁ $\frac{t}{Q'^2} \ll 1$, factorization can be applied. The real incoming photon scatters off a single quark, which
₁₅₂ emits a virtual photon. The leading order, leading twist diagram for TCS is given in Figure 1.1.

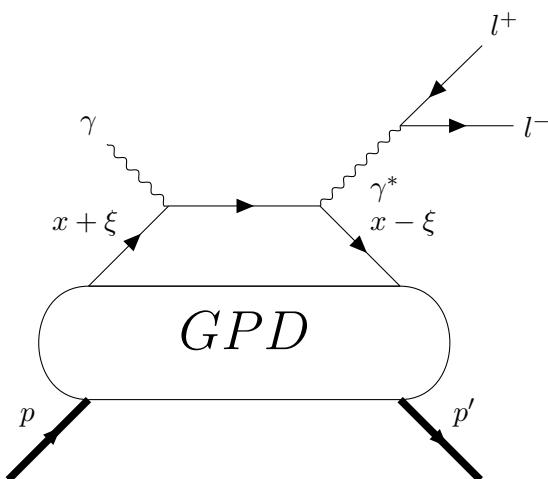


Figure 1.1: The Feynman diagram for TCS in the reaction $\gamma p \rightarrow p' l^+ l^-$

₁₅₃ The relevant variables to study the TCS reaction are shown in Figure 1.2. They are the virtuality
₁₅₄ of the outgoing photon $Q'^2 = (k + k')^2$, the transferred momentum to the nucleon $t = (p' - p)^2$, the

square of the γp Center-of-Mass (COM) energy $s = (p + q)^2$ (or equivalently the real photon energy E_γ), the azimuthal angle ϕ between the leptonic plane and the hadronic plane, and the angle of the outgoing electron in the lepton COM frame, θ . Finally, as for DVCS, the quark momentum imbalance can be defined as:

$$\xi = \frac{\tau}{2 - \tau}, \quad (1.1)$$

where $\tau = Q'^2/(s - m_p^2)$, which plays the symmetric role of x_B in DVCS.

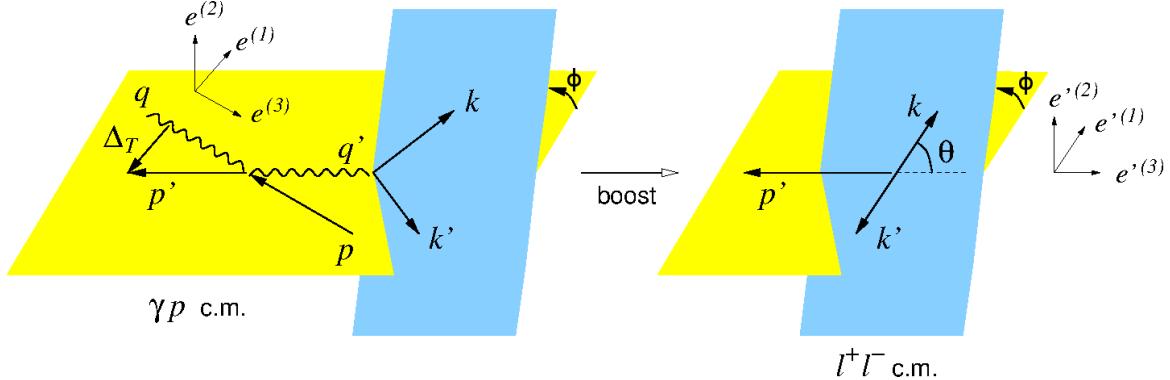


Figure 1.2: Frame definition and relevant variables for TCS, taken from [1]. The yellow plane containing the momenta of the target and recoil protons is called the hadronic plane. The blue plane containing the momenta of the two leptons is the leptonic plane. The angle between these planes is called ϕ . The angle θ is defined as the angle between the lepton with momentum k and the direction of the recoil proton momentum in the lepton pair COM frame.

As for DVCS, a Bethe-Heitler (BH) process also contributes to the $\gamma p \rightarrow p' l^+ l^-$ reaction and interferes with TCS. Figure 1.3 displays the two timelike BH diagrams, where the real photon decays in a lepton pair, from which one lepton then interacts with the proton.

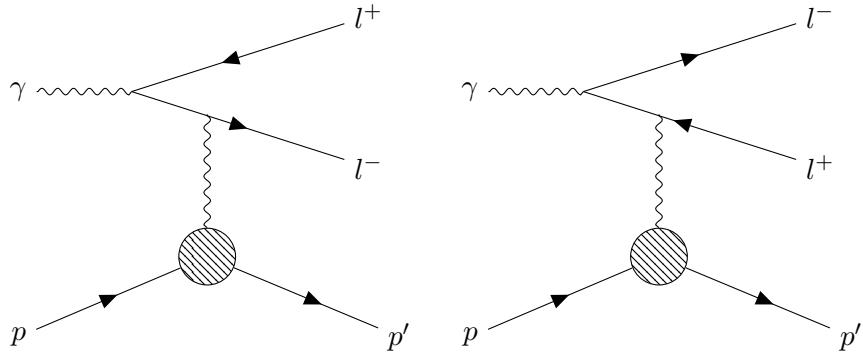


Figure 1.3: Feynman diagrams for the Bethe-Heitler processes that contribute to the $\gamma p \rightarrow p' l^+ l^-$ reaction.

1.1.1 The CFFs of TCS

Similarly to DVCS, the closed quark line in the TCS diagram forces GPDs to appear inside CFFs in the amplitude. The TCS CFFs have similar forms to the DVCS CFFs up to a sign and a complex conjugation. The two kinds of CFFs are related by the following relations:

$$\begin{aligned} \mathcal{H}_{TCS} &= \mathcal{H}_{DVCS}^* & \tilde{\mathcal{H}}_{TCS} &= -\tilde{\mathcal{H}}_{DVCS}^* \\ \mathcal{E}_{TCS} &= \mathcal{E}_{DVCS}^* & \tilde{\mathcal{E}}_{TCS} &= -\tilde{\mathcal{E}}_{DVCS}^* \end{aligned} \quad (1.2)$$

164 In this chapter, we use TCS CFFs unless specified otherwise.

165 **1.1.2 TCS cross section**

166 **Unpolarized cross section** The unpolarized cross section for $\gamma p \rightarrow p' l^+ l^-$ can be expanded as:

$$d^4\sigma(\gamma p \rightarrow p' e^+ e^-) = d^4\sigma_{BH} + d^4\sigma_{TCS} + d^4\sigma_{INT}. \quad (1.3)$$

167 Each term is written explicitly according to the formulas given in [1] in the following. The BH cross
168 section is parametrized by FFs only. It reads:

$$\frac{d^4\sigma_{BH}}{dQ'^2 dt d(\cos\theta) d\phi} = \frac{\alpha_{em}^3}{4\pi(s - m_p^2)^2} \frac{\beta}{-tL} \left[(F_1^2 - \frac{t}{4m_p^2} F_2^2) \frac{A}{-t} + (F_1 + F_2)^2 \frac{B}{2} \right], \quad (1.4)$$

169 where

$$\begin{aligned} A &= (s - m_p^2)^2 \Delta_T^2 - t(a + b) - m_p^2 b^2 - t(4m_p^2 - t) Q'^2 \\ &\quad + \frac{m_l^2}{L} \left[\left\{ (Q'^2 - t)(a + b) - (s - m_p^2) b \right\}^2 + t(4m_p^2 - t)(Q'^2 - t)^2 \right] \\ B &= (Q'^2 + t)^2 + b^2 + 8m_l^2 Q'^2 - \frac{4m_l^2(t + 2m_l^2)}{L} (Q'^2 - t)^2, \\ \beta &= \sqrt{1 - 4m_l^2/Q'^2}, \end{aligned} \quad (1.5)$$

170 with m_l the lepton mass and

$$a = 2(k - k') \cdot p', \quad (1.6)$$

$$b = 2(k - k') \cdot (p - p'), \quad (1.7)$$

$$L = \frac{(Q'^2 - t)^2 - b^2}{4}. \quad (1.8)$$

171 The BH cross section is plotted in Figure 1.4 for different θ and ϕ . One can see that the cross
172 section is largely enhanced around $\phi = 0^\circ$ for high values of θ , and around $\phi = 180^\circ$ for low values of
173 θ .

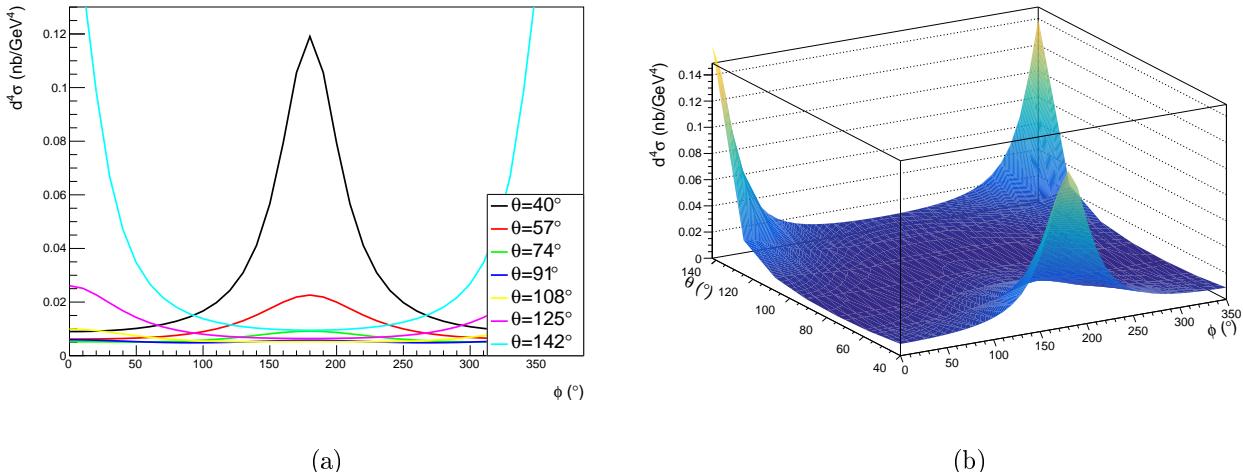


Figure 1.4: BH cross section at the kinematic point $Q'^2 = 4 \text{ GeV}^2$, $\xi = 0.1$ and $-t = 0.2 \text{ GeV}^2$, (a) for different values of θ as a function of ϕ , (b) as a function of both θ and ϕ .

174 The TCS contribution reads:

$$\frac{d\sigma_{TCS}}{dQ'^2 dt d(\cos \theta) d\phi} \approx \frac{\alpha_{em}^3}{8\pi s^2} \frac{1}{Q'^2} \frac{1 + \cos^2 \theta}{4} \sum_{\lambda, \lambda'} |M^{\lambda' -, \lambda -}|^2, \quad (1.9)$$

175 where

$$\begin{aligned} \frac{1}{2} \sum_{\lambda, \lambda'} |M^{\lambda' -, \lambda -}|^2 &= (1 - \xi^2) (|\mathcal{H}|^2 + |\tilde{\mathcal{H}}|^2) - 2\xi^2 \operatorname{Re} (\mathcal{H}^* \mathcal{E} + \tilde{\mathcal{H}}^* \tilde{\mathcal{E}}) \\ &\quad - \left(\xi^2 + \frac{t}{4M^2} \right) |\mathcal{E}|^2 - \xi^2 \frac{t}{4M^2} |\tilde{\mathcal{E}}|^2. \end{aligned} \quad (1.10)$$

As shown in [2, 1], the TCS contribution to the total cross section is two orders of magnitude less important than the BH one. Therefore measuring the TCS cross section is very challenging. One way to avoid this issue is to extract the BH-TCS interference term. This term reads:

$$\begin{aligned} \frac{d^4 \sigma_{INT}}{dQ'^2 dt d\Omega} &= -\frac{\alpha_{em}^3}{4\pi s^2} \frac{1}{-t} \frac{m_p}{Q'} \frac{1}{\tau \sqrt{1-\tau}} \frac{L_0}{L} [\cos(\phi) \frac{1 + \cos^2(\theta)}{\sin(\theta)} \operatorname{Re} \tilde{M}^{--} \\ &\quad - \cos(2\phi) \sqrt{2} \cos(\theta) \operatorname{Re} \tilde{M}^{0-} + \cos(3\phi) \sin(\theta) \operatorname{Re} \tilde{M}^{+-} + O(\frac{1}{Q'})], \end{aligned} \quad (1.11)$$

176 where

$$L_0 = \frac{Q'^2 \sin^2(\theta)}{4}, \quad (1.12)$$

177 and the \tilde{M} terms are CFFs combinations defined in [4]. At leading order and leading twist, the only 178 term contributing to the cross section is the one proportional to \tilde{M}^{--} . The CFF dependence of \tilde{M}^{--} 179 is given by:

$$\tilde{M}^{--} = \frac{2\sqrt{t_0 - t}}{m_p} \frac{1 - \xi}{1 + \xi} \left[F_1 \mathcal{H} - \xi(F_1 + F_2) \tilde{\mathcal{H}} - \frac{t}{4m_p^2} F_2 \mathcal{E} \right], \quad (1.13)$$

180 where

$$t_0 = \frac{4\xi^2 m_p^2}{(1 - \xi^2)}. \quad (1.14)$$

181 Furthermore, the dominant term in Equation (1.13) at JLab kinematics is the one containing the CFF 182 \mathcal{H} ($\tilde{\mathcal{H}}$ is suppressed by a factor $\xi \approx 0.3$, \mathcal{E} is suppressed by a factor $\frac{t}{4m_p^2} < 0.25$). The $\cos(\phi)$ modulation 183 of the interference cross section thus gives direct access to the real part of the CFF \mathcal{H} .

184 **Transversely polarized photon cross section** As mentionned for DVCS, the helicity-spin observables are a powerful tool to extract the imaginary part of CFFs. The same reasoning can be applied 185 to TCS. In the case of transversely polarized photons, the interference cross section is expressed as: 186

$$\begin{aligned} \frac{d\sigma_{INT}}{dQ'^2 dt d(\cos \theta) d\phi} &= \frac{d\sigma_{INT}|_{\text{unpol.}}}{dQ'^2 dt d(\cos \theta) d\phi} - \nu \frac{\alpha_{em}^3}{4\pi s^2} \frac{1}{-t} \frac{M}{Q'} \frac{1}{\tau \sqrt{1-\tau}} \frac{L_0}{L} \left[\sin(\phi) \frac{1 + \cos^2(\theta)}{\sin(\theta)} \operatorname{Im} \tilde{M}^{--} \right. \\ &\quad \left. - \sin(2\phi) \sqrt{2} \cos(\theta) \operatorname{Im} \tilde{M}^{0-} + \sin(3\phi) \sin(\theta) \operatorname{Im} \tilde{M}^{+-} + O\left(\frac{1}{Q'}\right) \right], \end{aligned} \quad (1.15)$$

187 where ν is the circular polarization of the incoming real photon. The additional polarization term 188 exhibits the same CFF content as the unpolarized cross section, except that it now depends on the 189 imaginary parts via $\sin(n\phi)$ factors. Extracting the $\sin(\phi)$ component of the polarized cross section 190 enables to access the imaginary part of \mathcal{H} . This is an important test of the universality of GPDs once 191 compared with DVCS data.

192 1.2 TCS observables

193 In this section, the TCS observables extracted in this analysis are presented. These observables are 194 the R ratio and the Forward/Backward asymmetry sensitive to the real parts of the CFFs, and the 195 photon polarization asymmetry sensitive to the imaginary parts of the CFFs.

196 1.2.1 R ratio

197 The R ratio, introduced in [1], is defined as:

$$R(\sqrt{s}, Q'^2, t) = \frac{\int_0^{2\pi} d\phi \cos(\phi) \frac{dS}{dQ'^2 dt d\phi}}{\int_0^{2\pi} d\phi \frac{dS}{dQ'^2 dt d\phi}}, \quad (1.16)$$

198 where

$$\frac{dS}{dQ'^2 dt d\phi} = \int_{\pi/4}^{3\pi/4} d\theta \frac{L}{L_0} \frac{d\sigma}{dQ'^2 dt d\phi d\theta}. \quad (1.17)$$

199 It has to be noted that the definition used in this manuscript differs by a factor 2 from the original
200 definition. This was chosen to be consistent with the CLAS TCS analysis in [3].

201 The R ratio is directly sensitive to the real part of \tilde{M}^{--} . The integration domain is set to $[\pi/4, 3\pi/4]$
202 to avoid kinematic regions where TCS is too small compared to BH.

203 1.2.2 Forward-Backward asymmetry

204 The idea of Forward-Backward Asymmetry (A_{FB}) was initially proposed for J/ψ threshold photo-
205 production studies in [5]. The A_{FB} is defined as:

$$A_{FB}(\theta, \phi) = \frac{d\sigma(\theta, \phi) - d\sigma(180^\circ - \theta, 180^\circ + \phi)}{d\sigma(\theta, \phi) + d\sigma(180^\circ - \theta, 180^\circ + \phi)} \quad (1.18)$$

206 where only the θ - ϕ dependence of the cross section is explicitly written.

207 The transformation ($\theta \rightarrow 180^\circ - \theta, \phi \rightarrow \phi + 180^\circ$) corresponds to inverting the vectors k and k' in the
208 COM frame of the lepton pair. This transformation leaves both L (Equation (1.8)) and L_0 (Equation
209 (1.12)) unchanged, as the term b appears squared in L . Moreover both $\frac{d^4\sigma_{BH}}{dQ'^2 dt d\Omega}$ and $\frac{d^4\sigma_{TCS}}{dQ'^2 dt d\Omega}$ remain
210 unchanged under this transformation. Assuming leading order and leading twist, only the $\cos(\phi)\tilde{M}^{--}$
211 term contributes in Equation (1.11) and the interference cross section is transformed as:

$$\frac{d^4\sigma_{INT}}{dQ'^2 dt d\Omega} \xrightarrow{FB} -\frac{d^4\sigma_{INT}}{dQ'^2 dt d\Omega}. \quad (1.19)$$

212 Finally the A_{FB} can be explicitly written as:

$$A_{FB}(\theta_0, \phi_0) = \frac{-\frac{\alpha_{em}^3}{4\pi s^2} \frac{1}{-t} \frac{m_p}{Q'} \frac{1}{\tau\sqrt{1-\tau}} \frac{L_0}{L} \cos\phi_0 \frac{(1+\cos^2\theta_0)}{\sin(\theta_0)} \text{Re}\tilde{M}^{--}}{d\sigma_{BH}}, \quad (1.20)$$

213 where we neglect the TCS contribution in the denominator. This observable is sensitive to the same
214 quantity as the R ratio. However, it is not integrated over a large phase space. It is therefore less
215 sensitive to detector acceptance effects. First predictions for the TCS A_{FB} , realized with the VGG
216 model, are presented in Section 3.7.

217 1.2.3 Photon polarization asymmetry

218 The photon polarization asymmetry $A_{\odot U}$ (also referred in the following as Beam Spin Asymmetry
219 (BSA) because of its similarity with the DVCS BSA) is defined as:

$$A_{\odot U} = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-}, \quad (1.21)$$

220 where indexes $+(-)$ refer to the right(left)-handed circular polarization of the incoming real photon,
221 and U to the unpolarized target. It can be written explicitly as:

$$A_{\odot U} = \frac{-\frac{\alpha_{em}^3}{4\pi s^2} \frac{1}{-t} \frac{m_p}{Q'} \frac{1}{\tau\sqrt{1-\tau}} \frac{L_0}{L} \sin\phi \frac{(1+\cos^2\theta)}{\sin(\theta)} \text{Im}\tilde{M}^{--}}{d\sigma_{BH}}. \quad (1.22)$$

222 Projections for this observable have been made in [2].

223 **Chapter 2**

224 **Analysis strategy, particle identification
225 and momentum corrections**

226 In this Chapter, we first introduce the general strategy of the analysis. Then the particle identifica-
227 tion framework is briefly presented. A dedicated positron identification algorithm based on multivariate
228 analysis was developed for this analysis. It is presented in a specific section, including various quality
229 checks of the approach as well as the determination of the remaining π^+ contamination.

230 This chapter also describes momentum corrections developed to correct CLAS12 data and simula-
231 tions. These corrections have been developed in the perspective of the TCS analysis but can be used
232 in any CLAS12 analysis. A full set of momentum corrections for protons was put in place. Subsection
233 2.4.1 presents the Monte-Carlo based corrections, and Subsection 2.4.2 the data-driven corrections in
234 the CD. Momentum corrections for leptons are also presented in Section 2.5. Fiducial cuts for leptons
235 are presented in Section 2.6. Finally, a data driven proton efficiency correction is presented in Section
236 2.8.

237 **2.1 Analysis strategy for TCS**

238 The first step of the analysis is to select the $\gamma p \rightarrow p' e^+ e^-$ reaction among all events recorded by
239 CLAS12. In order to select these events, the final-state particles must be identified. The pid given by
240 the Event Builder (EB) of CLAS12 is used to filter out potential events with the required final state
241 particles:

- 242 • exactly one proton,
- 243 • exactly one electron,
- 244 • exactly one positron,
- 245 • any other particles.

246 The event builder algorithm of CLAS12 is briefly presented in Section 2.2 of this note. An enhanced
247 positron identification algorithm, crucial in the TCS measurement, is also presented in this chapter.
248 This positron PID is applied right after the EB pid, in order to discard pions which have been mis-
249 identified as positron by the EB. This specific positron algorithm is explained in details in Section
250 2.3.

251 The second step of the analysis consists in correcting the data. These corrections aim at matching
252 the measured momenta of the detected particles with their actual momenta at the vertex.

253 Once the final state particles are well identified and their momenta are corrected, exclusivity cuts
254 are applied in order to make sure that each event is a TCS event (i.e. from the $\gamma p \rightarrow p' e^+ e^-$ reaction).
255 The observables are then computed with the kinematic variables of these good events. This step is
256 presented in Chapter3.

257 2.2 Particle identification

258 In this analysis, the standard CLAS12 EB pid was used in order to select events with the good final
259 state. The EB algorithm is described in details in the RG-A analysis note. In particular, we rely on
260 the proton and lepton pid from the EB.

261 For proton, the identification is based on the comparison of the measured time-of-flight (tof) to the
262 theoretical one. The measured tof t_{tof} of a hadron is given by:

$$263 \quad t_{tof} = t_{TOF} - t_S, \quad (2.1)$$

263 where t_{TOF} is the time associated with the detector in which the hadron was detected and t_S is the
264 start time given event by event (see RG-a note for full details). The expected tof from tracking t_{track}
265 is given by:

$$266 \quad t_{track} = \frac{P_L \sqrt{p^2 + m^2}}{pc}, \quad (2.2)$$

266 where m is the mass of the particle, p its momentum and P_L the path length from the vertex to the
267 interaction point in the TOF detector obtained from tracking. The mass hypothesis which minimizes
268 the difference $t_{tof} - t_{track}$ is assigned to the particle.

269 Electrons and positrons have very low mass and will likely be detected in the FD of CLAS12. The
270 timing resolution of the FTOF does not allow to separate leptons and pions at the CLAS12 kinematics.
271 As a consequence, the pid of leptons is solely based on the Sampling Fraction (SF) of the EC and the
272 number of photo-electrons in the HTCC. The SF is defined as:

$$273 \quad SF = \frac{E_{dep}}{P}, \quad (2.3)$$

273 where E_{dep} is the total energy deposited in the EC, and P the momentum measured by the DCs. The
274 EB of CLAS12 assigns the particle ID, in the Lund convention, ± 11 (i.e. electron or positron) for
275 particles that fulfill all the following requirements:

- 276
- 277 • A track in the DCs and an associated EC shower (the curvature of the track in the torus magnetic
field gives the charge of the particle)

278

 - A minimum deposited energy in the PCAL: $E_{PCAL} > 60$ MeV

279

 - A total measured SF, $SF_M(E_{dep})$, within 5σ of the parametrized SF (see common analysis note
for more details).

281

 - a minimum number of HTCC photo-electrons: $N_{PHE}(HTCC) > 2$.

282 The quality of pid provided by the EB is assessed via the pid χ^2 given for each particle. The pid χ^2
283 is defined as the deviation of the identifying quantity (tof for proton, SF for leptons) to its expected
284 value, normalized by its standard deviation (more details is given in the RG-A common analysis note).
285 One can cut on the χ^2 in order to improve the pid quality. In this analysis, we decided to only use
286 the standard EB cut on the χ^2 to maximize the number of events available to calculate the TCS
287 observables. The distributions of pid χ^2 for the three TCS final state particles (and separated between
288 the CD and FD for protons) is shown in Figure 2.1. The distributions on the first row are obtained
289 after selecting events with the good final state. The second row shows only events which are used for
290 the final analysis (after exclusivity cuts and for the phase space of interest, see Chapter 3 for more
291 details). One can see that events used in the final analysis have low χ^2 , and that no further cuts are
292 needed.

293

294 *Start of common analysis section*

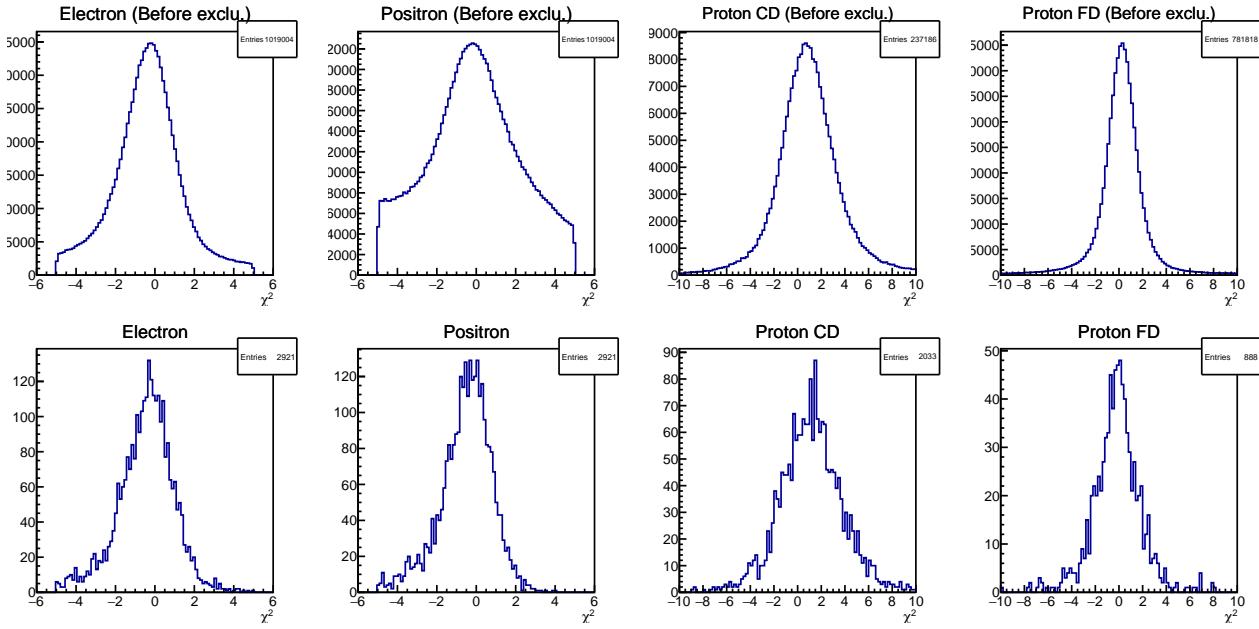


Figure 2.1: Pid χ^2 for the three final state particles of the TCS reaction. In the proton case, the protons detected in the CD or the FD are shown separately. One can see that the selected events have low χ^2 ensuring a good pid quality.

2.3 Positron identification

Measurements of TCS observables require the detection of a pair of leptons from the decay of a virtual photon. For momenta lower than 4.9 GeV, the HTCC of CLAS12 provides good pion/lepton separation [6]. The HTCC electron detection efficiency is estimated to be higher than 99%. For momenta higher than the HTCC threshold (4.9 GeV) data and simulation show a large contamination of π^+ in positron samples. In this section, evidence for pion contamination is shown and a multivariate analysis is proposed to reduce this contamination. The results of this approach are shown and assessed. An estimate of the remaining pion contamination is given.

In the following we consider particles that have been assigned ID -11 and identified as positrons by the EB. The electron case is treated at the end of this section.

2.3.1 Evidence of π^+ contamination

In order to measure TCS observables, we aim at achieving a clean identification of leptons. The positron ID is crucial to reach this goal. However the standard CLAS12 EB pid cuts are not sufficient to remove potential mis-identified π^+ from the positron sample. This contamination can be seen in both the data and simulations.

We use the expression "true-positron" for Generated positron-reconstructed positron and "mis-id. pion" for Generated pion-reconstructed positron in the simulation samples. For the data samples, analogous designations are used, replacing "Generated" by "Produced". Finally, we also refer to these categories as respectively "Signal" and "Background" when methods to distinguish both sets are described.

π^+ contamination from data

Positron momenta spectrum A first evidence of pion contamination is seen by investigating events with an electron, a proton, a positron and any other particles in CLAS12. One can see that there is an excess of positrons above the HTCC threshold. This is visible in Figure 2.2 where the polar angle of the positrons are plotted against their momenta.

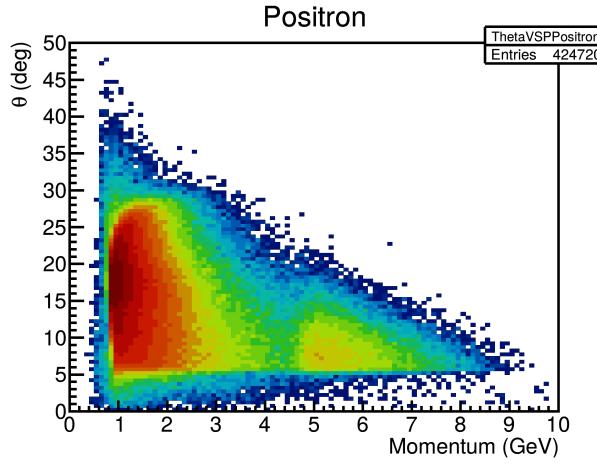


Figure 2.2: Polar angle θ versus momentum of positrons for events with one electron, one proton and one positron in CLAS12.

320 **Exclusive reaction** A second evidence of pion contamination is seen when investigating the exclusive
 321 reaction $ep \rightarrow e\pi^+n$. Events with at least a particle with ID -11 and momentum bigger than 4.4 GeV
 322 and an electron with momentum lower than 4.4 GeV were filtered from the CLAS12 dataset. Both
 323 particles are required to be detected in the FD. Cuts on the electron momentum ensure that it is a
 324 true electron. The particles with ID -11 are assigned the π mass. The missing mass of the system
 325 $ep \rightarrow ee_{m_\pi}^+ X$ is then calculated. The notation $e_{m_\pi}^+$ is used to refer to the particle with ID -11 and mass
 326 equal to the π mass. The missing mass spectrum obtained is shown in Figure 2.3.

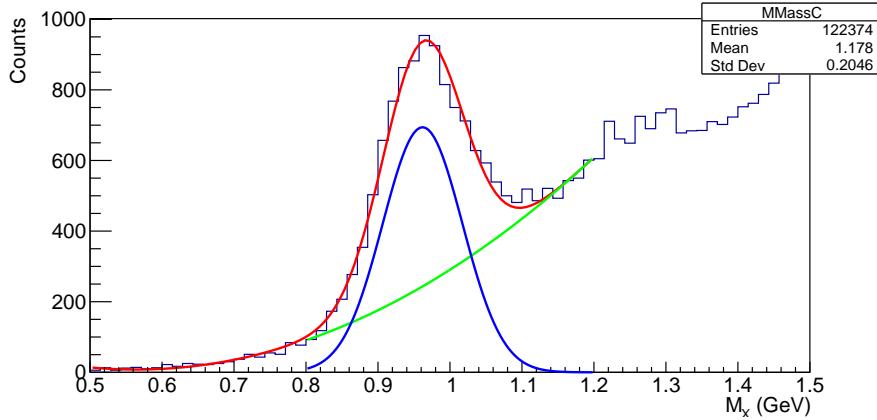


Figure 2.3: Missing mass spectrum of the reaction $ep \rightarrow ee_{m_\pi}^+ X$ in the neutron mass region.

327 A clear peak at the neutron mass is visible. This peak is produced by the reaction $ep \rightarrow e\pi^+n$ where
 328 the π^+ has been identified as a positron. These exclusive mis-identified pion events are a good way
 329 to quantify the pion contamination. Furthermore the momenta of the mis-id. π^+ cover most of the
 330 momentum range of interest, from 4.9 GeV to 10.6 GeV as shown in Figure 2.4. We will use these
 331 events later in Subsection 2.3.5 as a scale to quantify the pion background.

332 π^+ contamination from simulations

333 The π^+ contamination is also visible in simulations. Two test samples were generated, one sample
 334 of positrons and one of positively charged pions. They were passed to the GEMC CLAS12 simulation
 335 and reconstructed using the standard CLAS12 software. Particles were simulated within the CLAS12

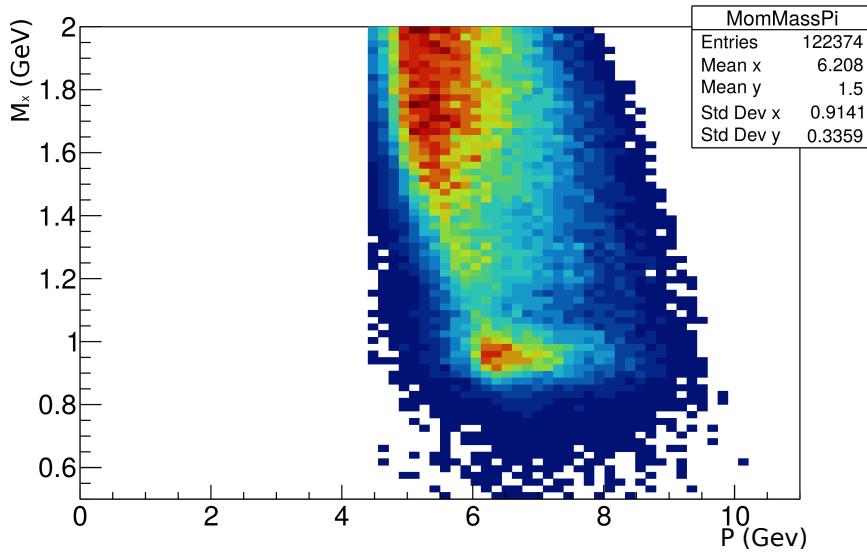


Figure 2.4: Mass of the missing particle versus momentum of the particle with ID -11 .

acceptance and within the range of momenta of interest ($4 \text{ GeV} < P < 10.6 \text{ GeV}$ and transverse momentum within $0.5 \text{ GeV} < P_t < 2 \text{ GeV}$). The output of both samples were then skimmed identically: only particles with ID -11 were kept. The kinematics distributions for both true-positrons and mis-id. pions are shown in Figure 2.5.

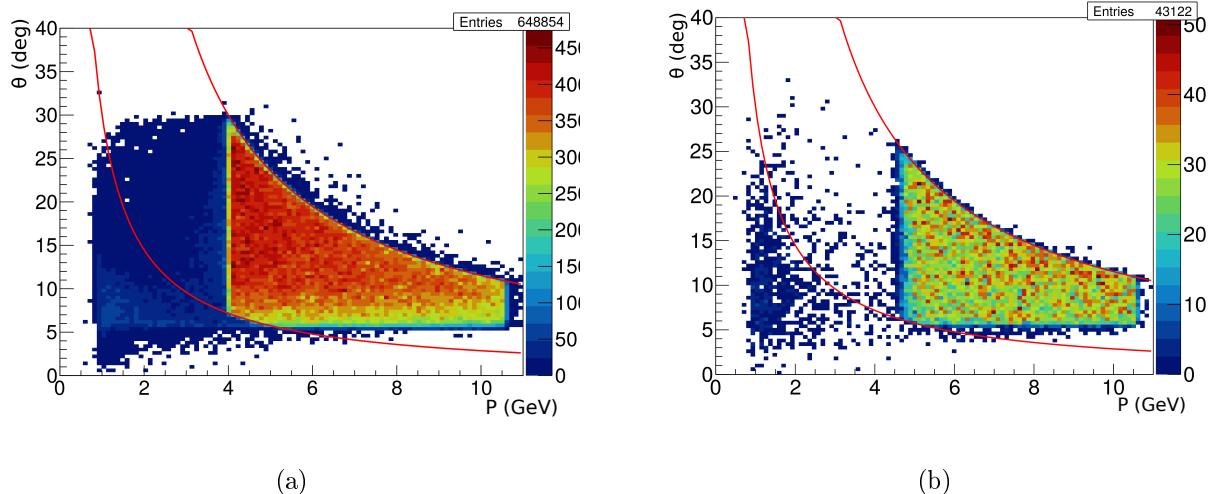


Figure 2.5: (a) Polar angle θ versus momentum of the simulated true-positrons (b) Polar angle θ versus momentum of the simulated mis-id. pions. The red lines represent the transverse momentum limits ($0.5 \text{ GeV} < P_t < 2 \text{ GeV}$) applied on the generated particles. The positrons detected at low momenta in Figure (a) are positrons which radiated a photon in the target material. The momenta of these positrons are corrected when the radiated photons are detected (see Section 2.5.2).

The behavior observed in the data (excess of positrons due to contamination from π^+) is reproduced in the simulations.

2.3.2 1D and 2D cuts from the simulations

The results of the simulations described above were used to explore simple cuts to try to remove the pion contamination. In this section, various pid cuts based on these simulations are defined.

345 **χ^2 cut**

346 The CLAS12 EB gives the deviation of the total sampling fraction from a parametrized model. This
 347 quantity is referred as χ^2 in the CLAS12 software (although the name commonly used in literature is
 348 *pull* value) and it is defined as:

$$\chi^2 = \frac{SF_M(E_{dep}) - SF_P(E_{dep})}{\sigma_P}. \quad (2.4)$$

349 The EB requires that $-5 < \chi^2 < 5$ to identify a particle as a lepton. The χ^2 of true-positrons and
 350 mis-id. pions is shown in Figure 2.6. One can see that mis-id. pions mostly populate the low χ^2
 351 region. Their sampling fraction is just big enough for them to be identified as positrons. From these
 352 distributions, two simple strategies can be tested: cutting on the absolute value of χ^2 ($|\chi^2| < c$) or
 353 cutting on low values of χ^2 ($c < \chi^2$). These strategies are referred as Symmetric and Asymmetric χ^2
 354 cuts in the following.

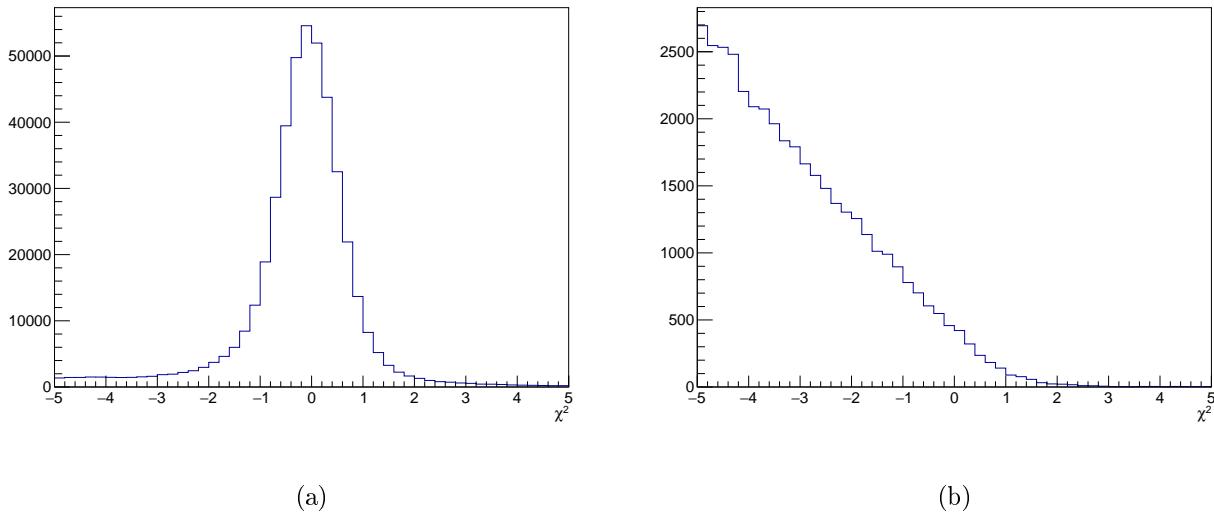


Figure 2.6: Sampling fraction χ^2 of (a) true-positrons and (b) mis-id. pions.

355 **2D Sampling Fraction cuts**

356 Another way to distinguish positrons from π^+ is to look at their partial SFs. The total sampling
 357 fraction can be decomposed according to the three layer of the EC (PCAL, ECIN and ECOUT):

$$SF_{PCAL} = \frac{E_{dep}(PCAL)}{P}. \quad (2.5)$$

358 Analogous equations can be written for ECIN and ECOUT. The longitudinal segmentation of the EC
 359 proves useful to distinguish positrons. Positrons are more likely to deposit all their energy in the first
 360 layers of the EC (PCAL and ECIN). On the contrary π^+ are Minimum Ionizing Particles (MIPs) and
 361 are more likely to deposit energy in all the layers of the EC. Figure 2.7 shows the SF in the ECIN
 362 versus the SF in the PCAL for simulated true-positrons and mis-id. pions. One can see that a 2D cut
 363 along the anti-diagonal of the distribution can be applied to separate them. In the following this cut
 364 is referred as "SF cut".

365 **2.3.3 Multivariate analysis approaches**

366 In the previous section, simple approaches involving cuts on 1 or 2 quantities were shown. However
 367 these approaches do not allow to use the full amount of information provided by the EC. Additionally

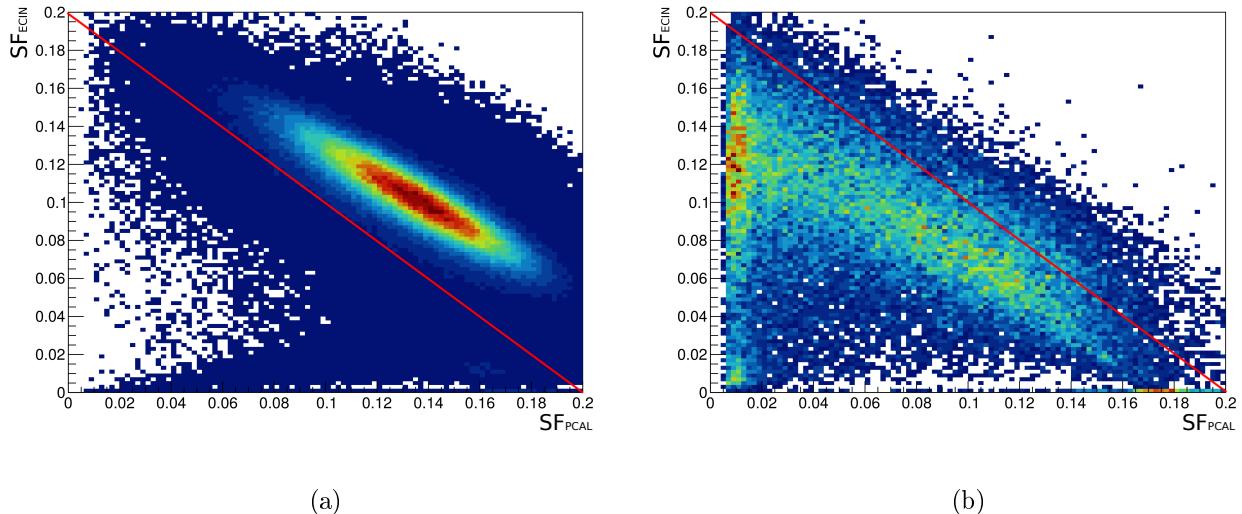


Figure 2.7: SF in the ECIN versus SF in the PCAL for simulated (a) true-positrons and (b) mis-id. pions. A possible cut to remove mis-id. pions is to keep only particles above the anti-diagonal represented by the red lines. Most mis-id. pions are removed while most positrons are kept.

368 to the SFs of all three EC layers, one can access the shape of the electromagnetic shower in each layer.
 369 The square of the width of the shower is defined for each coordinate (U,V,W) and for each layer as:

$$M_2 = \frac{\sum_{\text{strip}} (x - D)^2 \ln(E)}{\sum_{\text{strip}} \ln(E)}, \quad (2.6)$$

370 where D is the log-weighted mean position of the shower defined as:

$$D = \frac{\sum_{\text{strip}} x \ln(E)}{\sum_{\text{strip}} \ln(E)}, \quad (2.7)$$

and where x is the position of the EC hit along the considered coordinate and E is the deposited energy associated to the hit. There are potentially at least 12 variables (3 sampling fractions and 9 shower widths) to investigate to help separating π^+ and e^+ . A simple approach relying only on assessing correlations between each couple of variables one by one is not applicable here. A multivariate approach is needed. In this section some multivariate techniques are introduced. Their application to the π^+ / e^+ separation problem is presented.

The TMVA Root package [7] was used for this analysis. This package has been developed to train, test and compare a large range of multivariate tools. Multivariate Analysis (MVA) classifiers take several quantities as inputs and produce a single output on which one can then apply a cut. The value of the cut that maximizes the background rejection and the signal efficiency is then found and applied to the output distribution. In the following we only focus on three of them: Fisher discriminant, Boosted Decision Tree (BDT) and Multilayer Perceptron (MLP).

³⁸³ The MVA classifiers presented in the following were trained on the simulation samples described in
³⁸⁴ Subsection 2.3.1.

385 3D analysis

As a first step, the SFs of the three layers of the EC were considered as inputs. The distributions of these three input variables for true-positron (blue) and mis-id. pion (red) are shown in Figure 2.8. Three methods (Fisher, BDT, and MLP) were trained and tested on these three variables. This analysis is referred as 3D in the following.

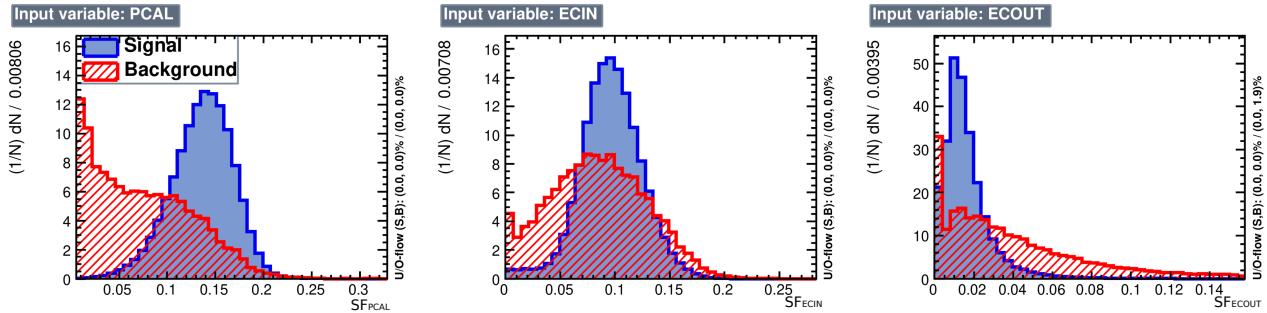


Figure 2.8: Input variables provided for the training of the multivariate classifiers. The leftmost plot shows the SF_{PCAL} distributions for the signal and the background. The middle plot SF_{ECIN} , the rightmost SF_{ECOUT} .

390 6D analysis

391 A 6D multivariate analysis was also studied. In this approach Fisher, BDT and MLP methods were
 392 applied to the three SFs and to the average width of the shower in each layer. The average of the
 393 square of the width in the PCAL is defined as:

$$M_{2/PCAL} = \frac{M_{2/U/PCAL} + M_{2/V/PCAL} + M_{2/W/PCAL}}{3} \quad (2.8)$$

394 where $M_{2/U/PCAL}$ is the square of the width of the shower along the U direction as defined in Equation
 395 (2.6). Similar equations apply for the V and W directions. The distribution of the input squared shower
 396 widths are shown in Figure 2.9. Figure 2.10 shows the architecture of the neural network used for the
 6D analysis.

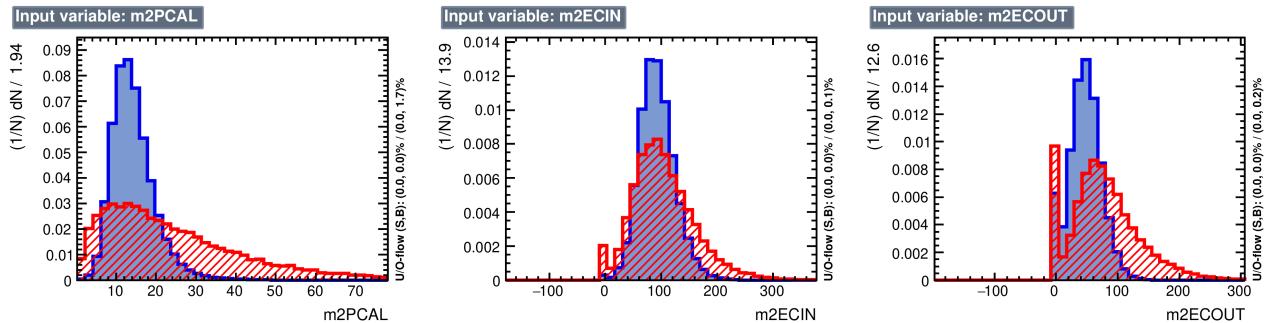


Figure 2.9: Average squared shower width for the PCAL (left), ECIN (middle), and ECOUT (right).

397 2.3.4 Training, testing and comparison of MVA classifiers on simulations

398 Fisher discriminant, BDT and MLP are trained on the sets of variables presented in the previous
 399 two subsections using the TMVA package. A first series of tests is also performed on simulations. The
 400 complete input sample is divided in two randomly selected subsets: a training and a testing sample.
 401 For each classifier, the training is performed on the training sample. The classifiers are then tested on
 402 the test sample. The distributions of classifiers outputs for the testing set are superimposed on the
 403 training distributions. An example is shown in Figure 2.11. These checks indicate that it is possible
 404 to apply reliably these trained classifiers on simulations. Training and testing outputs show good
 405 agreement for both 3D and 6D approaches and for all the classifiers. One can also look at indicators
 406 specific to each technique in order to assess the quality of the classifiers. An example is shown in
 407 Figure 2.12 where the convergence of the MLP with three input variables is shown.
 408

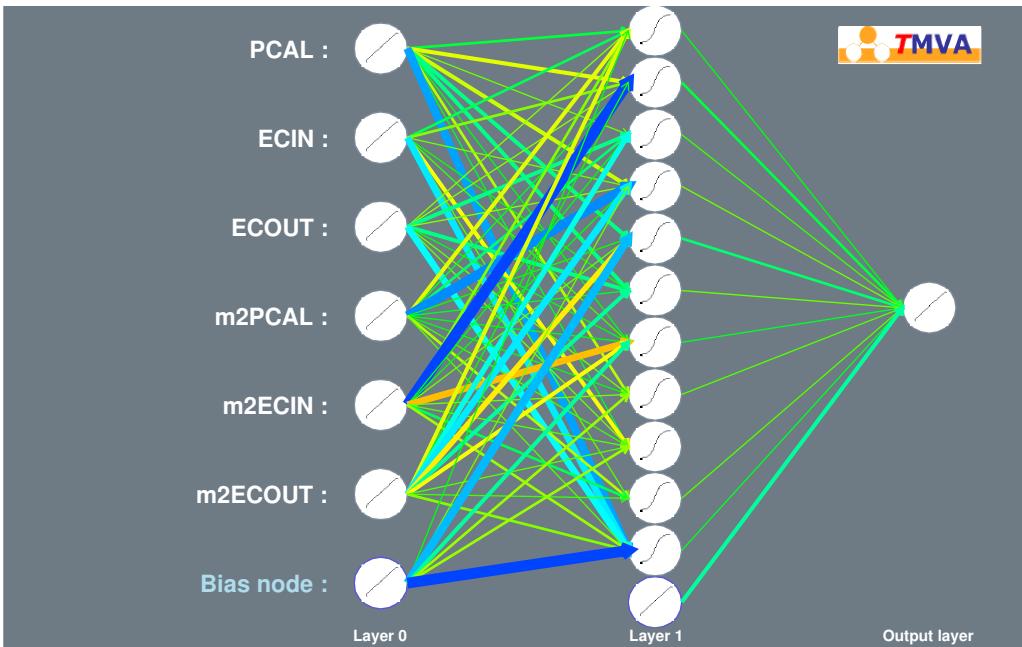


Figure 2.10: MLP using six input variables developed for the positron identification algorithm.

409 One can then compare the strength of different classifiers using Receiver Operating Characteristic
 410 (ROC) curves. ROC curves display the signal efficiency (fraction of signal kept) of a classifier versus
 411 its background rejection (fraction of background removed). The method which achieves higher signal
 412 efficiency and higher background rejection is preferred. The ROC curves for four 3D and three 6D
 413 classifiers are shown in Figure 2.13.

414 One can clearly see that 6D classifiers are more powerful than 3D ones. This behavior is expected
 415 as more information is taken into account by the classifiers. Indeed as shown in Figure 2.14, shower
 416 widths are not fully correlated with the deposited energy.

417 Figure 2.14a shows the correlations of the input variables for true positrons, while Figure 2.14b is
 418 for π^+ that are mistakenly identified as positrons. One can see that the sampling fraction in the PCAL
 419 is strongly anti-correlated with the sampling fraction in the ECIN and in the ECOUT, in both cases.
 420 This means that if the sampling fraction is high in the PCAL, it will be low in the ECIN/ECOUT;
 421 or the opposite situation. This behavior is explained by the fact that positrons tend to deposit most
 422 of their energy in a single calorimeter layer. We also observe a large correlation between the sampling
 423 fraction in the PCAL and the shower m_2 in this layer, depicting the relation between a high energy
 424 deposition and a large shower radial size.

425 2.3.5 Test and comparison of MVA classifiers on data

426 Once the classifiers were trained and tested on simulations, they were used on CLAS12 real data and
 427 their performances were compared. In this section, the method used to assess the performance of the
 428 classifiers on CLAS12 data is presented.

429 A benchmark to assess the π^+ / e^+ separation on data

430 Clean signal and background samples are needed to evaluate the performance of classifiers. In
 431 simulation samples, this is easily achievable as the nature of each simulated particle is known. In data,
 432 a prior knowledge is necessary. The background sample (mis-id. pion) is described in Subsection 2.3.1.
 433 The neutron peak events presented in Figure 2.3 were used. As shown in this figure, the missing mass
 434 spectrum is fitted with a Gaussian for the peak and with a second order polynomial for the underlying
 435 background. The integral of the Gaussian gives a scale of the number of mis-id. pions in the sample.
 436 In the following this sample is denoted "neutron sample".

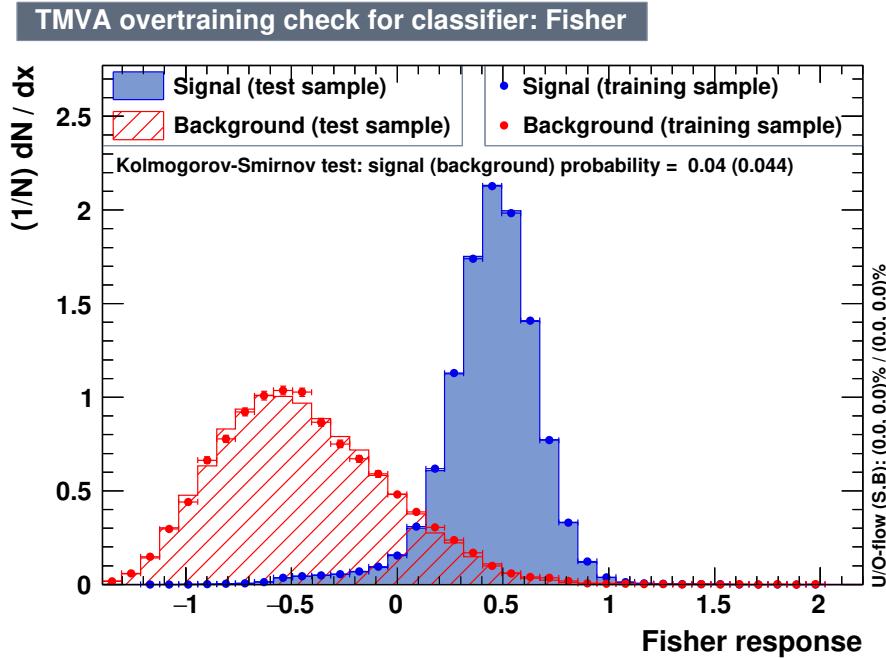


Figure 2.11: Training and testing outputs of the 3D Fisher classifier. Both distributions match very well for both the background and the signal.

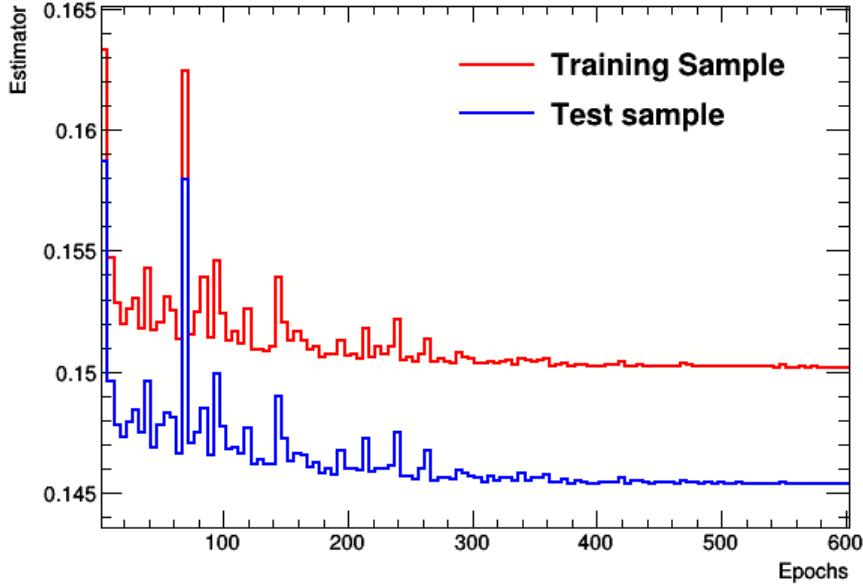


Figure 2.12: MLP 3D convergence test for the training and testing samples. The MLP error function summed over all the events in the training (red) and testing (blue) sets, respectively, is shown as a function of the training iterations. One should make sure that the convergence is reached after a certain number of training cycles.

On the contrary, defining a clean signal sample (true-positron) from data is more difficult. TCS events (at least one electron, one positron and one proton; missing transverse momentum fraction < 5%, and missing mass squared < 0.4 GeV², ensuring photo-production) with a positron with momentum bigger than 4 GeV were used and are referred as the "TCS sample" (see Chapter 3). This sample is not completely clean, it should a priori be a mixture of true-positrons and mis-id. pions. However the

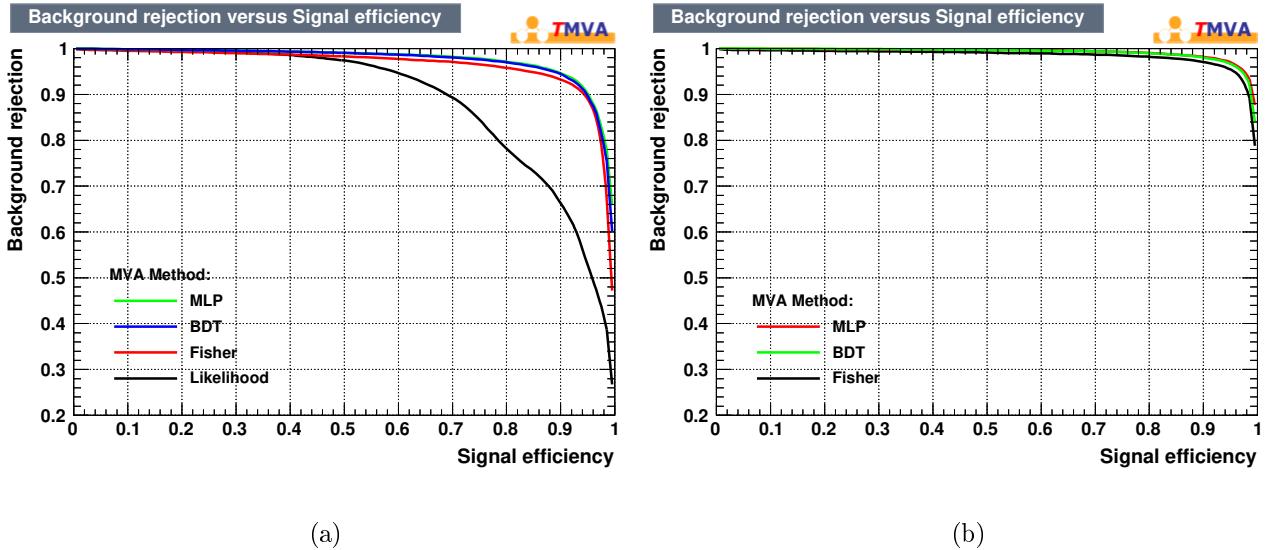


Figure 2.13: ROC curves for four classifiers (a) in the 3D case and (b) in the 6D case. In the 3D case, a Likelihood classifier was also tested, but not kept for later analysis due to its poor performance. The ROC curves of the 6D classifiers are consistently above 3D ones.

requirement of the e^-e^+p final state as well as the exclusivity cuts should enhance the true-positron fraction. This hypothesis is tested in the following section. To evaluate the "Signal+Background" in the TCS sample, the number of events is counted.

Once the two data test samples are defined, the different cuts presented in the previous sections are applied. For χ^2 and MVA approaches, one can vary the cut applied on the output variable and compare the number of remaining events in the neutron sample and in the TCS sample. This is realized by varying the cuts on the output variable in the specific output range and evaluating the integral of the neutron peak and the number of remaining TCS events. An example of the results of this procedure is shown in Figure 2.15, where the neutron peak integral and the number of TCS events are plotted against the value of the applied cut for the 3D BDT.

One can define a pseudo-ROC curve by associating these two curves. For each value of the cut, the number of TCS events is plotted against the corresponding number of neutrons. The curve obtained is not a ROC curve as the number of TCS events is a mixture of the signal and the background. However this pseudo-ROC curve still allows to compare the proposed π^+/e^+ separation strategies and evaluate their performances.

Strategy comparison and choice

The procedure presented in the previous section allows to draw for each identification technique a pseudo-ROC curve. All the pseudo-ROC curves are plotted in Figure 2.16. Both axes have been normalized to one. The blue circle denotes the case where no cuts are applied. The pink triangle represents the PCAL/ECIN SF anti-diagonal cut described in Subsection 2.3.2.

The pseudo-ROC curves shown in Figure 2.16 exhibit two regimes. At high background strength and large number of TCS events, the curves show a linear trend. This means that for a given number of mis-id. pions removed in the neutron sample, a proportional number of them is also removed from the TCS sample. On the contrary, when the background strength approaches zero, the number of TCS events decays dramatically. This happens because the cut applied on the output of the classifier is too hard and starts to remove true-positrons and mis-id. pions indistinctly. One should apply a cut on the classifier output that maximizes the background rejection while keeping most signal (at the limit between the two regimes). Figure 2.16 also permits to select the best approach to separate positrons from pions in CLAS12 data. One should use the method with the best pion rejection power while conserving most of the signal events. This is achieved by picking the method for which the

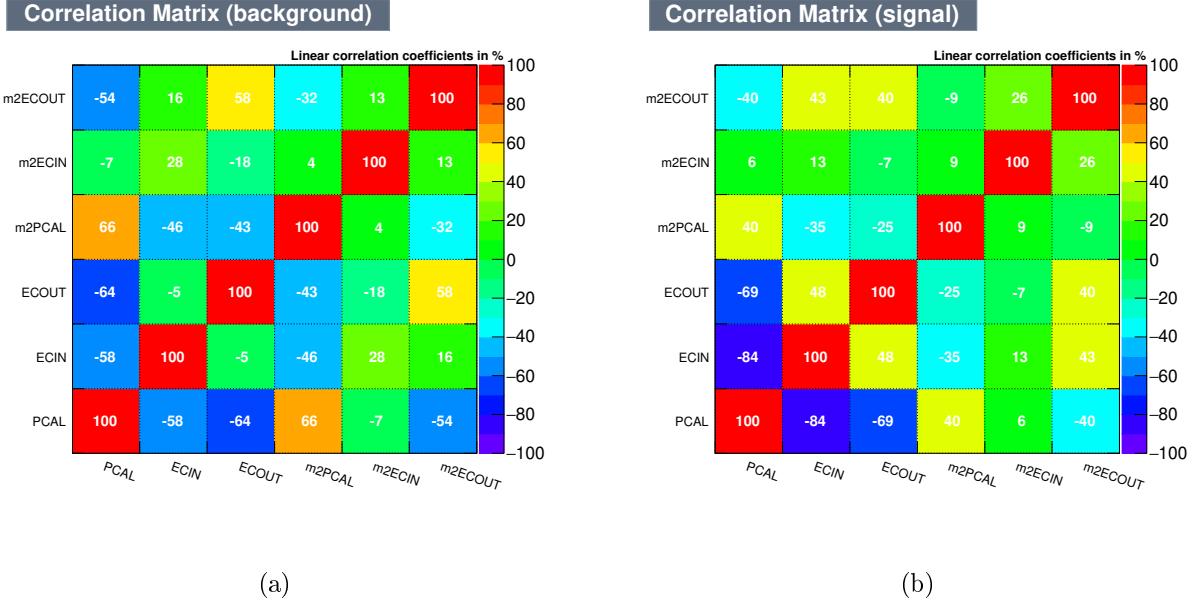


Figure 2.14: Linear correlation $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - E[X])(Y - E[Y])]}{\sqrt{E[(X - E[X])^2]}\sqrt{E[(Y - E[Y])^2]}}$ of the 6 input variables for (a) Background (mis-id. pions) and (b) Signal (true-positrons)

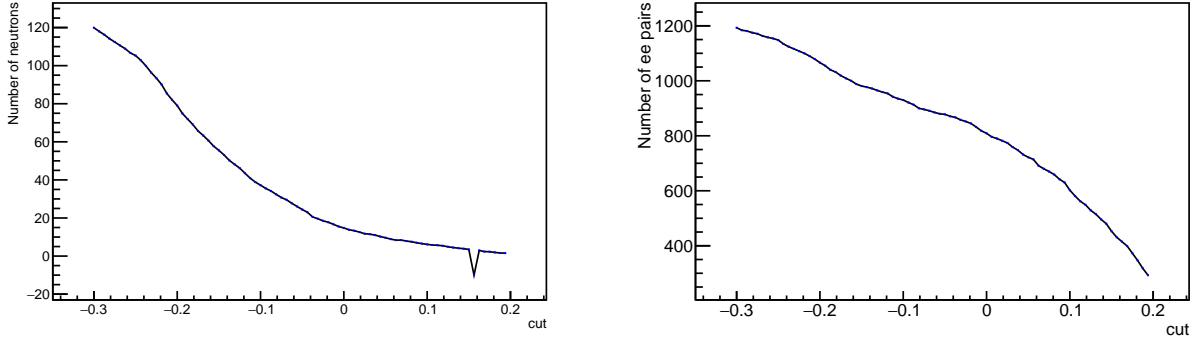


Figure 2.15: Number of neutrons as a function of the cut applied on the output of the 3D BDT (left). Number of TCS events as a function of the cut applied on the output of the 3D BDT (right).

472 pseudo-ROC curve gets closer to the (0,1) point. The MLP 6D classifier was chosen according to this
473 criterion.

474 Adding skewness to the classifier

475 It is clear from Figure 2.16 that adding the width of the shower to the analysis increases the
476 positron/pion separation efficiency. In order to test if adding more variables to the classifier training
477 further increases its power, the 3rd moment of showers in the EC, the skewness μ , was added to
478 train a MLP and a BDT. The skewness of a shower is defined as:

$$\mu = \frac{\sum_{\text{strip}} (x - D)^3 \ln(E)}{M_2^{3/2} \sum_{\text{strip}} \ln(E)}. \quad (2.9)$$

479 The pseudo-ROC curves for these two classifiers are compared to the 6D MLP and BDT in Figure
480 2.17. One can see that all curves superimpose. Adding the skewness to the inputs of the classifiers
481 does not improve their separation power.

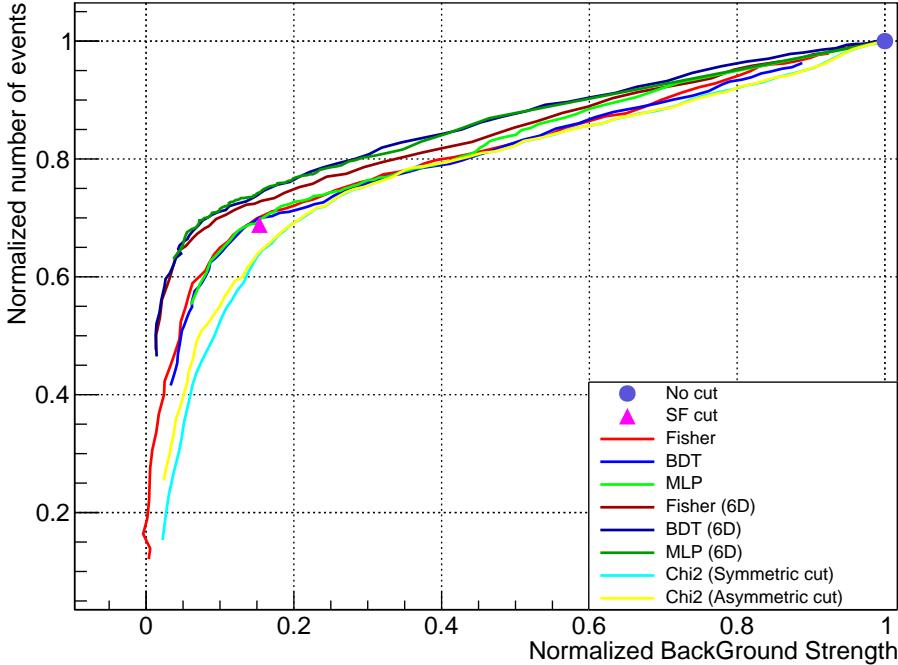


Figure 2.16: Pseudo-ROC curves obtained from data for different classifiers (3D classifiers are referred by their name only) and for the symmetric and asymmetric χ^2 cuts. The number of events in the TCS sample is plotted against the integral of the neutron peak in the neutron sample. Both axes are normalized to 1 when no cut is applied. The anti-diagonal SF cut is represented by the pink triangle. The blue dot represents that case were no additional cuts besides EB ones are applied.

2.3.6 Remaining background estimation

The pseudo-ROC curves in Figure 2.16 show a linear behavior at high background strength. This behavior can be explained by considering that mis-id. pions in the TCS sample are removed at the same rate as mis-id. pions in the neutron sample when the cut is varied. True positrons might also be removed in the process and we can suppose this removal is small in the region where the background is important. If the classifier behave well, few signal events will be removed while most of the background will be cut away.

Let y be the variable describing the *normalized number of TCS events* and x describing the *normalized background strength* in Figure 2.16. In the region $0.1 < x < 1$ the pseudo-ROC curves in Figure 2.16 are linear. Let $B(x)$ and $S(x)$ respectively be the number of background (mis-id. pions) and signal events (true-positrons) in the TCS sample for the corresponding x background strength.

As stated before, the number of background events in the TCS sample is linear with the background strength x :

$$B(x) = \beta x, \quad (2.10)$$

where β is the number of background events in the TCS sample when no cut is applied. The function $S(x)$ is unknown but we assume it is increasing with x (when background is removed, signal events might also be removed by mistake), and does not vary much with x (signal events should not be removed by the classifier, the loss is estimated on simulations and is expected to be less than 1%, as shown in Figure 2.20).

The normalized number of TCS events can then be written in the linear region as:

$$y(x) = \frac{S(x) + B(x)}{S(1) + B(1)} = \frac{S(x) + x\beta}{S(1) + \beta}. \quad (2.11)$$

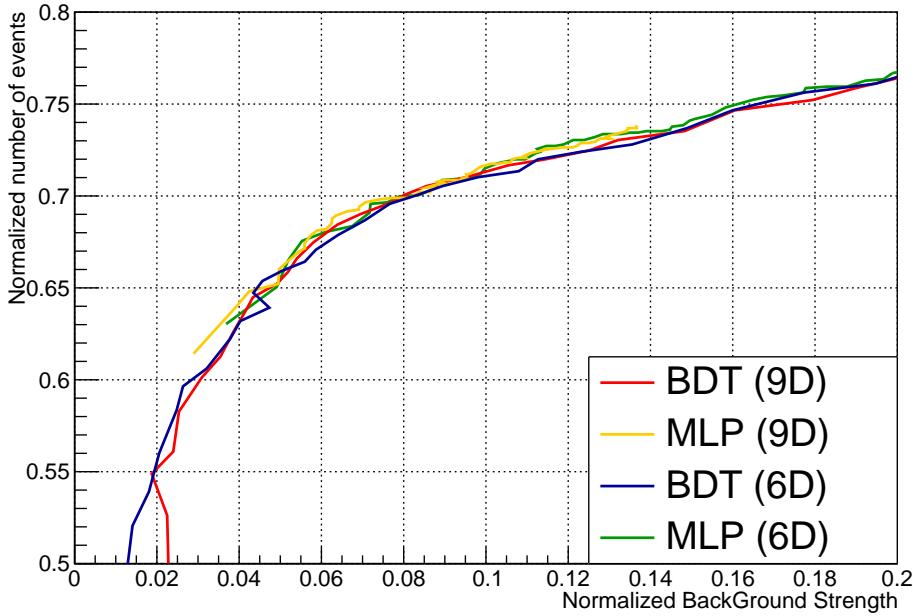


Figure 2.17: Pseudo-ROC curve obtained from data for 6D (SFs, shower widths) and 9D (SFs, shower widths and shower skewness) classifiers. The plot is zoomed in the region where the background is minimal while the signal is maximal. Adding the skewness information in the input of the classifiers does not improve their performances.

501 Although this formula is only applicable in the linear region, we can extrapolate it to $x = 0$:

$$y(0) = \frac{S(0)}{S(1) + \beta}. \quad (2.12)$$

502 At a given value x_0 of the normalized background strength achieved with the chosen cut, the corre-
503 sponding normalized number of TCS events is:

$$y(x_0) = \frac{S(x_0) + x_0\beta}{S(1) + \beta}. \quad (2.13)$$

504 Solving Equation (2.13) for $B(x_0)$ and assuming that the signal is almost constant $S(x_0) = S(0)$, one
505 can then obtain an estimate of the background/signal ratio at x_0 :

$$\frac{B}{S}(x_0) \approx \frac{y(x_0)}{y(0)} - 1. \quad (2.14)$$

506 The full derivation is given in Appendix A. In order to get a simple reading of the background/signal
507 ratio as a function of x , the pseudo-ROC curves of the 6D BDT and MLP are fitted with a linear
508 function in the range $0.1 < x < 1$. The pseudo-ROC curves $y(x)$ are then transformed as:

$$\frac{B}{S}(x) = \frac{y(x)}{y(0)} - 1, \quad (2.15)$$

509 where $y(0)$ is the intercept of the fit. The obtained curves are shown in Figure 2.18, from which one can
510 get an estimation of the background/signal ratio as a function of the normalized background strength
511 in the region where the function is linear.

512 The B/S ratio is estimated to be close to 0.05 for a cut at 0.5 on the MLP output corresponding
513 to a Background strength of 0.15 (see Figure 2.19 for values of the background as a function of the cut
514 applied on the output of the 6D MLP classifier). Finally the pion contamination C_{π^+} above 4 GeV is
515 estimated as:

$$C_{\pi^+} = \frac{B}{S+B} = \frac{1}{1+\frac{S}{B}} \approx 5\%. \quad (2.16)$$

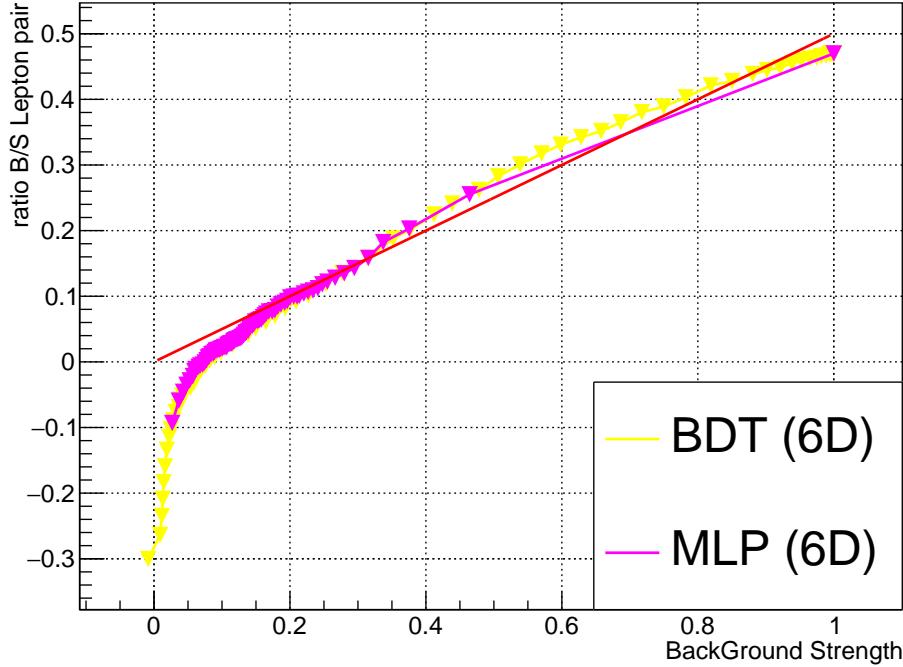


Figure 2.18: Background/Signal ratio in the TCS sample as a function of the background strength evaluated with the neutron sample. The red line corresponds to the linear fit of the BDT curve between 0.1 and 1.

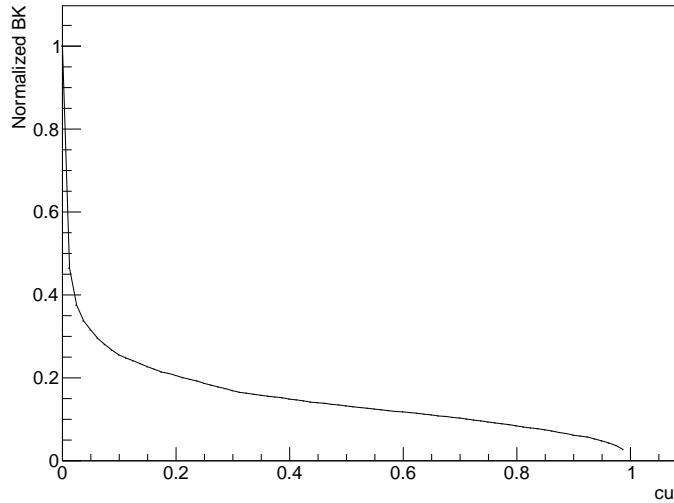


Figure 2.19: Normalized remaining background as a function of the cut applied to the output of the 6D MLP classifier.

516 2.3.7 Systematic checks on simulations

517 A second systematic check of the efficiency of the 6D MLP classifier was performed on simulations.
 518 A sample of BH-weighted events (see Chapter 3 for details on the TCS simulations) was used to test
 519 the classifier. This sample is completely uncorrelated from the training sample. It also assumes no
 520 hypothesis on momentum, polar angle and transverse momentum, apart from the ones arising from
 521 the event generation. The positron signal efficiency is shown in Figure 2.20. This figure shows that

the loss of true positrons in simulations when a cut is applied at 0.5 on the 6D MLP classifier output is less than 1%. Further tests on the kinematic distribution were also performed. Figures 2.21a, 2.21b and 2.22 show the distributions of momentum, polar angle and azimuthal angle of simulated positrons before and after the 0.5 cut was applied. No large systematic variation is seen, although some events do not fall in the kinematic limits of the training sample described in Subsection 2.3.1. This confirms that this classifier can be applied to events in the kinematics region of TCS events without losing signal events. One can see in Figure 2.22 that most true positrons that are identified as mis-id. pions are located on the outside edge of sector 3 (ϕ between 100° and 150°) and 5 (between -150° and -100°) of the FD of CLAS12. These two sectors have LTCC modules located between the DCs and the FTOF/EC carriage. The showers that initiate in LTCC module walls could be a reason for these mis-identifications.

Finally a similar check is performed on pion simulations. A sample of π^+ with flat kinematics is simulated and the kinematic distribution of mis-id. pions before and after the 6D MLP 0.5 cut are compared. Figure 2.23 shows the momentum and θ distributions of π^+ before and after the cut. No large systematic efficiency shift is seen.

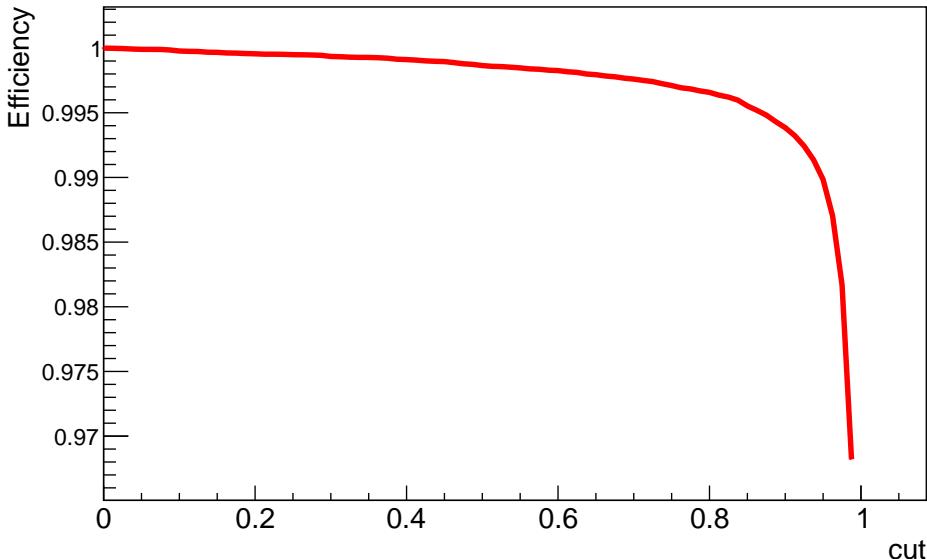


Figure 2.20: Signal efficiency for simulated TCS events (inbending electron) using the 6D MLP classifier as a function of the cut applied on the classifier output. At cut=0.5, the fraction of signal which is lost is less than a 1%.

2.3.8 Effect on data

The 6D MLP classifier was finally applied to the full available TCS data set. The output of the classifier and the signal efficiency on high-momentum positron candidates are shown in Figures 2.24 and 2.25, respectively. One can clearly see the mis-id. pions (around 0) and the true positrons (around 1) in the classifier output distribution. The effect of the classifier on the momentum distribution is assessed in Figure 2.26. One can clearly see an excess of positrons above the HTCC threshold in raw data. After applying the cut, the excess is removed.

A further check is performed by comparing the spectra obtained from simulation and data of the 6 variables included in the neural network. For the simulation samples the spectra are the ones shown in Figures 2.8 and 2.9. For data, we use the TCS data, apply a cut on the output of the MLP at 0.5 and plot the obtained spectra. Figures 2.27 and 2.28 show respectively the comparison for signal and background. One can see that both sets of spectrum are compatible providing further validation of the neural network method.

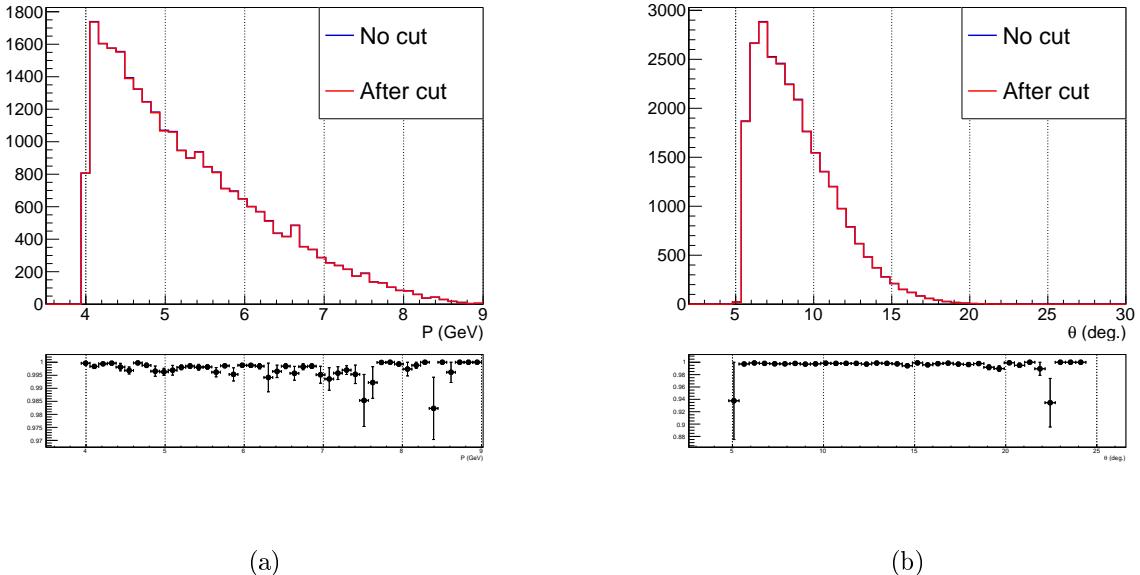


Figure 2.21: (a) Momentum spectrum of positrons for simulated TCS events. (b) θ spectrum of positrons for simulated TCS events. In both cases the histogram before and after MVA pid cut, of value 0.5, superimpose. Ratios of both histograms are also shown. The number of cut events per bin is always lower than 3%.

550 2.3.9 What about electron PID?

551 So far only the case of positron identification was treated. Figure 2.29 shows the distribution of the
 552 momentum of electrons versus the momentum of positrons for TCS events without any other cuts than
 553 the ones from EB and the ones described in Subsection 2.3.5. The π^+ contamination, mostly coming
 554 from photo-production events $ep \rightarrow e' p' \pi^+ (\pi^-)$, is visible in the positron high-momentum region. On
 555 the contrary no clear sign of π^- contamination in the electron high-momentum region is visible. This
 556 can be explained by the fact that high-momentum electrons $P_{e^-} > 4.9$ GeV are produced along with
 557 low momentum positrons $P_{e^+} < 4.9$ GeV. In this kinematics the positron is identified with a 99%
 558 efficiency by the HTCC. Lepton number conservation imposes that an electron is produced with the
 559 positron. Thus the simplest events that would contribute to the π^- contamination in the TCS data are
 560 the ones with final state $pe^+\pi^- (e^- e_{scat}^- \pi^+)$, where the electrons and π^+ would be undetected. Such
 561 events are suppressed by, at least, a factor of $\alpha_S^4 < 1\%$.

562 In addition to these considerations, the 6D MLP positron classifier was tested on electrons with
 563 momentum higher than 4.9 GeV from the TCS sample. This classifier is assumed to provide good
 564 results on electrons as the shower mechanism does not depend on the charge of the initiating particle
 565 at high energies. The signal efficiency as a function of the cut applied to the classifier output for
 566 simulation and data electrons is shown in Figure 2.30. One can see that simulated electrons are
 567 suppressed at most up to 3% for a 0.8 cut. The same behavior is seen for data electrons, showing that
 568 the background is less prominent than in the positron case. No further pid cuts for electrons than the
 569 EB ones are added in the analysis.

570 Finally the double pion contamination (where both leptons are mistaken for pions) is absent in the
 571 high-momentum region. The kinematic region were both leptons have momenta higher than 4.9 GeV
 572 is not accessible in CLAS12 kinematics.

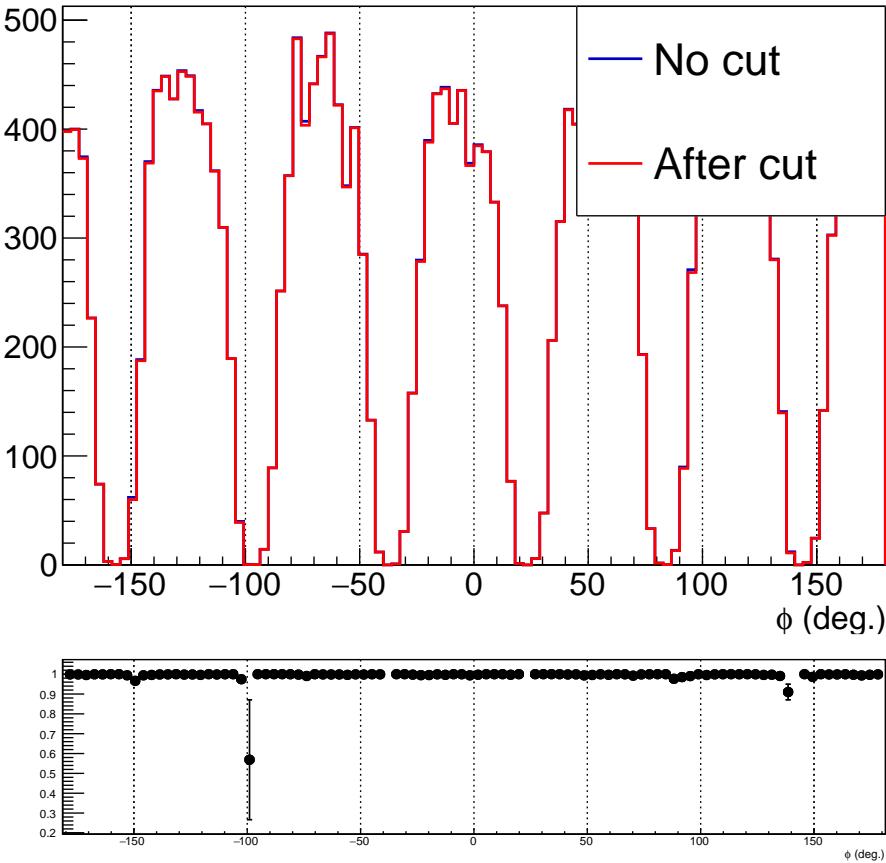


Figure 2.22: ϕ spectrum of positrons for simulated TCS events. The loss of positrons happens at very specific ϕ angles corresponding to the edges of sectors which accommodate an LTCC module.

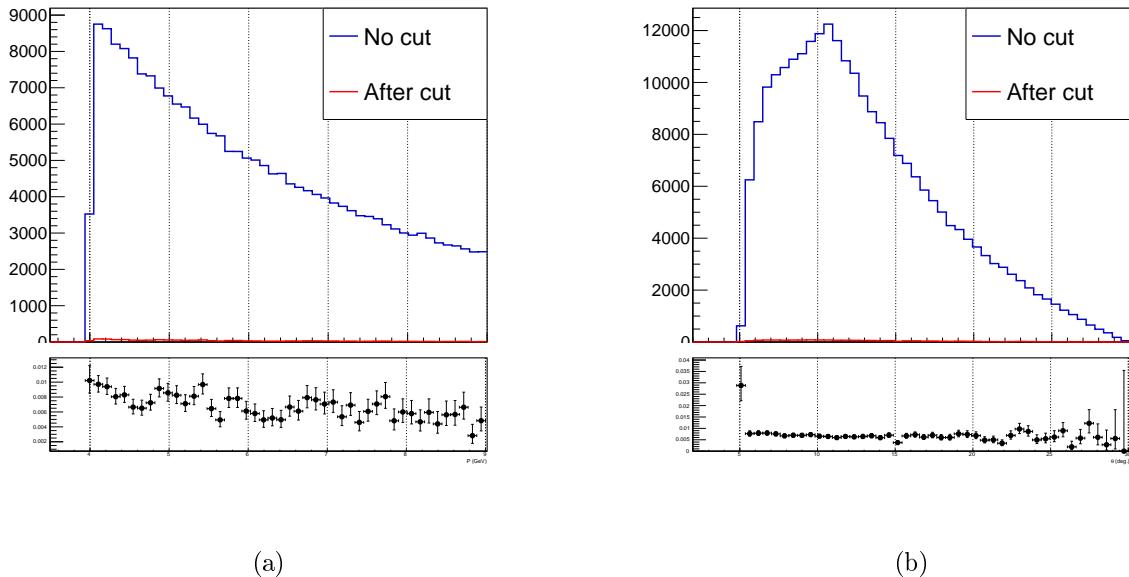


Figure 2.23: (a) Momentum spectrum of mis-id. pions. (b) θ spectrum of mis-id. pions. The histograms before and after the MLP cut at 0.5 are shown.

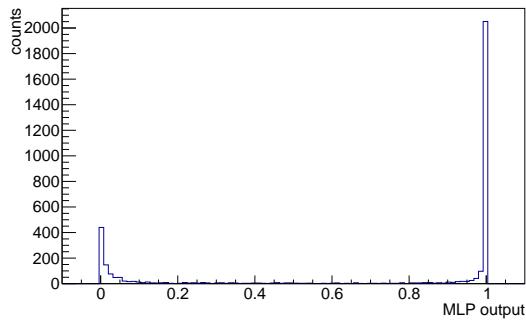


Figure 2.24: Output of the 6D MLP neural network on the TCS events used in the analysis.

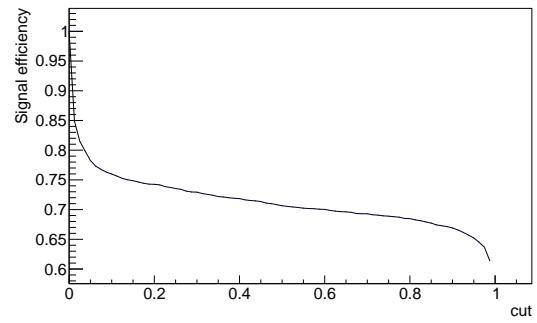


Figure 2.25: Signal efficiency of the cut applied on the MLP output. This corresponds to the proportion of remaining events for a given cut.

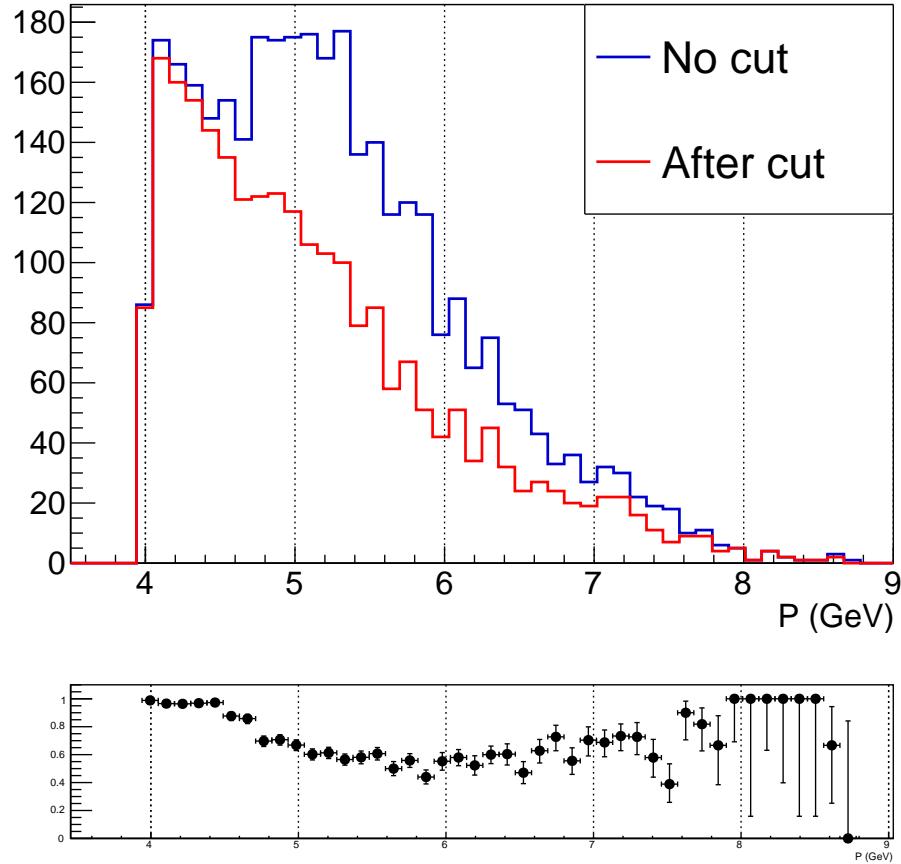


Figure 2.26: Momentum of positrons for TCS events. The blue histogram corresponds to all events. The red histogram corresponds to events that passed a 6D MLP output cut of 0.5.

573 **2.3.10 Crosscheck using electrons in the outbending torus magnetic field config-**
 574 **uration**

575 A further cross check of the positron PID was performed using electrons detected with CLAS12 in the
 576 outbending torus magnetic field configuration. As, at typical Jlab energies, electrons and positrons
 577 have very similar behaviours in the calorimeters one can cross check the identification of outbending

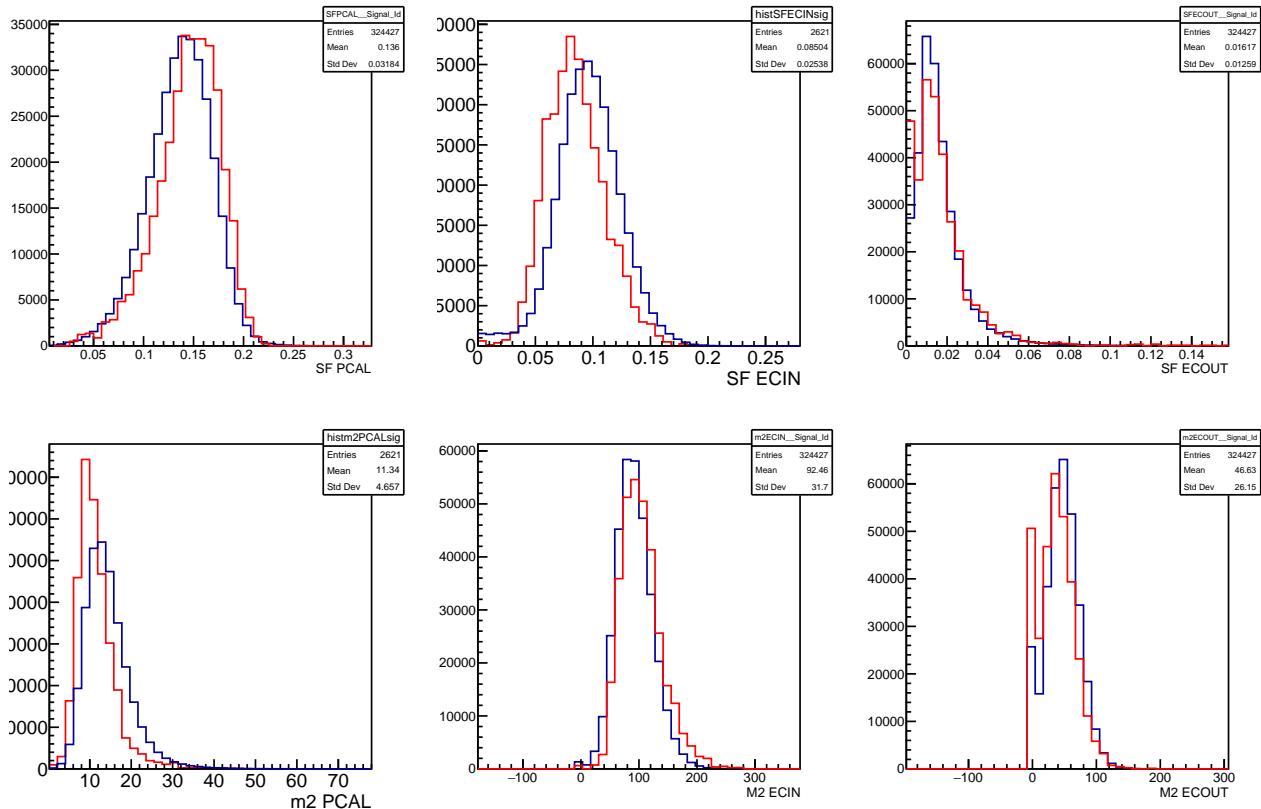


Figure 2.27: Comparison of the data and simulation spectra of the 6 variables used in the neural network for signal events. The red curves corresponds to data, the blue curves to simulation. The curves are normalized to have equal integrals.

positrons using inbending electrons. TCS events are used for this studies (one proton, one electron, one positron and exclusivity cuts). The classifier is applied on the electron response in each layer of the calorimeter. Figure 2.31 shows the signal efficiency for outbending electrons for the corresponding cut values on the output of the neural network. For a cut at 0.5, the signal efficiency is 98.5%. Figures 2.32 shows the kinematic dependance of the remaining events after cuts. No strong kinematic dependance is seen. Finally combining the curves in Figures 2.31 and 2.19 one can plot the ROC curve of the neural networks estimated with data only. This ROC curves is shown in Figure 2.33. Additionally, the yellow square on the plot shows the signal efficiency and background strength at the standard value of the cut used in the analysis. Furthermore, the green curves show the variation of these two quantities when the systematic check (0.5 ± 0.3) is performed, outside of this range the signal efficiency drops dramatically or the net background rejection (from missing neutron events) goes below 80%. Finally, the ROC curves obtained from simulation (see Figure 2.13b) is superimposed. One can see that the performance of the classifier on data and simulations are very similar with little discrepancy at high background rejection. However we only utilize the region which is well reproduced by simulations.

2.3.11 Choice of the cut on the output of the neural network

In this subsection the choice of the value of the cut applied to the output of the neural network is developped. For each set of input variable, the neural network provides an output between 0 and 1. If the output is close to 0, the particle is probably background (misidentified pion), while if it is 1 it is signal (a true positron). One should then choose what value delimits the background and signal regions. A naive choice is to select 0.5 as previously mentionned in the previous sections. This cut yields a 98.5% signal efficiency and a 87% background rejection.

2.3. Positron identification

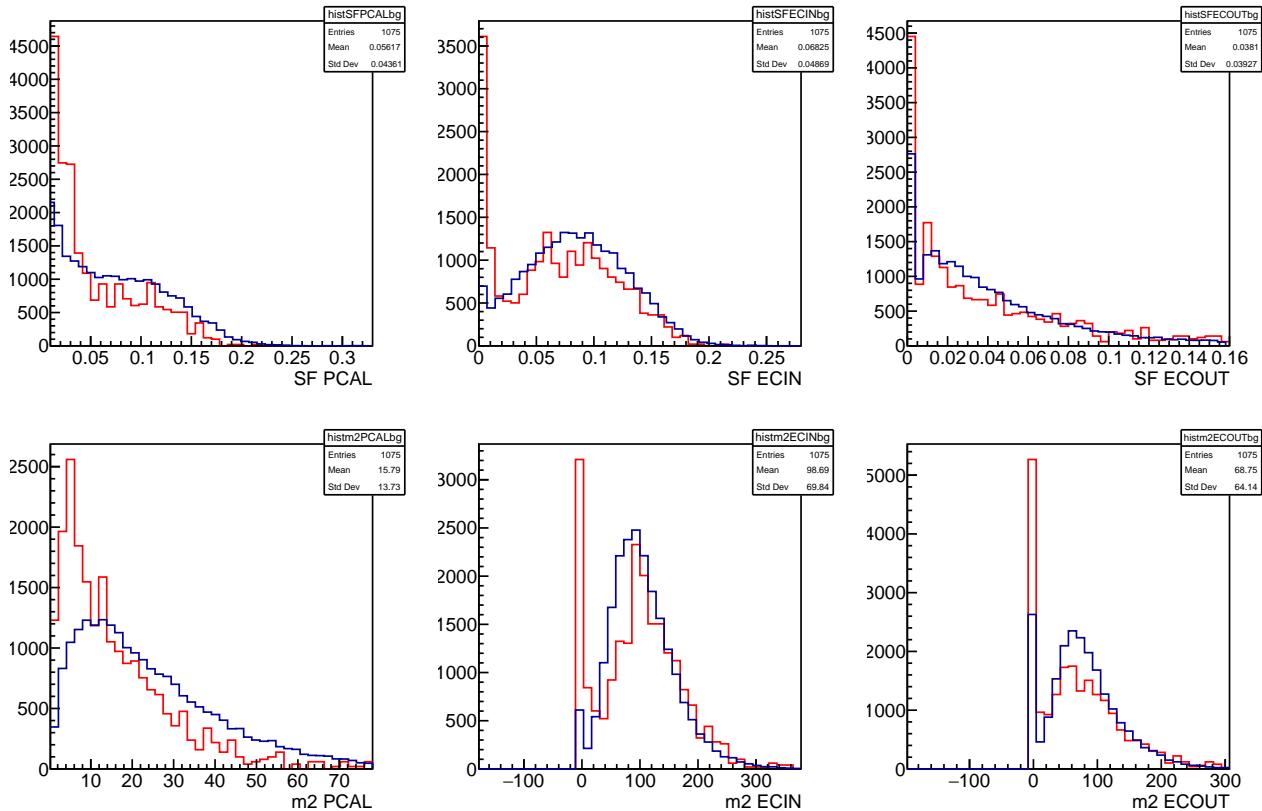


Figure 2.28: Comparison of the data and simulation spectra of the 6 variables used in the neural network for background events. The red curves corresponds to data, the blue curves to simulation. The curves are normalized to have equal integrals.

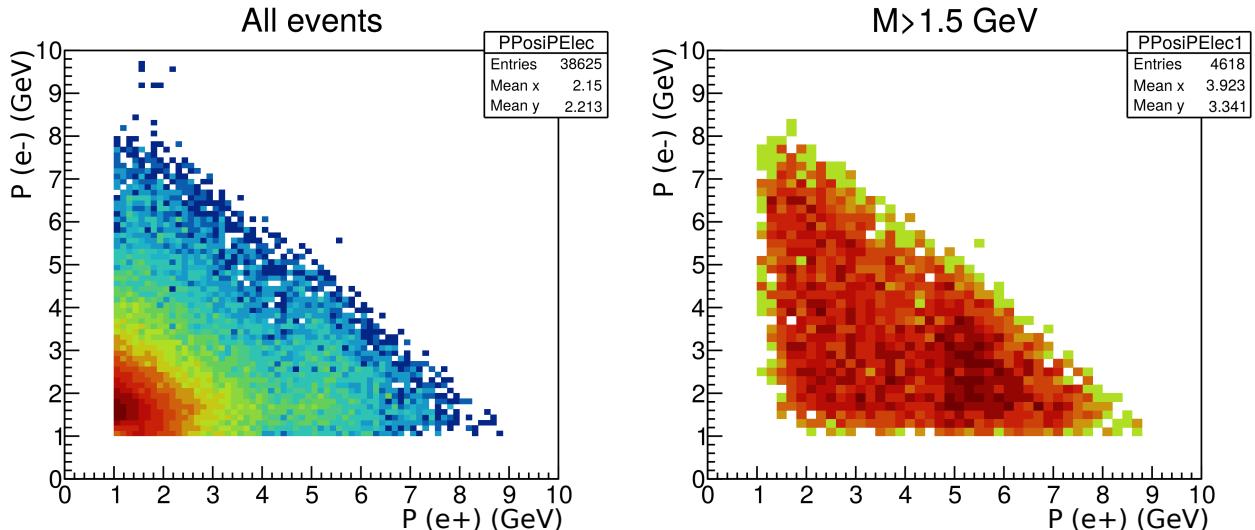


Figure 2.29: Momentum of the positron versus momentum of the electron in e^-e^+p events for all invariant masses (left) and only invariant masses higher than 1.5 GeV (right). No additional PID cuts than the EB ones are used in these two plots. The π^+ contamination is clearly visible in the $P(e^+) > 4.9$ GeV region in the left figure. On the contrary no π^- contamination is visible in the $P(e^-) > 4.9$ GeV.

599 A way to quantitatively choose the value of the cut is to calculate the significance define as:

$$S = \frac{S(\epsilon_S)}{\sqrt{S(\epsilon_S) + B(\epsilon_B)}} = \sqrt{S_i} \frac{\epsilon_S}{\sqrt{\epsilon_S + k\epsilon_B}}, \quad (2.17)$$

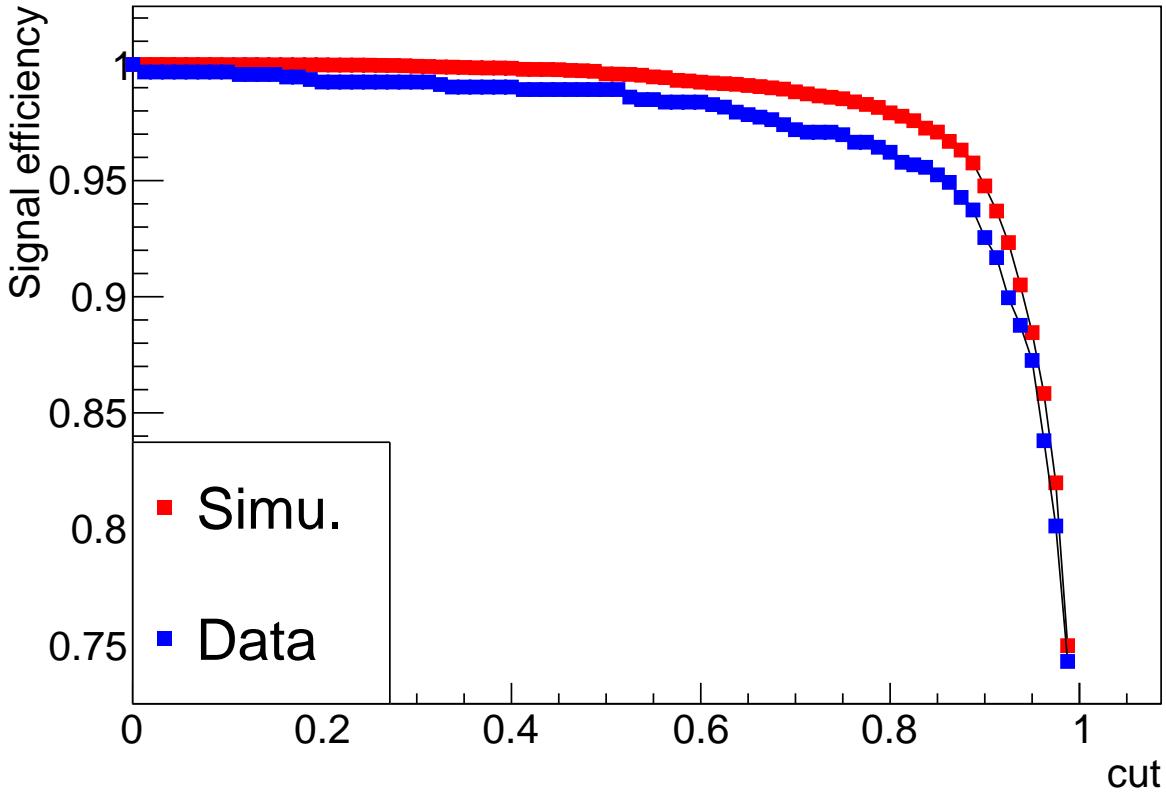


Figure 2.30: Signal efficiency of the MLP classifier for electrons with momentum bigger than 4 GeV, for simulated TCS events (red) and data (blue).

600 where

$$B_i = kS_i, \quad (2.18)$$

601 $S(\epsilon_S)$ is the number of signal events for a given cut, $B(\epsilon_B)$ the number of background events for a given
 602 cut, S_i the number of Signal events without cut, B_i is the number of background events without cut,
 603 ϵ_S the signal efficiency and $\epsilon_B = 1 - r_B$ the background efficiency (r_B being the background rejection
 604 shown on ROC curves). From Figure 2.16 we estimated that $k = 50\%$. In the following we set ξ_i to 1
 605 as the overall normalization of the significance is not of interest here. The goal is to find the cut that
 606 maximize the significance.

607 The significance curve is shown in Figure 2.34. The maximum of significance is 0.967 obtained for
 608 a cut at 0.825. This optimal value of the cut is different from the one we used in the analysis. However
 609 one can see on Figure 2.34 that the significance is mostly flat in the cut region between 0.2 and 0.8. We
 610 can compare the values of significance for the cuts used in this analysis (0.5 for the standard cut, 0.2
 611 and 0.8 for the systematics checks). The minimum of significance, for values of the cut preserving some
 612 signal, is obtained when the cut is 0. In this case the significance is 0.816. The range of significance
 613 for values of cuts between 0 and 0.825 is thus $0.967 - 0.816 = 0.151$. Table 2.1 shows, for different
 614 values of the cut, the associated significance and the variation to the maximum relative to the maximal
 615 variation of the significance. For the chosen cut at 0.5 the lose of significance is 4%. The range of cut
 616 covered bu the systematic yields significance which are at most 17% lower than the maximum. Outside
 617 of this range the significance fall quickly.

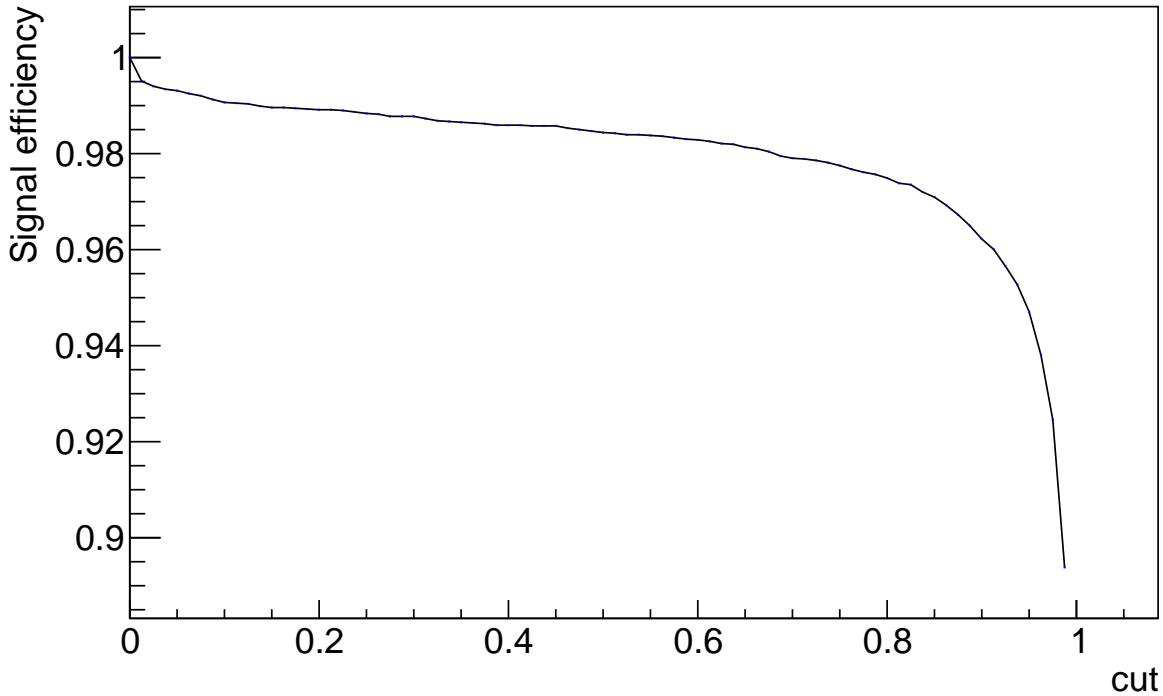


Figure 2.31: Signal efficiency of the MLP classifier for outbending electrons in TCS events with momentum bigger than 4 GeV.

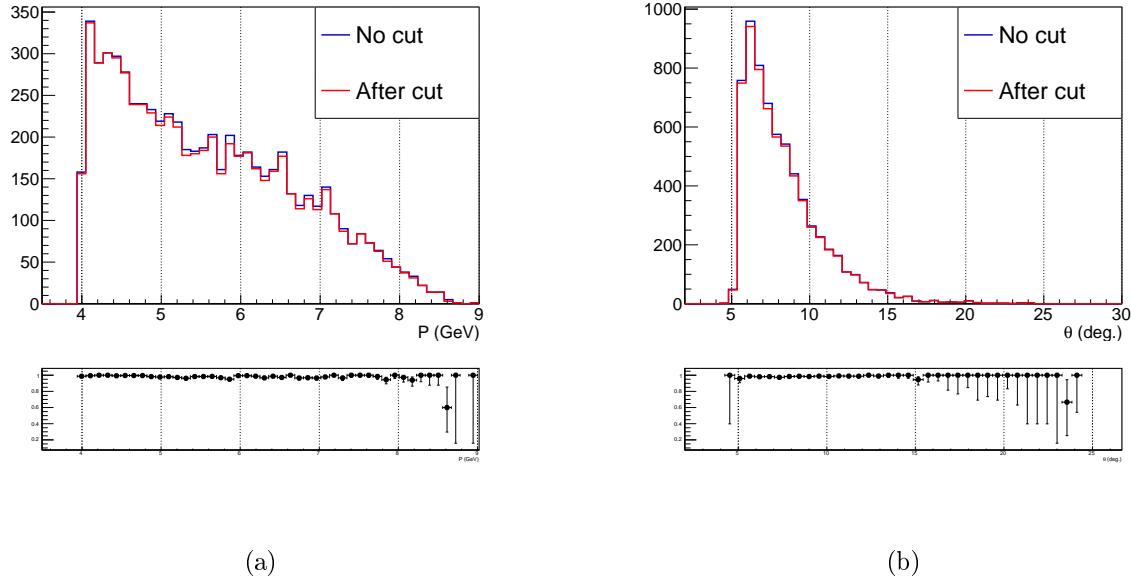


Figure 2.32: (a) Momentum spectrum of outbending electrons in TCS events. (b) θ spectrum of outbending electrons in TCS events.. The histograms before and after the MLP cut at 0.5 are shown.

Cut c	Significance $S(c)$	Variation from the maximum $(S(c) - S_{Max})/0.151$
0.825	$0.967(S_{Max})$	0%
0.5	0.96	-4%
0.2	0.947	-17%
0.8	0.967	0%

Table 2.1: Value of the significance for different values of the cut.

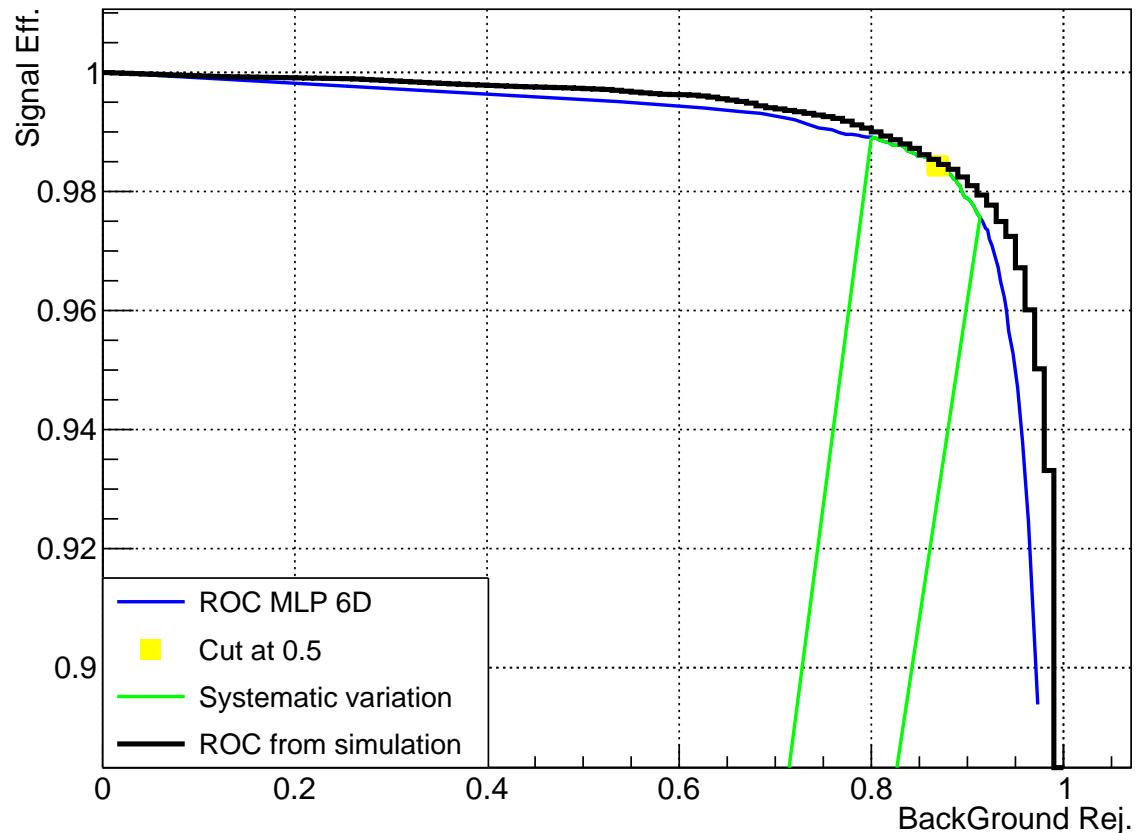


Figure 2.33: ROC Curves of the neural network classifier calculated with data only. The signal efficiency is calculated using outbending electrons, while the background rejection comes from missing neutron events.

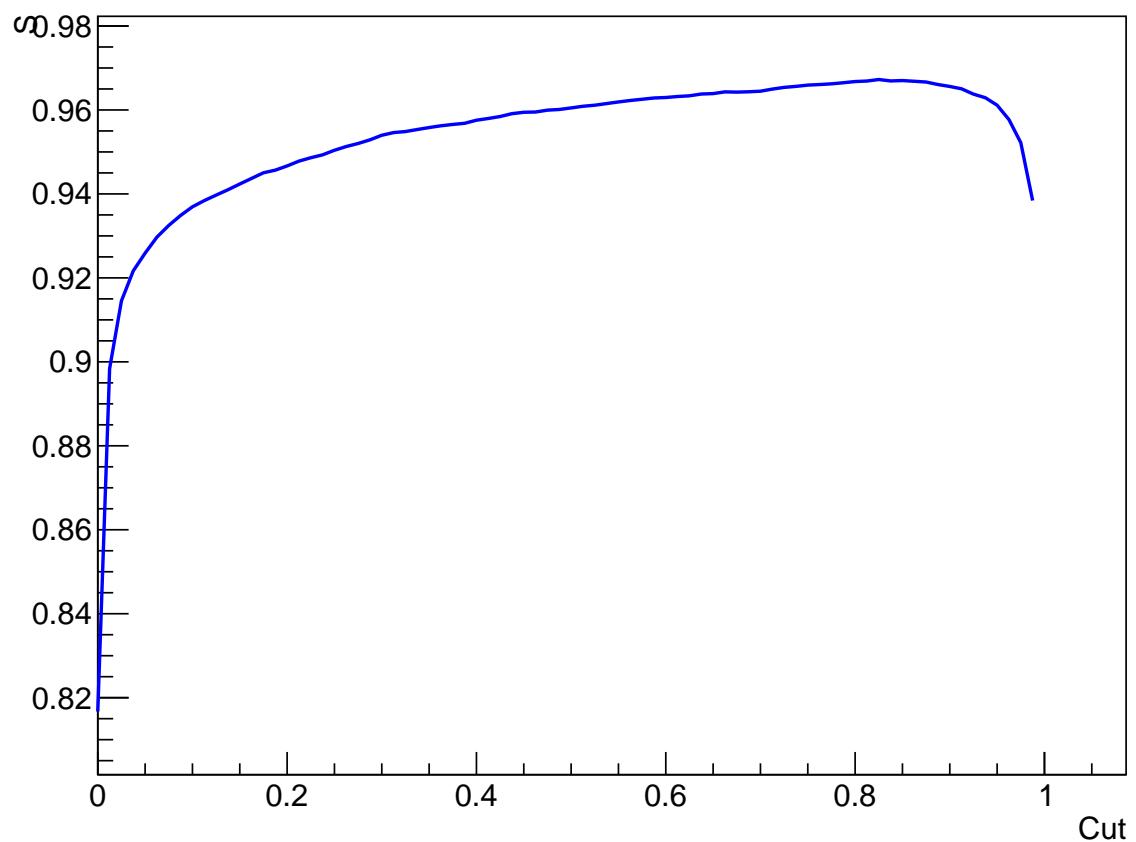


Figure 2.34: Significance curve obtained using data samples. The y-axis values are not relevant. The maximum significance is obtained for a cut at 0.825.

618 2.4 Proton momentum corrections

619 The proton momentum corrections are split in two main contributions. The first corrections, pre-
 620 sented in Subsection 2.4.1, are determined by comparing Monte-Carlo generated and reconstructed
 621 proton kinematics. The shifts observed in this case is attributed to the energy lost by the proton while
 622 propagating in the various detector materials. The determination of the parameters of this correction
 623 solely relies on simulations. The second contribution is a data-only based correction, which aims at
 624 correcting mis-alignments and inefficiencies of the actual detectors, not accounted for by the simula-
 625 tion. In particular, data-driven corrections are developed for the momenta of the protons in the CD
 626 in Subsection 2.4.2.

627 2.4.1 Monte-Carlo corrections

628 The Monte-Carlo (MC) momentum corrections for the proton are derived using BH simulations (see
 629 Section 3.1 for details). The goal of these corrections is to match the momenta of the generated protons
 630 with the momenta of the reconstructed protons. These corrections account for the energy lost by the
 631 proton while crossing the various detector parts of CLAS12.

632 The difference between the generated and reconstructed momenta,

$$\Delta P = P_{Gen.} - P_{Rec.}, \quad (2.19)$$

633 is studied as a function of the polar angle of the proton, θ . The plot in Figure 2.35 shows the difference
 634 between the generated and reconstructed proton momenta as a function of θ for protons detected in the
 635 FD. The momenta difference shows different behaviors in two distinct regions. Below 27° protons cross
 636 little material before being detected and the momentum difference is small (below 20 MeV). Above 27°
 637 the material budget between the target and the DCs is larger, especially due to the HTCC and the
 638 forward CTOF light-guides. In this region the momentum resolution is degraded and the momentum
 639 difference can reach up to 80 MeV.

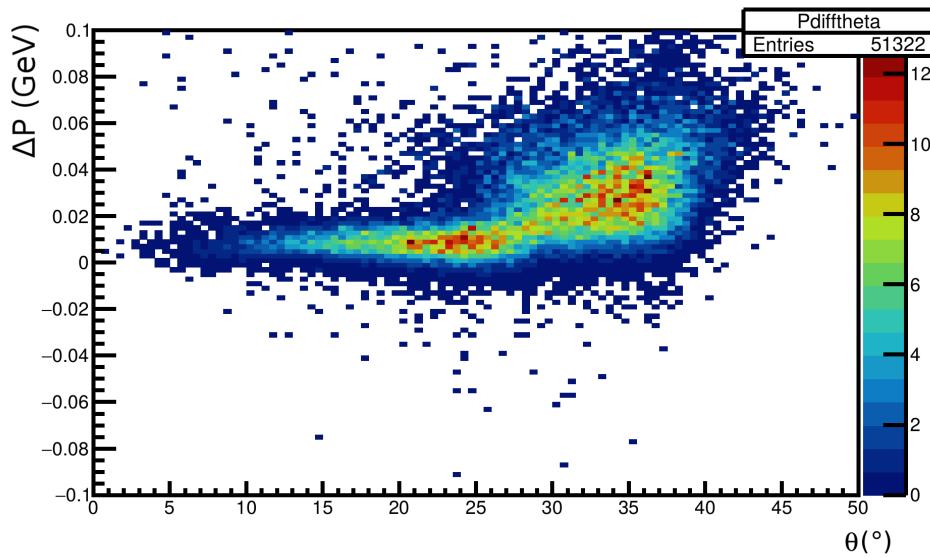


Figure 2.35: Difference between the generated and reconstructed momenta for protons detected in the FD of CLAS12. One can see two distinct regions, below and above $\theta = 27^\circ$.

640 The MC corrections are derived in three different CLAS12 regions: the two regions in the FD
 641 described above and one region in the CD. In each region the momenta difference is parametrized as
 642 a function of the reconstructed momentum as shown in Figure 2.36. These corrections are at most of
 643 the order of 4% for low-momenta proton (~ 0.45 GeV) in the high-polar angle region of the FD. The
 644 corrections are applied to protons in both the simulations and the data.

2.4. Proton momentum corrections

645 The corrections functions are second order polynomial in the FD and first order in the CD, defined
 646 as:

$$f_{MC}(P) = a_0 + a_1 \cdot P + a_2 \cdot P^2, \quad (2.20)$$

647 where the coefficient a_i are given in Table 2.2. The corrected momentum then reads:

$$P_{Corr. MC} = P_{Uncorr.} + f_{MC}(P). \quad (2.21)$$

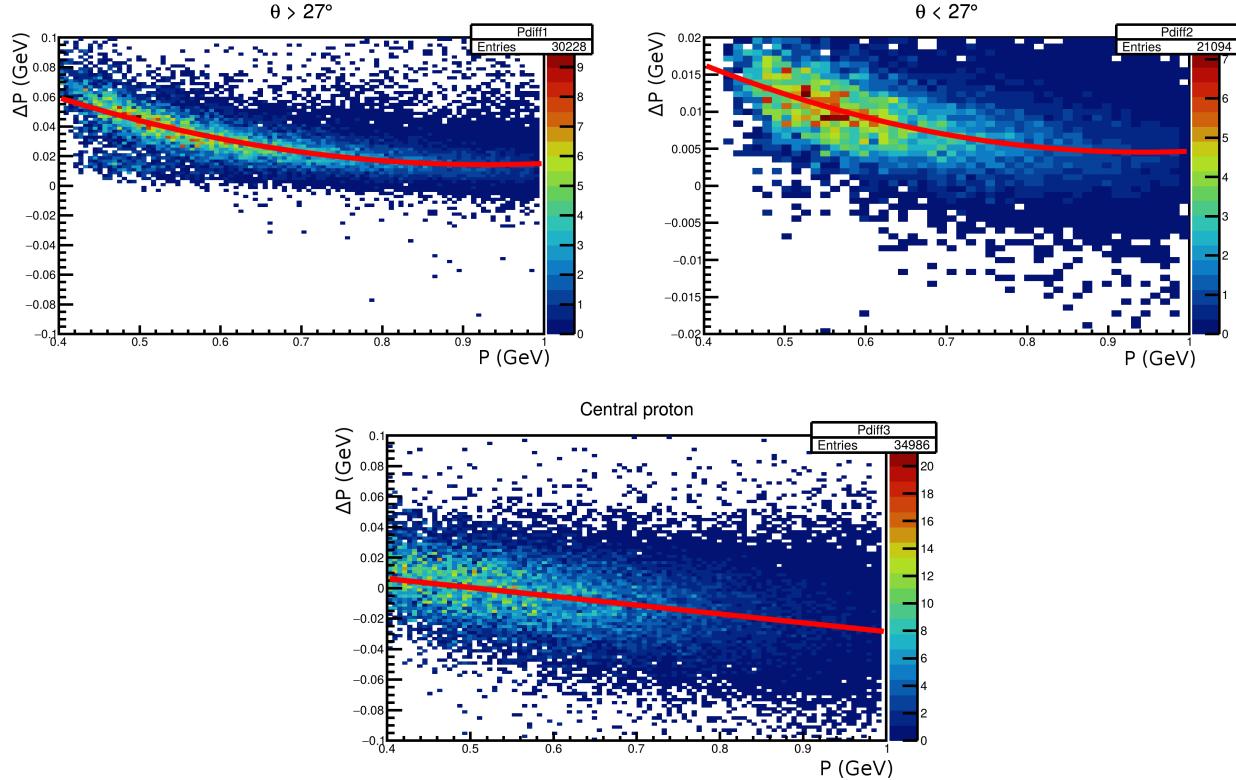


Figure 2.36: Top left: momenta difference (defined in Equation (2.19)) as a function of the reconstructed momentum for polar angles above 27° in the FD. The distribution of the mean of each momentum slice (obtained by a gaussian fit) is fitted with a 2nd order polynomial. The result of this fit is superimposed to the original distribution (red line). Top right: Corresponding figure for protons with polar angles below 27° in the FD. Bottom: Corresponding figure for protons in the CD.

Region	a0	a1	a2
FD ($\theta > 27^\circ$)	0.1533 ± 0.0015	-0.2989 ± 0.0042	0.1607 ± 0.0028
FD ($\theta < 27^\circ$)	0.03989 ± 0.0009	-0.0748 ± 0.0027	0.03957 ± 0.0018
CD	0.02929 ± 0.0006	-0.05779 ± 0.0009	-

Table 2.2: Values of the coefficients used in the MC-based momentum for proton.

2.4.2 Data-driven momentum corrections

649 Data-driven momentum corrections for the proton are motivated by the fact that the simulations
 650 depict an "ideal" detector, and therefore do not perfectly reproduce the data. Because of issues mainly
 651 due to the reconstruction software of the CVT, the reconstructed momentum in the CD can be shifted
 652 from its actual value. To investigate this issue a method using exclusive two-pion production events
 653 was developed. This method relies on the exclusive measurement of the $ep \rightarrow e'p'\pi^+\pi^-$ reaction, where
 654 the scattered electron and the pions are detected in the FD. The kinematics of the scattered proton can
 655 then be studied in two different ways. In one case the proton can be detected by CLAS12, in the other

case its kinematics can be inferred by calculating the missing 4-momentum of X in the $ep \rightarrow e'\pi^+\pi^-X$ reaction.

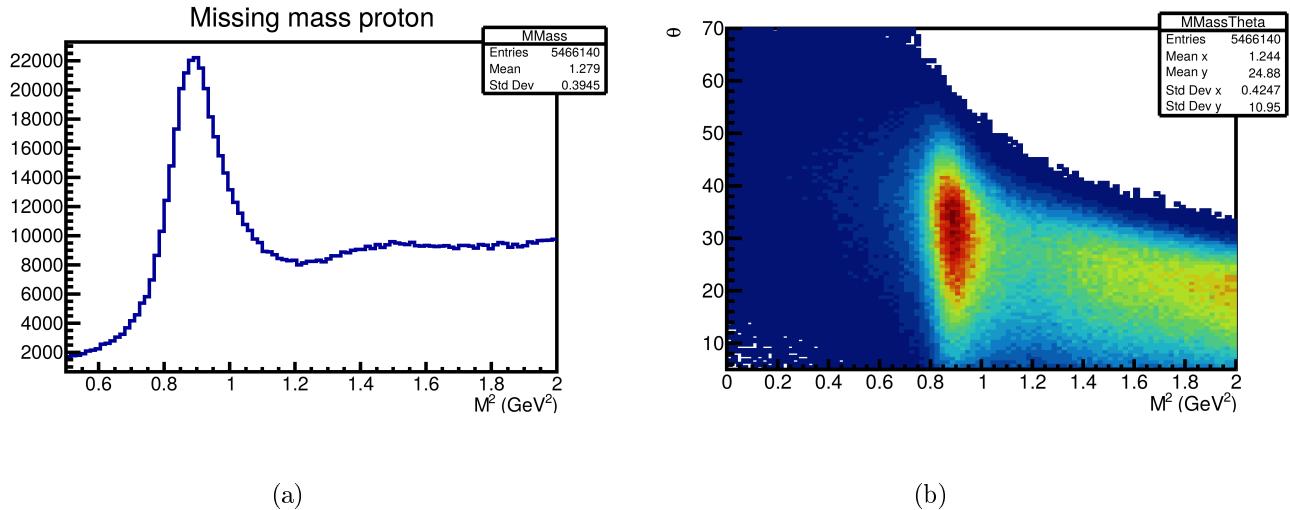


Figure 2.37: (a) Squared missing mass of the $ep \rightarrow e'\pi^+\pi^-X$ reaction. One can see a clear peak at the proton mass and a higher-mass continuum. (b) Missing polar angle as a function of the squared missing mass for the same reaction. Once can see that the high-polar-angle region, corresponding to topologies where the missing proton goes in the CD, is free of high-mass background.

The missing mass spectrum obtained from the latter analysis is shown in Figure 2.37. One can see that the missing mass shows a clear peak at the proton mass. Furthermore, looking at the dependence of the missing mass as a function of the missing polar angle in Figure 2.37b, one can see that at high polar angles (above 35°) the high-mass component is suppressed and the missing mass spectrum has a contribution only from the scattered proton. This allows to compare directly the kinematics of the missing proton to the kinematics of the detected proton.

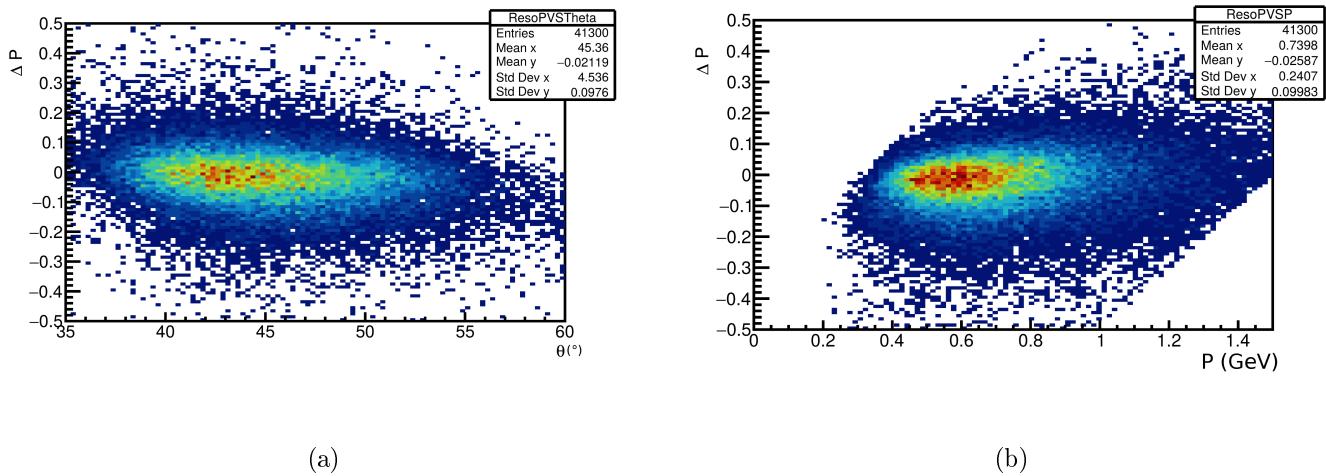


Figure 2.38: (a) Data-driven momentum difference as a function of the polar angle for protons in the CD. (b) Momentum difference as a function of the momentum.

Figure 2.38 shows the momentum difference:

$$\Delta P = P_{\text{Rec.}} - P_{\text{Missing.}} \quad (2.22)$$

as a function of the detected polar angles (Figure 2.38a) and the detected momenta (Figure 2.38b).

666 No large dependencies are seen. The momentum resolution defined as:

$$\frac{\Delta P}{P} = \frac{P_{Rec.} - P_{Missing.}}{P_{Rec.}} \quad (2.23)$$

667 is also plotted as a function of the local azimuthal angle in the last layer of the CVT, ϕ_{CVT} , as shown
668 in Figure 2.39.

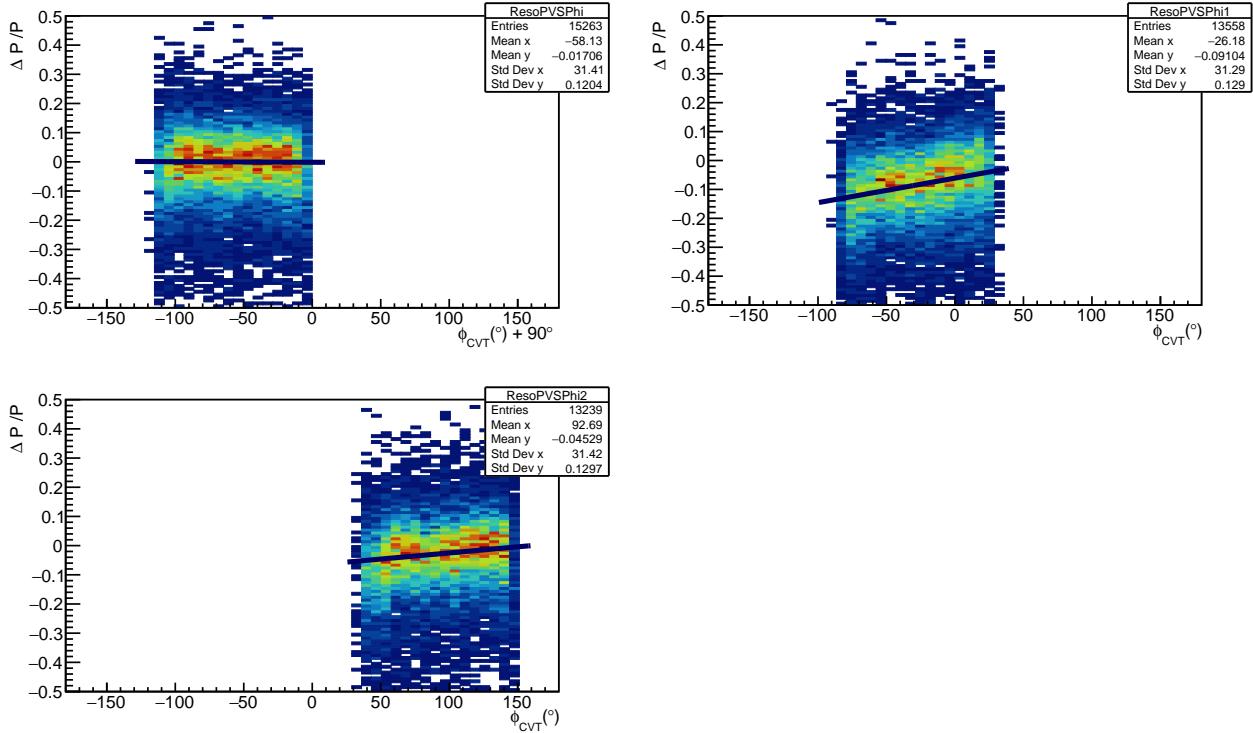


Figure 2.39: Momentum resolution as a function of the local azimuthal angle for protons in the CD, for the three regions of the CVT. The superimposed black line is the correction function for each region.

669 Each subplot corresponds to one of the three CVT regions. The azimuthal coordinate of the last
670 layer of the CVT is used for this correction (layer ID 12 in the CLAS12 reconstruction nomenclature).
671 The distribution of the gaussian means of each ϕ_{CVT} slices is fitted with a linear function, for each
672 subplot. The obtained coefficients are detailed in Table 2.3, where the functions can be written:

$$f_{DATA}(\phi_{CVT}) = c_0 + c_1 \cdot \phi_{CVT}. \quad (2.24)$$

673 It has to be noted that for the range $-210^\circ < \phi_{CVT} < -90^\circ$ a 90° shift is used for the angle to be in the
674 $[-180^\circ, 180^\circ]$ range. The coefficient given in Table 2.3 are given after this translation. The resulting
675 corrections are applied to the proton momenta in the data. The corrected momentum reads:

$$P_{Corr. DATA} = P_{Uncorr.} \cdot (1 - f_{DATA}(\phi_{CVT})). \quad (2.25)$$

676 The corrections range from almost zero for protons in the region $-210^\circ < \phi_{CVT} < -90^\circ$ to up to
677 14% at the lower edge of the $-90^\circ < \phi_{CVT} < 30^\circ$ region. These corrections are performed after the
678 MC corrections presented in the previous subsection.

ϕ_{CVT} range	c0	c1
$-120^\circ < \phi_{CVT} + 90^\circ < 0^\circ$	-0.0014 ± 0.0031	$-2.137e^{-5} \pm 5.240e^{-5}$
$-90^\circ < \phi_{CVT} < 30^\circ$	-0.060 ± 0.0013	$0.000849 \pm 3.42e^{-5}$
$30^\circ < \phi_{CVT} < 150^\circ$	-0.0670 ± 0.0044	$0.00041 \pm 4.61e^{-5}$

Table 2.3: Values of the coefficients used in the data based momentum for proton in the CD.

679 *End of common analysis section*

681 **2.5 Lepton momentum corrections**

682 In this section we present the corrections applied to leptons. First, in Subsection 2.5.1, the shift be-
 683 tween the Monte-Carlo generated and the reconstructed kinematics is analyzed. Second, in Subsection
 684 2.5.2, a specific detected-photon correction is presented.

685 **2.5.1 Monte-Carlo corrections**

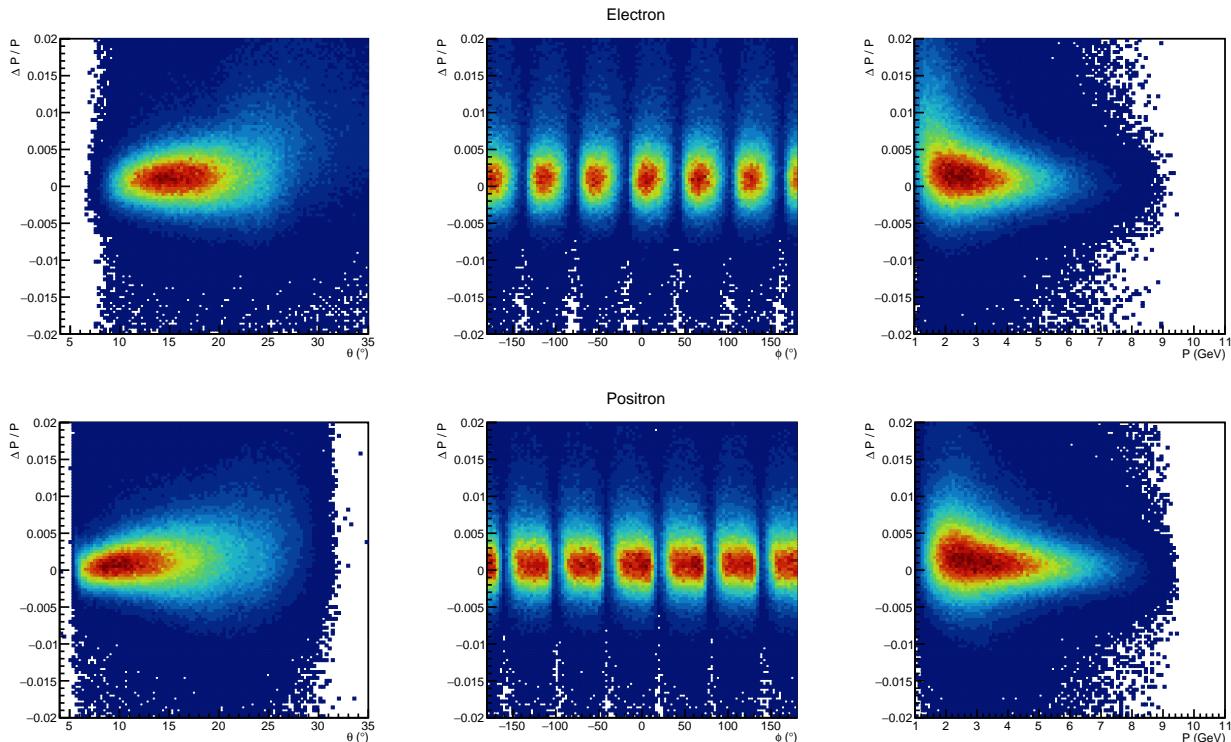


Figure 2.40: Reconstructed momentum resolution for electrons (top) and positrons (bottom), as a function of the lab angles θ , ϕ and the momentum.

686 As for proton, BH-simulation events are used to investigate the impact of the energy lost by the lepton
 687 in the material of CLAS12 before being detected. This energy loss ultimately affects the reconstruction
 688 of the kinematic variables, especially the momentum. Figure 2.40 shows, for the electrons and the
 689 positrons, the momentum resolution:

$$\frac{\Delta P}{P} = \frac{P_{Gen.} - P_{Rec.}}{P_{Rec.}}, \quad (2.26)$$

690 where $P_{Gen.}$ is the generated momentum and $P_{Rec.}$ the reconstructed one, as a function of their polar
 691 and azimuthal angles in the lab reference frame, θ and ϕ , and their momentum. These plots show that
 692 the shift in momentum due to the energy loss of leptons is always smaller than 1%. For this reason,
 693 we decided not to add MC-driven corrections to the momentum of leptons in the subsequent analysis.

694 **2.5.2 Detected radiated photon correction**

695 Leptons propagating through the target material can lose energy by radiating photons. These
 696 photons are emitted at low angles around the direction of the lepton. The process is represented by
 697 the diagram of Figure 2.41.

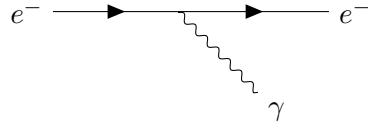


Figure 2.41: Diagram representing the radiation of a photon from an electron.

698 One way to recover the initial momentum of the lepton, before any radiative energy loss, is to use
 699 the detected photon in the CLAS12 EC. Figure 2.42a shows, in the CLAS12 data, the uncorrected
 700 momenta of electrons versus the difference in the polar angle at the vertex between electrons and
 701 detected photons, $\Delta\theta_\gamma$, while Figure 2.42b shows the cone angle between electrons and detected photons
 702 (the angle between the 3-momentum of the electron and the 3-momentum of each photon). One can
 703 see that in both plots there is a significant number of photons detected in the close vicinity of the
 704 electron. The same behaviour is seen for positrons. Similar results are also obtained for simulations,
 705 as seen in Figure 2.43.

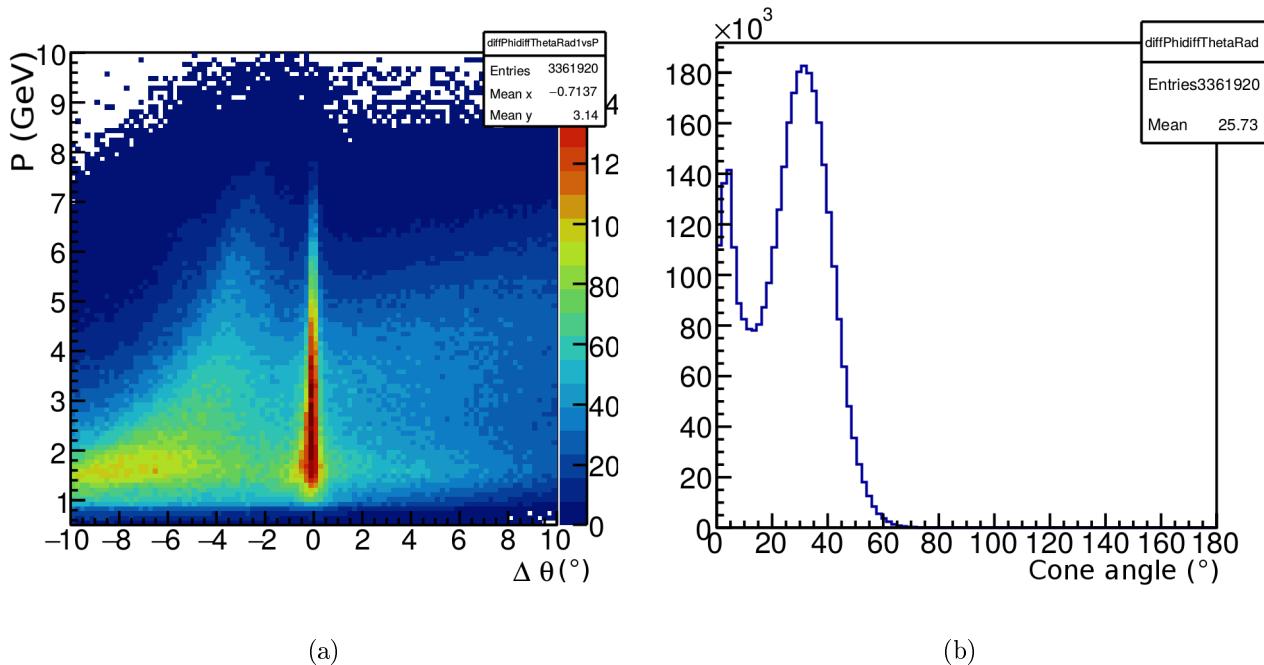


Figure 2.42: (a) Electron momentum as a function of the difference of polar angle between the electron and the photons detected in each event, (b) Cone angle between the electron and same event photons. These plots are produced with real CLAS12 data.

706 The 4-momenta of photons within $-1.5^\circ < \Delta\theta_\gamma < 1.5^\circ$ and with a cone angle below 10° are added
 707 to the reconstructed 4-momentum of the corresponding lepton. The effect of this correction can be
 708 seen on simulation in Figures 2.44 and 2.45. The simulations used for this study are described in
 709 Section 3.1. The improvements in the momentum and angular resolutions are shown for the electrons
 710 for which this correction is applied.

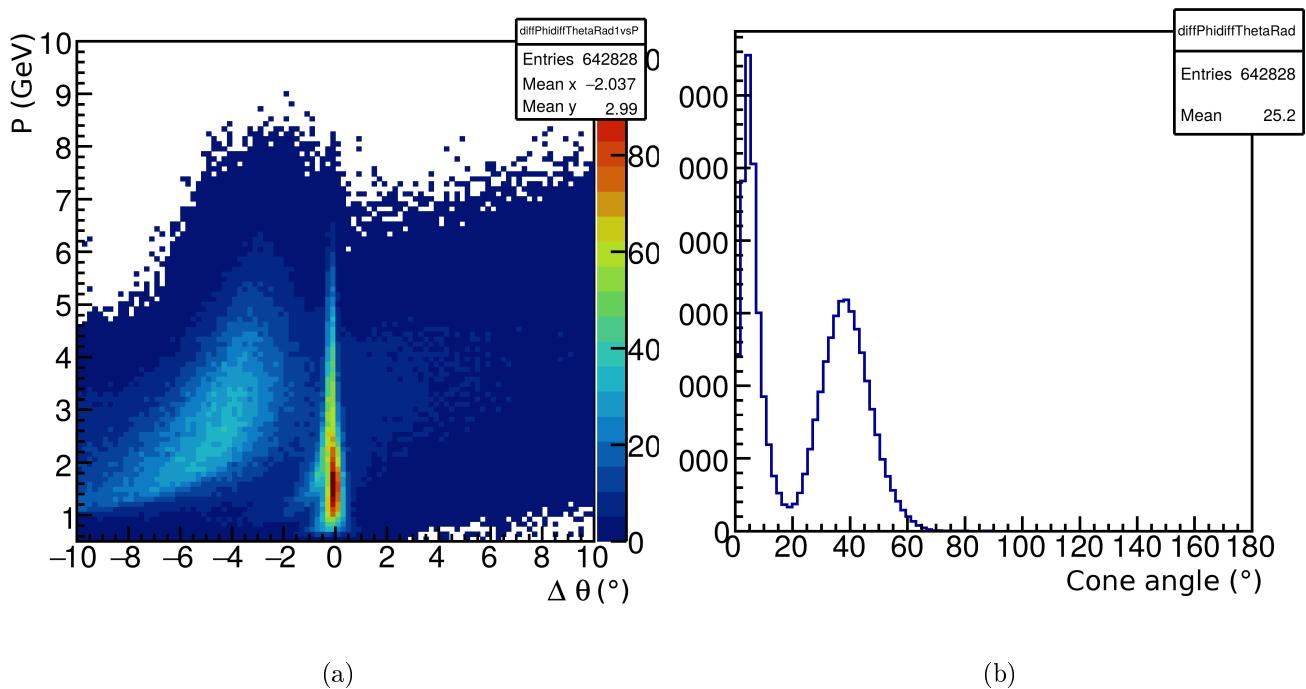


Figure 2.43: (a) Electron momentum as a function of the difference of polar angle between the electron and the photons detected in each event, (b) Cone angle between the electron and same event photons. These plots show simulated electrons.

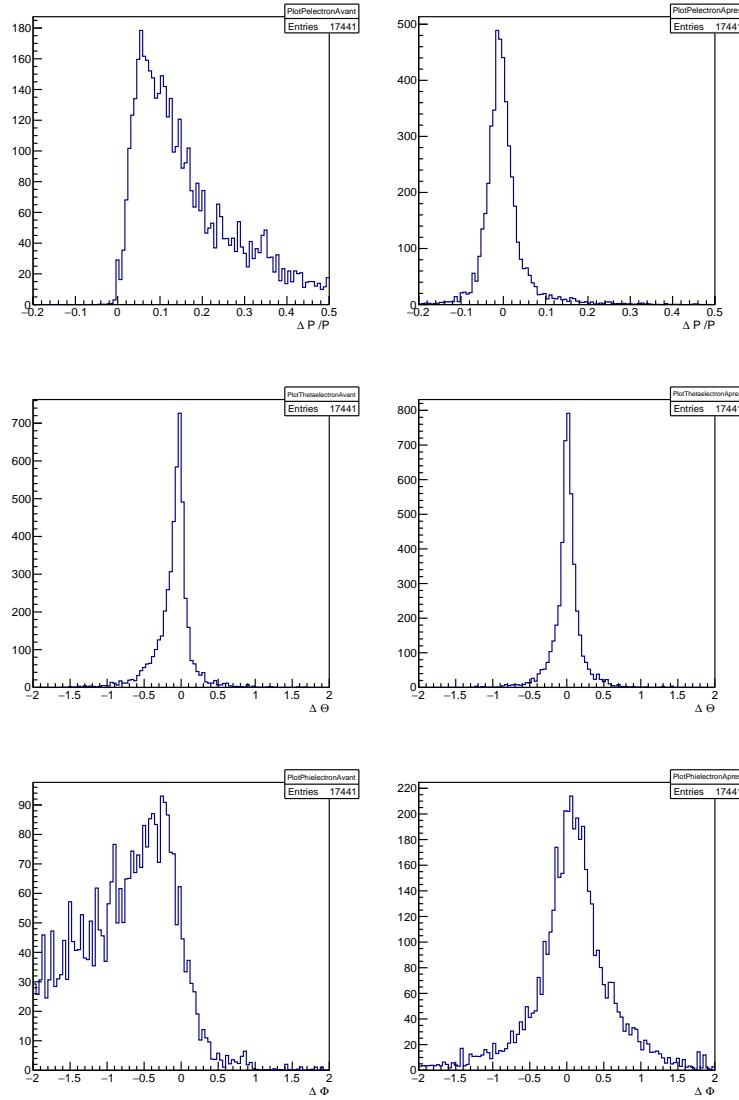


Figure 2.44: Distributions for electrons of the momentum resolution (top), the θ difference (middle) and the ϕ difference (bottom) before and after applying the detected radiated photon correction to the simulated electrons (in the left and right plots, respectively).

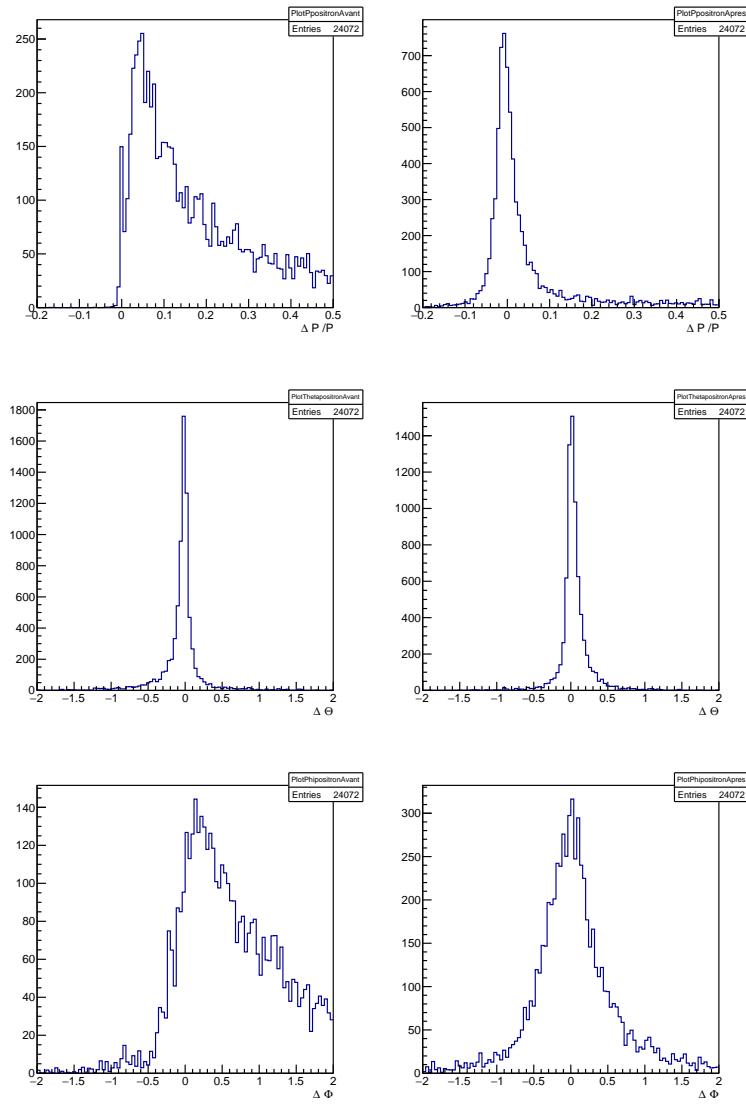


Figure 2.45: Distributions for positrons of the momentum resolution (top), the θ difference (middle) and the ϕ difference (bottom) before and after applying the detected radiated photon correction to the simulated positrons (in the left and right plots, respectively).

2.6 Fiducial cuts

Fiducial cuts consist in excluding regions where the detection efficiency is not well reproduced by the simulations. This mismatch between the data and the simulations mainly occurs on the edge of the detectors, where the detection efficiency varies fast. For calorimeters, this occurs in regions where the energy deposition is incomplete. In this analysis we develop a fiducial cut for the PCAL. This cut removes parts of the detector where electromagnetic showers are reconstructed too close to the edge of the active region. This ensures that most of the energy of an electromagnetic shower reconstructed in the PCAL is measured.

The PCAL fiducial cut developed in this section is based on the width of the measured electromagnetic shower. The definition of the square of the shower width in the PCAL is given in Formula (2.6). The mean shower size $\sqrt{M_2}$ is calculated for each of the three calorimeter coordinates (U , V and W). Figures 2.46 and 2.47 show the distributions, for sectors 1 to 3, of the mean shower size as a function of respectively low and high values of U , V and W . The shower mean size is estimated away from the edge of the calorimeter (red lines). The value of the cut is then set at a distance equal to the mean shower size from the edge of the distribution (black lines).

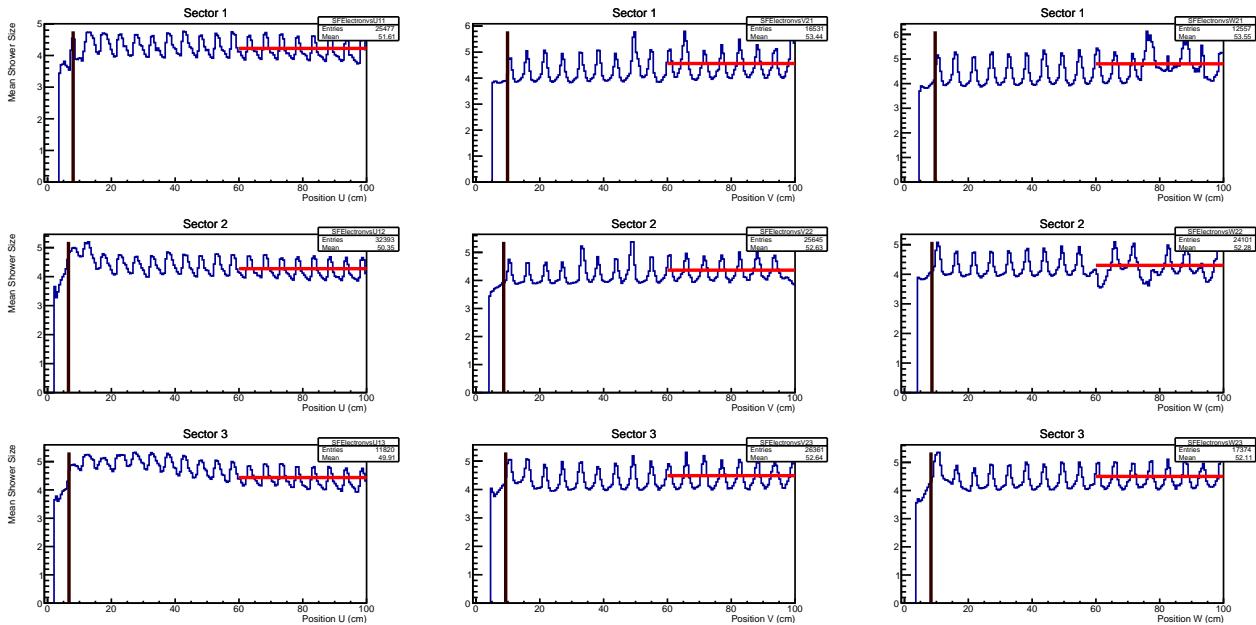


Figure 2.46: Mean shower width as a function of U , V and W (low values) for three sectors of CLAS12. The red line corresponds to the mean shower size calculated away from the edge of the distribution. The black line corresponds to the fiducial cut.

The cut values used in the TCS analysis are summarized in Table 2.4, where the maximum and minimum values of V and W are given for each sector of CLAS12. Only the cuts along V and W are used, as cutting on the U bars was found to be redundant.

sector	V_{\min}	V_{\max}	W_{\min}	W_{\max}
1	9.78924	402.06	9.47359	393.895
2	8.62768	402.389	8.57818	402.064
3	9.23112	403.875	8.23956	403.622
4	19.2814	403.021	8.26354	392.355
5	8.73336	402.915	9.28017	403.634
6	9.12088	403.681	8.13996	403.886

Table 2.4: Values of the fiducial cuts used. The minimum and maximum values for the U and V coordinates in the PCAL are given for each sector.

The effect of the PCAL fiducial cuts on the electron distribution is shown in Figure 2.48. The

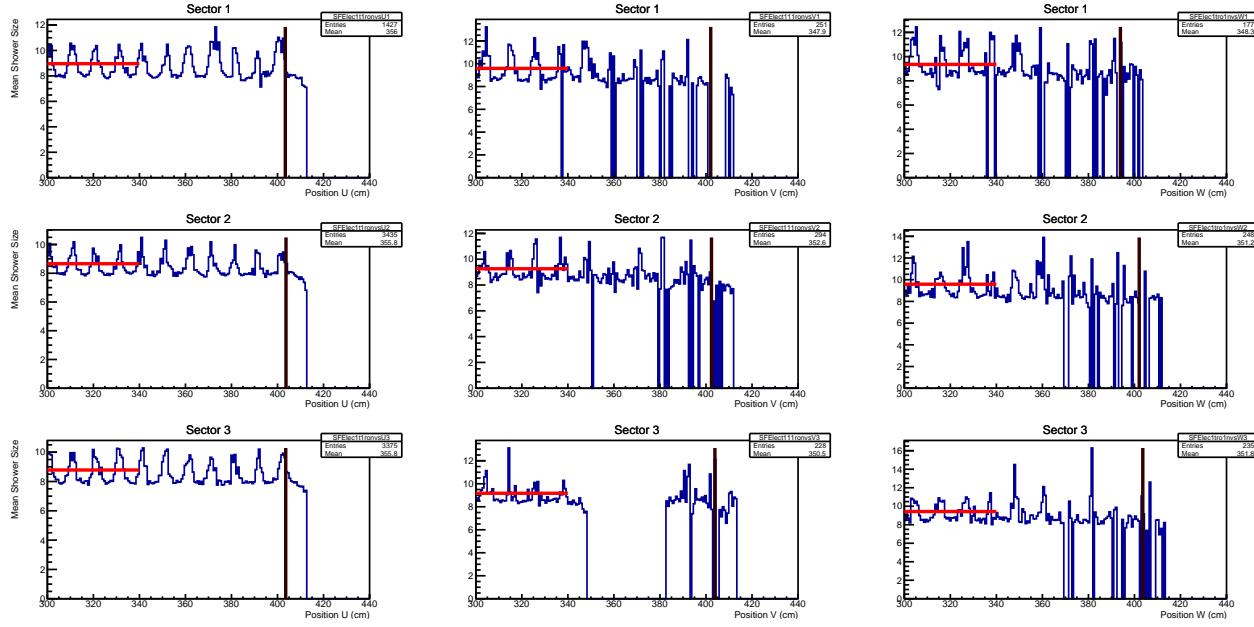


Figure 2.47: Same plots as in Figure 2.46 for the case of high values of U, V and W.

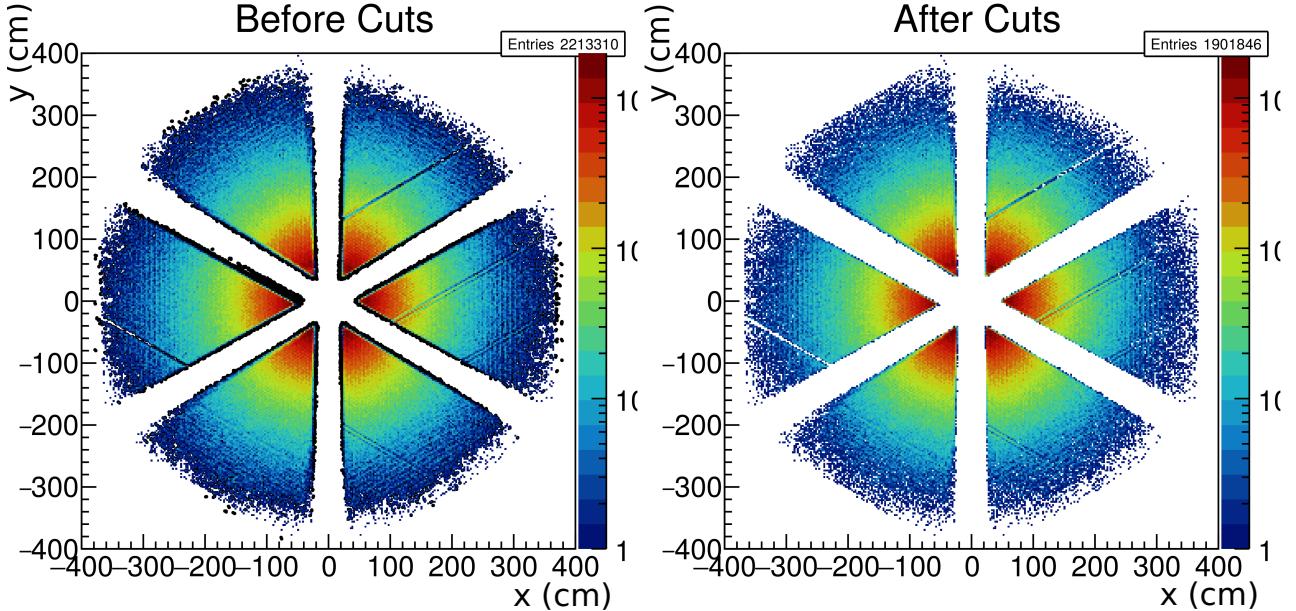


Figure 2.48: Effect of the fiducial cuts on the electron distribution in the natural coordinate plane of the PCAL. The left plot shows the electrons which are removed in black. The plot on the right shows all the electrons that passed the PCAL fiducial cuts.

730 left plot shows the removed electrons in black. The plot on the right shows the regions kept for the
 731 analysis. The fraction of electrons lost by these fiducial cuts was estimated on inclusive events (at least
 732 one electron detected in CLAS12) from the inbending data set. It ranges from 5% at 1 GeV to up to
 733 30% at 10 GeV.

734 Other approaches have been used to define fiducial cuts on the EC of CLAS12 [8]. The values of
 735 the cuts derived with these methods are similar to the ones presented here. For this reason, the TCS
 736 analysis presented in the next chapter adopts the values of the cuts defined in this section.

737 2.7 Background merging

738 The background merging consists in mixing data events recorded with random trigger with simulation
 739 events. The random trigger events are recorded regularly (with a frequency of few hundreds hertz)
 740 during the data taking. For each simulation event, the ADCs and TDCs of the CTOF, FTOF, DCs,
 741 SVT, MVT, EC and HTCC from a random trigger event are added to the list of ADCs and TDCs
 742 obtained from the GEMC simulation. The merged events are then reconstructed with the standard
 743 reconstruction software. This background merging procedure aims at better reproducing the detection
 744 efficiencies in the simulation by mimicking the backgrounds present during the actual data taking.

745 In the TCS analysis, the background merging procedure is used in the acceptance calculation in
 746 order to reproduce detector efficiencies.

747 748 *Start of common analysis section*

749 2.8 Proton efficiency correction

750 In order to take into account the differences in the proton detection efficiency between real data and
 751 simulations, a proton efficiency correction to the GEMC simulation is implemented. This correction is
 752 derived using the same data sample as for the proton momentum corrections discussed in Subsection
 753 2.4.2. The $e p \rightarrow e(p')\rho \rightarrow e(p')\pi^+\pi^-$ reaction is selected by applying a cut on the invariant mass of the
 754 two pions, $0.6 \text{ GeV} < M_{\pi^+\pi^-} < 1 \text{ GeV}$. The same reaction is generated using the *genev* event generator
 755 [12] and passed through the GEMC and the CLAS12 reconstruction softwares. The kinematics of the
 756 missing proton are assumed to be well reconstructed and are used to derive the correction. The MC and
 757 data driven corrections described in Section 2.4 are applied prior to the computation of this correction.
 758 The proton efficiency is measured for data and simulations as:

$$A \cdot Eff^{\text{Data/Simu.}}(\Omega_{Mis.}) = \frac{N_{Rec.}^{\text{Data/Simu.}}(\Omega_{Mis.})}{N_{Mis.}^{\text{Data/Simu.}}(\Omega_{Mis.})}, \quad (2.27)$$

759 where $N_{Mis.}^{\text{Data/Simu.}}(\Omega_{Mis.})$ is the number of events with a missing proton (ie. the mass of the missing
 760 particle is compatible with the mass of a proton) in the kinematic bin $\Omega_{Mis.} = P_{Mis.}; \theta_{Mis.}; \phi_{Mis.}$,
 761 $N_{Rec.}^{\text{Data/Simu.}}(\Omega_{Mis.})$ is the corresponding number of events with a detected proton, and A is the geo-
 762 metrical acceptance of CLAS12. The proton efficiency correction is then encoded in the ratio:

$$Eff_{Corr} = \frac{A \cdot Eff^{\text{Data}}}{A \cdot Eff^{\text{Simu.}}} = \frac{Eff^{\text{Data}}}{Eff^{\text{Simu.}}}. \quad (2.28)$$

763 The correction is computed in the CD and in the FD independently, using similar procedures described
 764 in the next two subsections.

765 2.8.1 Efficiency correction in the central detector

766 As shown in Figure 2.37b, there is very little background under the missing-proton mass peak in
 767 the high-polar-angle region. The number of events with a missing proton or a detected proton is then
 768 given by the number of events in each bin. The integrated efficiencies as a function of the momentum,
 769 the polar and the azimuthal angles of the missing proton are shown in Figure 2.49. The efficiency
 770 calculated in the simulations case is higher than for the data. The efficiency correction is calculated
 771 as a function of the three variables, with 2 bins in θ (from 37° to 45° and from 45° to 65°), 4 bins
 772 in momentum (spanning the 0.4 to 1.5 GeV range evenly) and 18 bins in ϕ (from -180° to 180° , 20°
 773 bins). The limits of the binning are driven by the variation of the correction as a function of each
 774 variable. Figure 2.50 shows the value of the corrections.

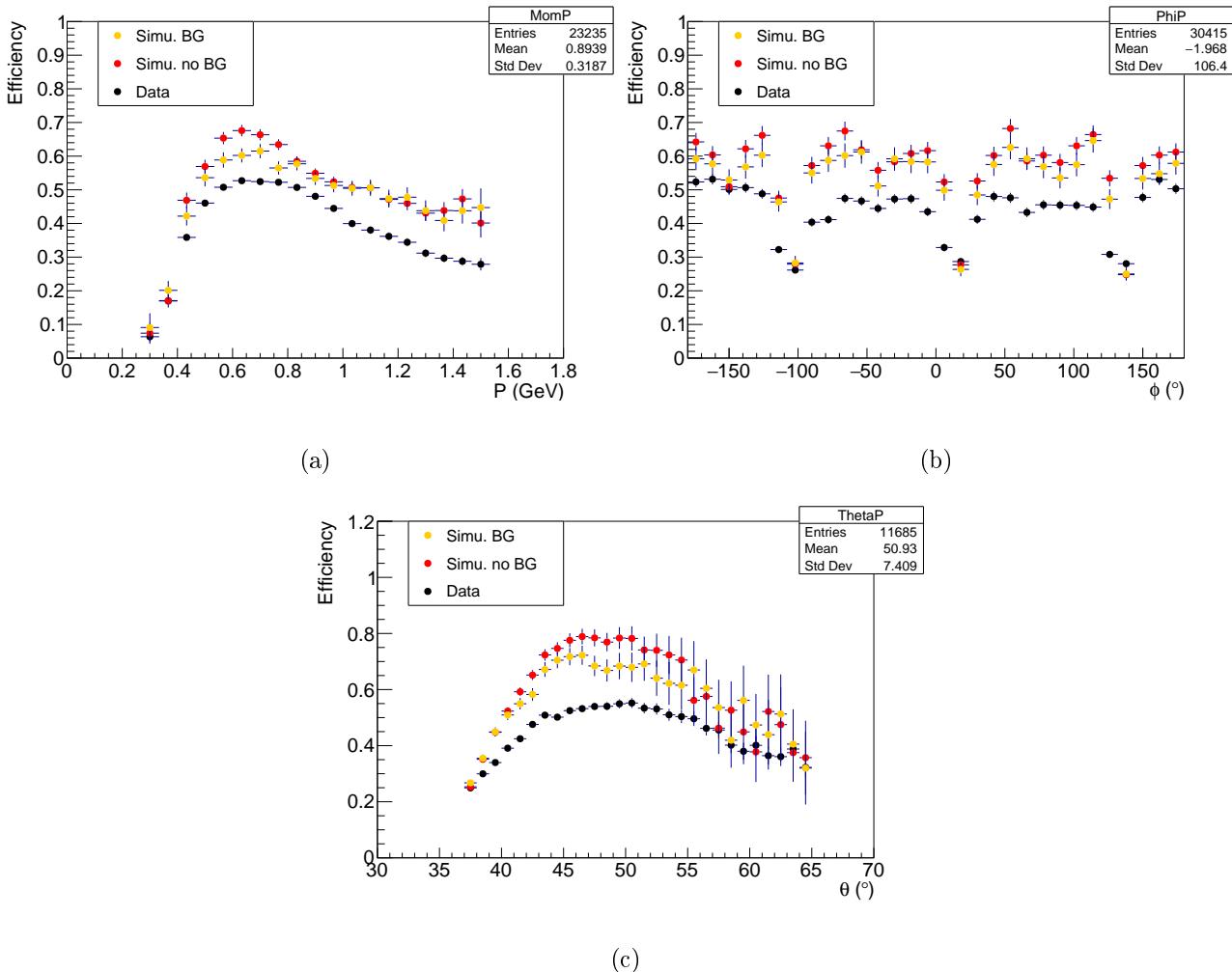


Figure 2.49: Proton efficiency in the CD, as a function of (a) the momentum, (b) the azimuthal angle and (c) the polar angle of the missing proton; for simulations without BG merging in red, with BG merging in yellow and data in black.

2.8.2 Efficiency correction in the forward detector

The proton efficiency is also derived for FD protons. Unlike in the case of the CD, where there is no background under the proton peak, there is a large high-mass background in the missing mass spectrum for protons at polar angles below 37° . In this region the number of events with a missing or reconstructed proton is calculated by fitting the missing proton peak with a gaussian plus a linear background as shown in Figure 2.51. The integral of the gaussian defines the number of events. The fit on the missing mass peak is done in both the missing-proton case and the reconstructed-proton case. In the case of the reconstructed proton, only the events with a reconstructed proton are kept in the missing mass spectrum fit.

Figure 2.52 shows the efficiency as a function of the momentum, the azimuthal and the polar angles of the missing proton. As the ratios between simulations and data efficiencies are fairly constant in θ and ϕ , a single differential correction in momentum is applied according to the efficiencies shown in Figure 2.52a. The correction values are given in Table 2.5. Note that the momentum bin $[0.3, 0.42]$ is assigned the same correction factor as the next bin as the number of data points is too low to allow correct extraction of the efficiency. This change has no effect on the TCS analysis, as most protons have momenta above 0.4 GeV in the CD.

End of common analysis section

Momentum (GeV)	0.3	0.42	0.54	0.66	0.78	0.9	1.02	1.14	1.26	1.38
Eff_{Corr}	0.665*	0.665	0.839	0.957	0.957	1.01	0.963	0.944	0.937	0.916

Table 2.5: Values of the efficiency correction in the FD as a function of the proton momentum.

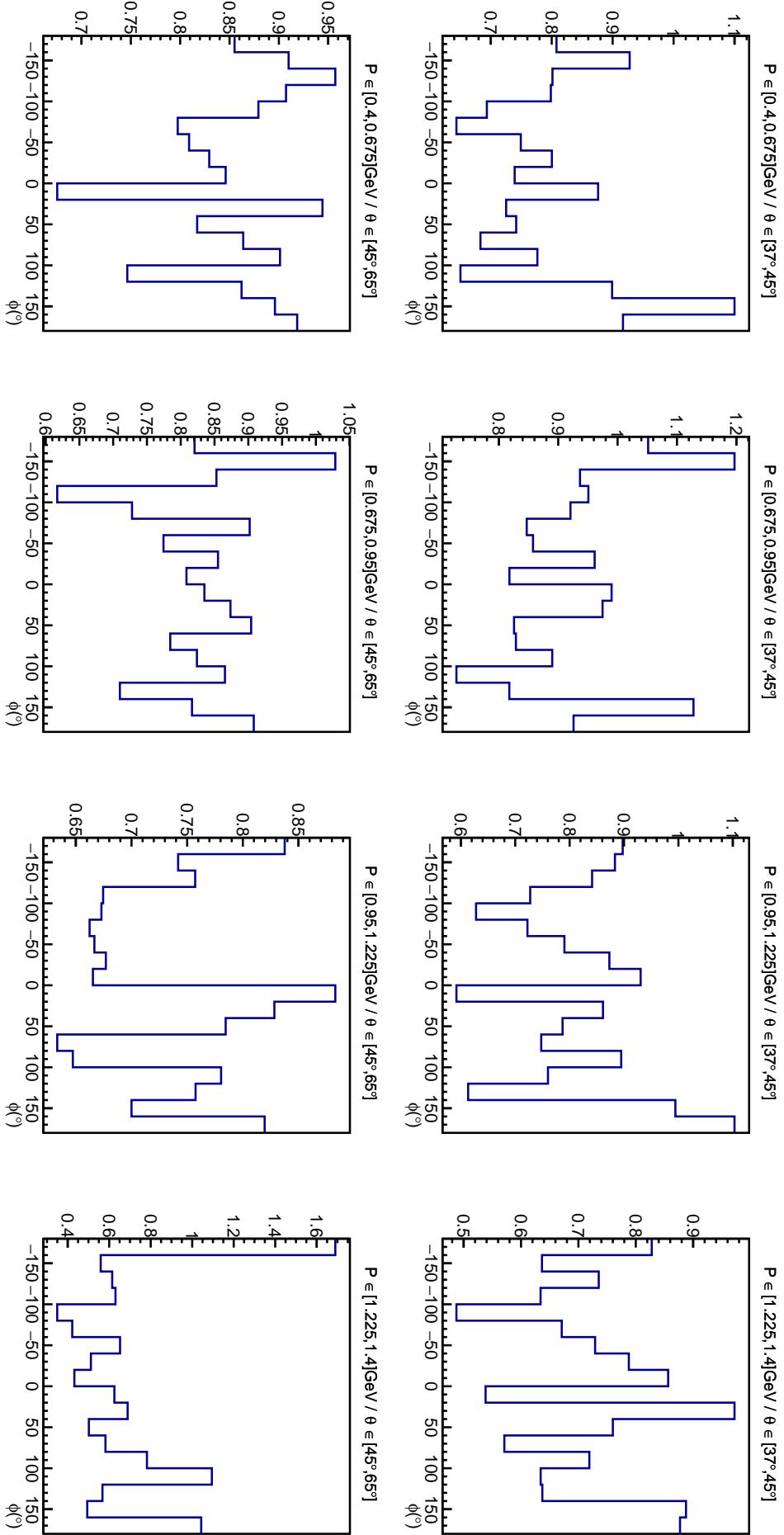


Figure 2.50: Efficiency correction in the CD for two bins in polar angle θ and 4 momenta bins. The correction is given as a function of azimuthal angle of the proton ϕ .

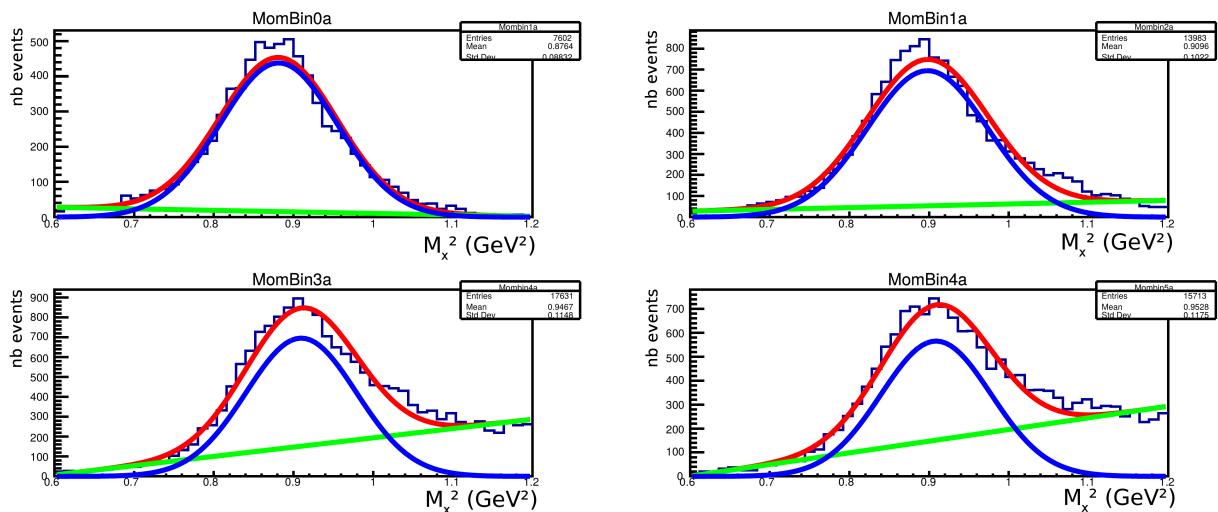


Figure 2.51: Examples of fits performed for the proton efficiency analysis in the FD. The missing mass peak is fitted with a gaussian and a linear background.

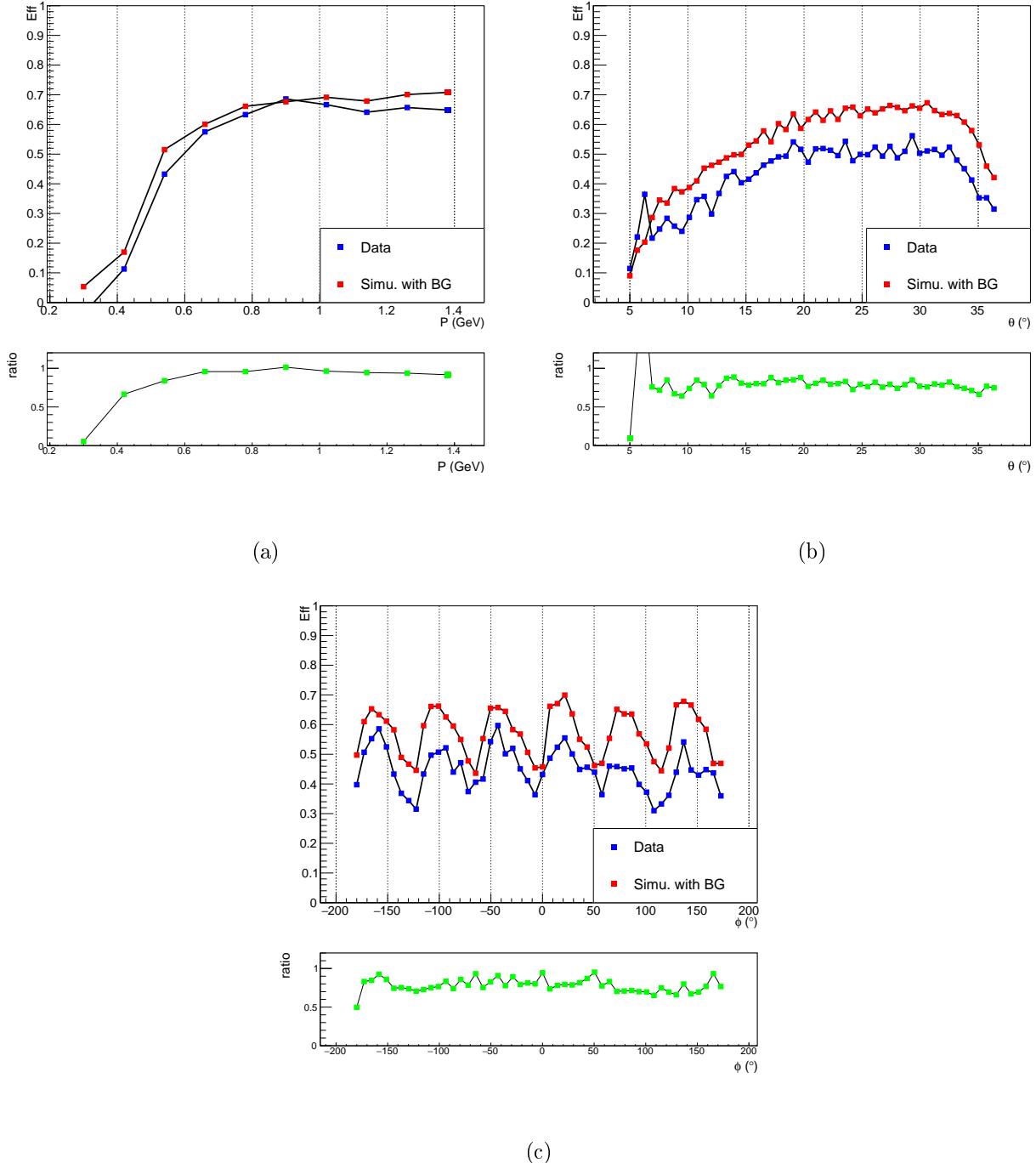


Figure 2.52: Proton efficiency in the FD, as a function of (a) the momentum, (b) the azimuthal angle and (c) the polar angle of the missing proton; for simulations in red and data in blue.

793 **Chapter 3**

794 **Simulations and extraction of the TCS
795 observables**

796 In this chapter, the various steps towards the experimental measurement of the TCS observables are
797 presented. The simulation software is presented in Section 3.1. Exclusivity cuts based on simulations
798 and used for event selection are justified in Section 3.2. The comparison of the kinematic distributions
799 of data and simulation is reported in Section 3.3. The acceptance study is displayed in Section 3.4. The
800 background reactions for the TCS events are discussed in Section 3.5. The formulae used to calculate
801 the TCS observables from data are detailed in Sections from 3.6 to 3.9. The binning used for data is
802 displayed in Section 3.11. Finally the systematic errors are estimated in Section 3.12. Results from
803 this analysis will be presented in the next chapter.

804 **3.1 TCS event generator and simulations**

805 Simulations of the $\gamma p \rightarrow e^- e^+ p'$ reaction are necessary for the analysis of TCS. They are used
806 especially to determine relevant exclusivity cuts as well as estimating the acceptance of CLAS12 for
807 this reaction. Two independent generators are used in this analysis. The first one, *TCSGen*, was
808 developed during the exploratory TCS analysis of CLAS data [3]. A second generator developed for
809 HERA, called *GRAPE* is also used to cross-check *TCSGen*, as well as to explore pair production
810 background and interference between final-state electrons (see Subsection 3.5.1).

811 **3.1.1 GRAPE**

812 The *GRAPE* Generator [9] was developed for experiments running at HERA, at the DESY laboratory
813 in Hamburg. It is a di-lepton generator to study the electromagnetic background mainly in J/Ψ and Υ
814 measurements. It is a full generator including the kinematics of the beam electron producing the real
815 photon. It also includes pair production from the incident electron as well as interferences between
816 leptons in the final state. The diagrams which can be included in the generator are shown in Figure
817 3.12. It is a non-weighted event generator.

818 **3.1.2 TCSGen**

819 *TCSGen* is a generator developed by R.Paremuzyan for the first CLAS analysis of TCS, and com-
820 pleted and corrected for the purpose of this analysis. It generates weighted events, with the possibility
821 to use BH-only weights or to include the TCS-BH interference cross section. Contrary to *GRAPE*,
822 the initial electron from the beam is not included and no possible interferences are taken into account.
823 Each event is assigned a weight w equal to:

$$w = psf \cdot \sigma \cdot flux_{\gamma}, \quad (3.1)$$

824 where psf is the phase space factor, σ is the cross section of the considered processes, and $flux_{\gamma}$ is
825 the equivalent photon flux given in [10]. The phase space factor is the product of the ranges allowed

826 for each kinematic variable. First the energy of the incoming real photon is randomly picked between
 827 a minimum energy $E_{\gamma \text{ Min}}$, defined by the user of the generator, and the maximum possible energy,
 828 equal to the electron beam energy E_b . The photon energy phase-space factor is $\text{psf}_{E_{\gamma}} = E_b - E_{\gamma \text{ Min}}$.
 829 Second, the range allowed for the square of the transferred momentum to the proton $-t$ is fully defined
 830 by the value of E_{γ} previously picked. The limit t_{\min} and t_{\max} of the possible range for $-t$ are given
 831 by Equation (5.10) in page 121 of [11]. The associated phase space factor is $\text{psf}_t = t_{\max} - t_{\min}$. After
 832 randomly picking a value for $-t$, the kinematically accessible invariant mass is fully defined. The
 833 associated phase space is $\text{psf}_M = M_{\text{Max}} - M_{\text{Min}}$. The limits $M_{\text{Min/Max}}$ are also given in [11]. Finally
 834 the center-of-mass angles θ and ϕ are randomly chosen, in the range $[0 - \pi]$ and $[0 - 2\pi]$, respectively.
 835 The final-state particles are then boosted to the lab frame and a final azimuthal rotation is performed.
 836 The formula for the phase-space factor is:

$$\text{psf} = \text{psf}_{E_{\gamma}} \cdot \text{psf}_t \cdot \text{psf}_M \cdot \text{psf}_{\theta \text{ CoM}} \cdot \text{psf}_{\phi \text{ CoM}} \cdot \text{psf}_{\phi \text{ Lab.}} \quad (3.2)$$

837 The cross section σ can be either the BH one or include the TCS interference term. The formulae
 838 used are those from [1]. In the following, the simulations were produced using the BH cross section
 839 alone. This was chosen as the CFFs tables available at the time in the generator were not up to date.
 840 In this case, the BSA and the A_{FB} should be vanishing.

841 The kinematic distributions of events generated with *TCSGen* and with *GRAPE* in the same kine-
 842 matic region and with BH cross section only are compared in Appendix B. The two generators agree,
 843 up to a normalization constant. For the TCS analysis, the acceptance calculations were performed
 844 using *TCSGen* as it allows to cover the whole phase space with less iterations than *GRAPE*, as it is a
 845 weighted generator and no generated events are discarded.

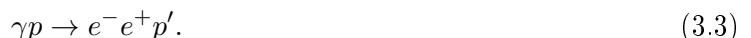
846 3.2 Event selection

847 3.2.1 Final state selection

848 The initial step of the extraction of TCS observables from the complete RGA dataset is the event
 849 selection. First the final state of interest is selected. The complete dataset is skimmed to select events
 850 with at exactly one proton, one electron, one positron and any other particles. We allow any other
 851 particles in the event to avoid removing good events where accidental TOF/EC hits or false tracks
 852 are recorded. The particles are selected according to the pid given by the CLAS12 reconstruction
 853 software (see Chapter 2), and using the dedicated positron identification neural network described in
 854 Chapter 2. The cut applied on the output of the neural network is set at 0.5. The momenta of the
 855 detected particles are corrected and fiducial cuts are applied. Finally a cut on the lepton momentum is
 856 also applied (leptons with momenta greater than 1 GeV are kept). This cut is motivated by the poor
 857 momentum reconstruction for tracks with large curvature (low momentum) in the FD.

858 3.2.2 Exclusivity cuts

859 Once the events with the relevant final state are retrieved, exclusivity cuts are applied. The reaction
 860 of interest is :



861 However the RGA dataset was not taken with a beam of photons but with electrons impinging directly
 862 on the target. Therefore quasi-real photo-production events are used in this analysis. A quasi-real
 863 photon is emitted by an electron from the beam and it interacts with the proton, producing a $e^+ e^-$
 864 pair in the final state, as:



865 The corresponding conservation of momentum equation is:

$$p_{\text{beam}} + p_{\text{target}} = p_{\text{scat.}} + p_{\gamma} + p_{\text{target}} = p_{\text{scat.}} + p_{e^+} + p_{e^-} + p_p. \quad (3.5)$$

3.2. Event selection

866 The 4-momenta of the scattered electron and initial real photon are fully defined by the measurement
 867 of the 4-momenta of the final state particles:

$$p_\gamma = p_{e^+} + p_{e^-} + p_p - p_{\text{target}}, \quad (3.6)$$

868

$$p_{\text{scat.}} = p_{\text{beam}} + p_{\text{target}} - (p_{e^+} + p_{e^-} + p_p). \quad (3.7)$$

869 Two exclusivity cuts are applied on the missing particle (X) of the pe^+e^-X system. The mass and
 870 the transverse momentum fraction of the missing particle X are constrained to be close to zero. The
 871 mass cut ensures that the missing particle is an electron. The transverse momentum cut ensures the
 872 low virtuality of the photon. Indeed, the virtuality of the incoming photon can be written:

$$Q^2 = 2E_b E_X (1 - \cos(\theta_X)), \quad (3.8)$$

873 where E_b is the energy of the electron beam, E_X is the energy of the undetected scattered electron,
 874 θ_X is its scattering angle in the lab frame given by:

$$\sin(\theta_X) = \frac{Pt_X}{P_X}, \quad (3.9)$$

875 and P_X and Pt_X are, respectively, the momentum and transverse momentum of the missing particle.
 876 The values of these cuts are motivated by simulations. Figure 3.1a shows the distribution of simulated
 877 events in the transverse missing momentum fraction plane (left plot) and the transverse momentum
 878 fraction as a function of the missing mass (right plot). Photo-production events correspond to events
 879 at small transverse momentum fraction and small missing mass. The same distributions are shown for
 880 the data in Figure 3.1b. A similar behavior is seen, with the addition of high missing mass background.

881 The cuts used in the analysis are:

$$\left| \frac{Pt_X}{P_X} \right| < 0.05 \quad (3.10)$$

882 and

$$|M_X^2| < 0.4 \text{ GeV}^2. \quad (3.11)$$

883 The data/MC comparison shows compatible resolutions for transverse momentum fraction (left
 884 plot of Figure 3.2) while the width of the missing mass peak is not as well reproduced (see right plot
 885 of Figure 3.2). This is likely due to the fact that our protons are mainly in the CD and that the
 886 simulation is not yet well tuned to reproduce the CLAS12 performances. The MC was nonetheless
 887 necessary to guide the choice of the cuts. The values of the exclusivity cuts were picked in order to
 888 minimize the loss of good events on the MC, while at the same time cut out the background from non-
 889 photoproduction events on the data. To illustrate this statement, Figure 3.2 shows the distribution of
 890 the exclusivity variables for simulation and data as well as the cuts used in the analysis (green lines
 891 depict the standard cuts, red lines the tight cuts and blue lines the loose cuts).

892 The resulting distribution of incoming photon virtuality, calculated using Equation (3.8), is shown
 893 in Figure 3.3. The variation of the extracted quantities with different exclusivity cuts is investigated
 894 and accounted for in the systematic uncertainties (see Section 3.12).

895 Finally one can assess the exclusivity of the selected events by looking at two other constrained
 896 quantities. Assuming the real photon emitted by the beam propagates along the beam axis, one can
 897 directly relate the energy of the real photon with the z-component of the sum of momenta of the final
 898 state particles. The difference between the two ways to determine the photon energy is defined as:

$$\Delta E = p_z p' + p_z e^+ + p_z e^- - E_\gamma \quad (3.12)$$

899 where E_γ is obtained from Equation (3.6).

900 Another constrained quantity that was investigated is the square of the transferred momentum to
 901 the proton given by $t = (p'_p - p_p)^2$. Assuming the real incoming photon is along the beam direction
 902 and that its energy is given by $E_\gamma \simeq p_z p' + p_z e^+ + p_z e^-$ as discussed above, one can calculate t using

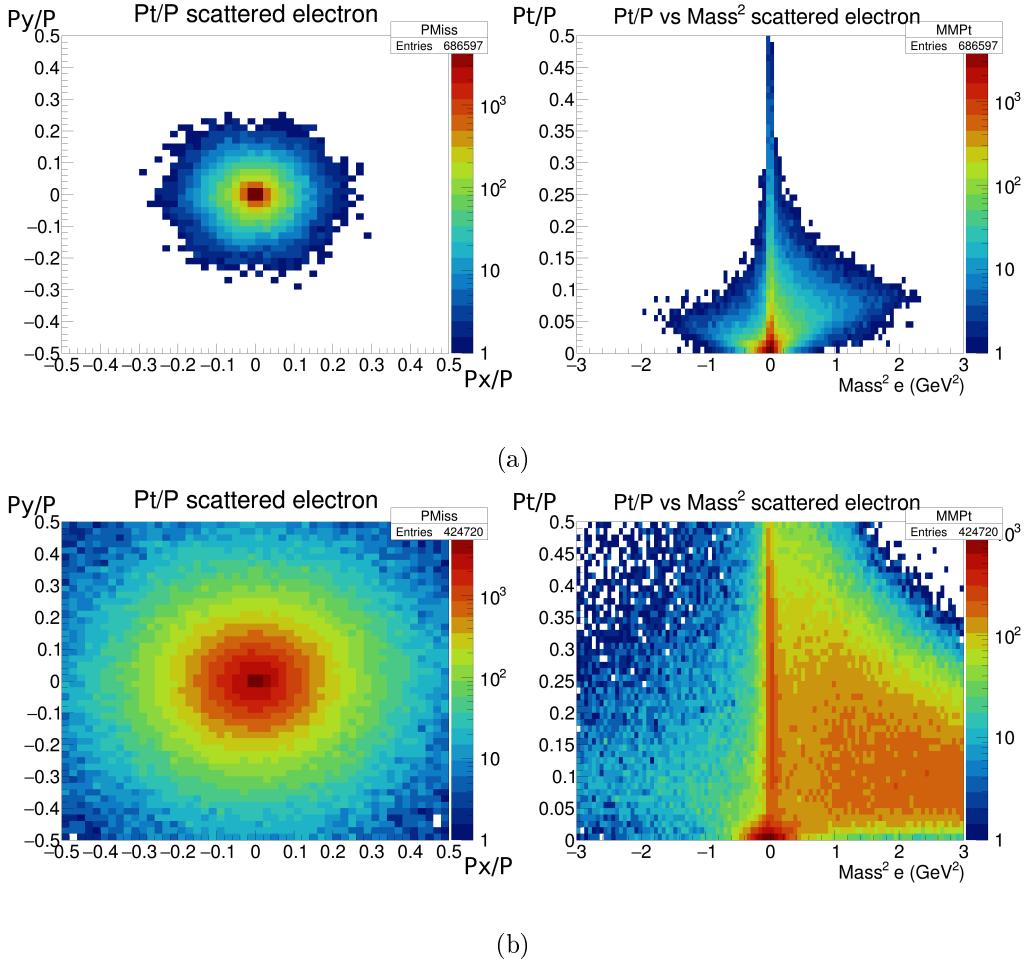


Figure 3.1: Distributions of the exclusivity variables for (a) TCS simulations and (b) the data.

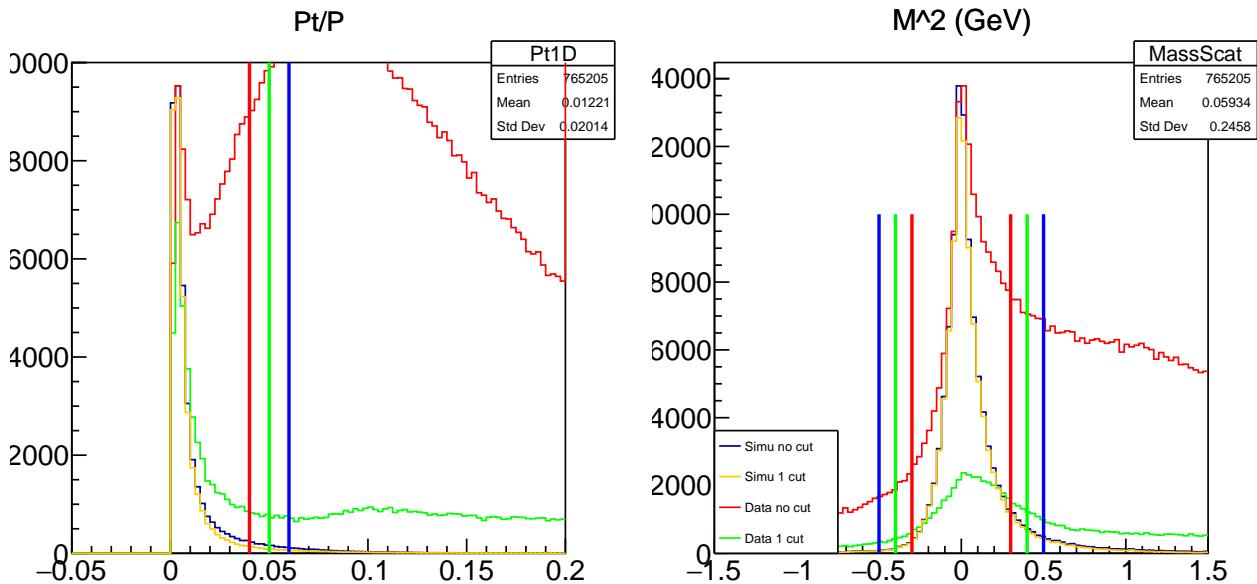


Figure 3.2: 1-Dimensional distributions of the exclusivity variables used in the analysis. The color code of the histograms is given in the legend box. The bold lines denotes the three sets of cuts used in the analysis. The green lines depicts the standard set of cuts, the red and blue lines depicts the tight and loose cuts respectively.

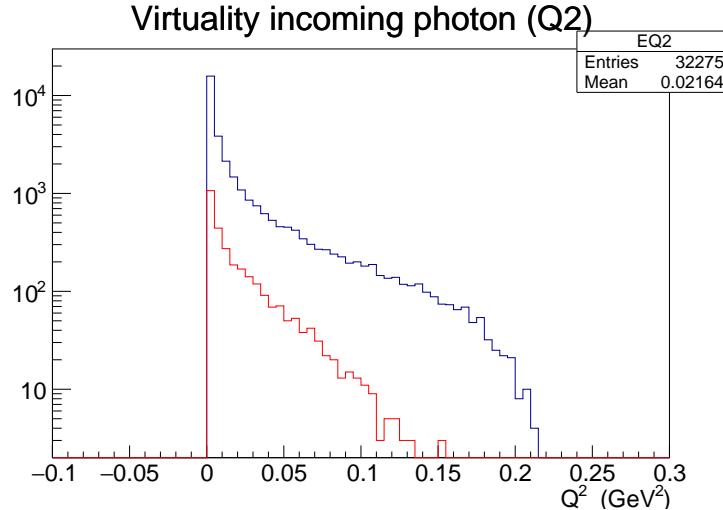


Figure 3.3: Virtuality of the incoming photon for all the events passing the exclusivity cuts (blue) and the events included in the analysis, ie within the kinematic cuts detailed in Section 3.3 (red).

903 the measured momenta of the final state leptons and only the z-component of the measured proton as
 904 $t = (p_{\gamma \text{ approx.}} - p_{e^+} - p_{e^-})^2$. The difference between the two calculations given by:

$$\Delta t = (p_{p'} - p_p)^2 - (p_{\gamma \text{ approx.}} - p_{e^+} - p_{e^-})^2 \quad (3.13)$$

905 allows to assess the exclusivity of the final events.

906 Figure 3.4 shows ΔE and Δt for all events with the TCS final state, events that passed the
 907 exclusivity cuts and final events that are included in the analysis after kinematic cuts (more details in
 908 Section 3.11). One can see that the exclusivity cuts strongly constrain these quantities.

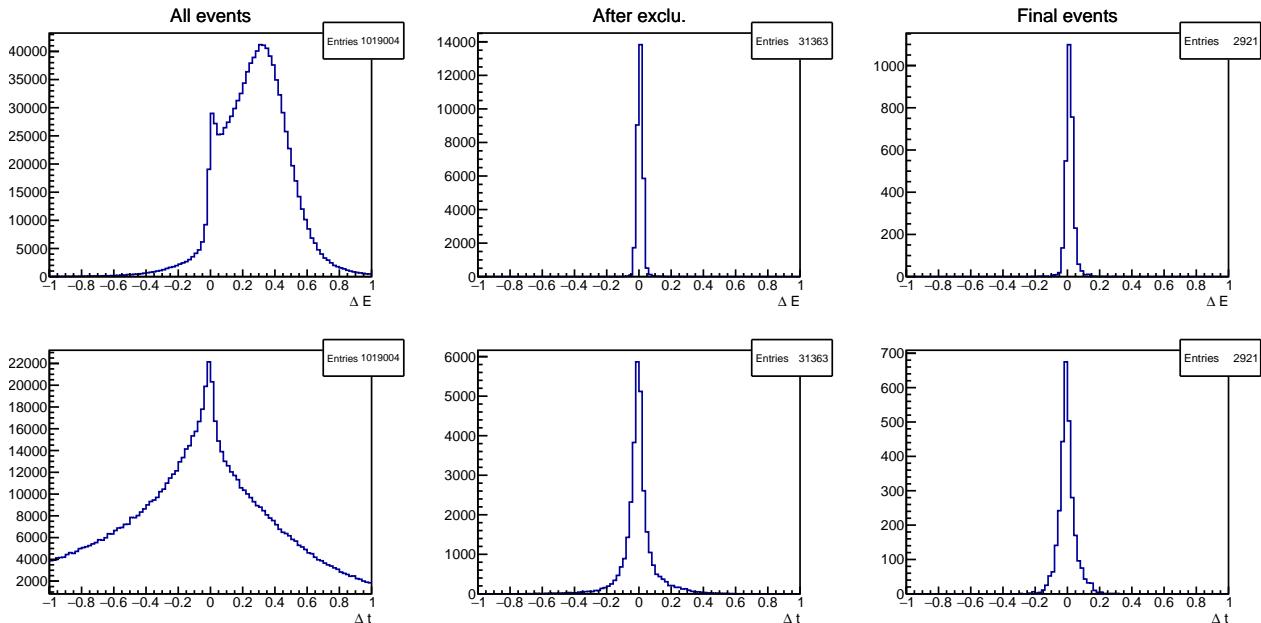


Figure 3.4: ΔE and Δt distributions as defined in Equations (3.12) and (3.13), for all events with the final state of interest in the inbending data set (left), after exclusivity cuts (middle), and for events included in the final analysis after kinematic cuts (right).

3.3 Phase space of interest and Simulations/Data comparison

After applying the exclusivity cuts detailed above, the invariant mass of the lepton pair is extracted. Figure 3.5 shows the invariant mass spectrum obtained for the full RGA Fall 2018 dataset. Meson resonances decaying into an electron-positron pair are visible.

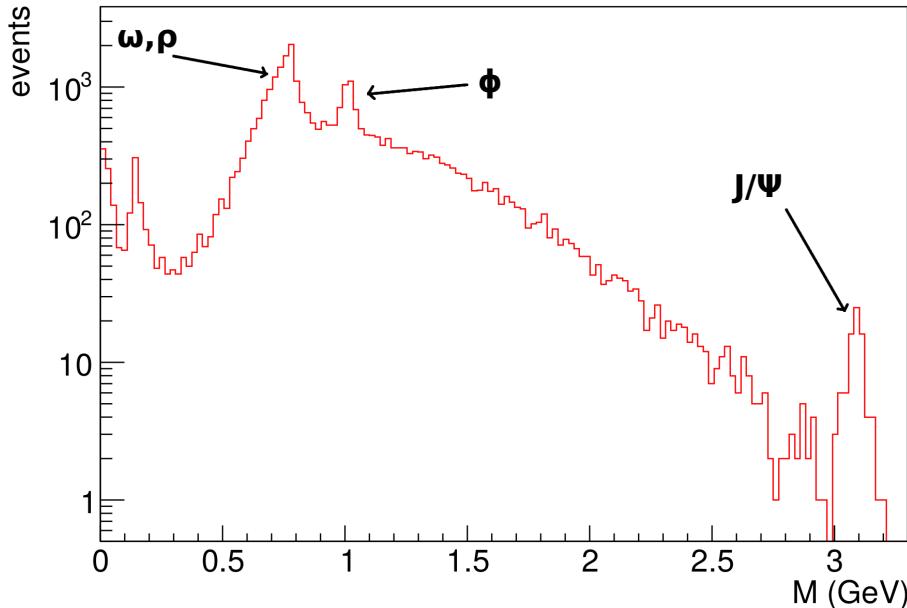


Figure 3.5: Lepton pair invariant mass spectrum after exclusivity cuts, extracted from data. Meson resonances (ω/ρ (770/782 MeV); ϕ (1020 MeV) and J/ψ (3096 MeV)) are visible. The peak at zero mass is due to photon conversion in the target material ($\gamma \rightarrow e^+e^-$) and π^0 Dalitz decay ($\pi^0 \rightarrow \gamma e^+e^-$). The peak at 0.1 GeV is an artifact induced by noise in the PMTs of the HTCC associated with two DC tracks in the same sector of CLAS12.

The mass region between 1.5 GeV and the J/ψ mass (3096 MeV) is selected for the measurement of TCS. The first reason for this choice is theoretical: the hard scale of TCS is given by the invariant mass of the lepton pair and has to verify $Q'^2 \gg m_p$ and $Q'^2 \gg -t$ in order to meet the conditions for the GPD factorization. As shown in the later analysis, a typical value of proton momentum transfer in this analysis is 0.3 GeV², which satisfy the previous condition. The second reason is that the mass range above 2 GeV is free from vector-meson resonances decaying into a di-lepton pair. The range between 1.5 GeV and 2 GeV has contributions from the wide vector mesons ρ (1450) and ρ (1700). In order to check the impact of these resonances, the mass spectrum obtained from the data is compared with BH-weighted simulations, between the ϕ and the J/ψ mass, in Figure 3.6. In this figure the simulated spectrum is normalized in order to have equal integral as the data spectrum, between 1.1 and 3 GeV. According to this plot there is no obvious meson contamination in the mass range of interest. The effect of this contamination on the observables is studied by extracting them in the two mass ranges ([1.5 GeV – 2 GeV] and [2 GeV – 3 GeV]).

Eventually, the phase space selected for the analysis is:

- $1.5 \text{ GeV} < M = \sqrt{Q'^2} < 3 \text{ GeV}$
- $0.15 \text{ GeV}^2 < -t < 0.8 \text{ GeV}^2$
- $4 \text{ GeV} < E_\gamma < 10.6 \text{ GeV}$.

Note that, unless specified otherwise, the observables are integrated over these kinematic ranges.

The distributions of the three variables $-t$, E_γ and M for data and simulations in the selected phase space are compared in Figure 3.7. The simulation distributions are normalized to have the same

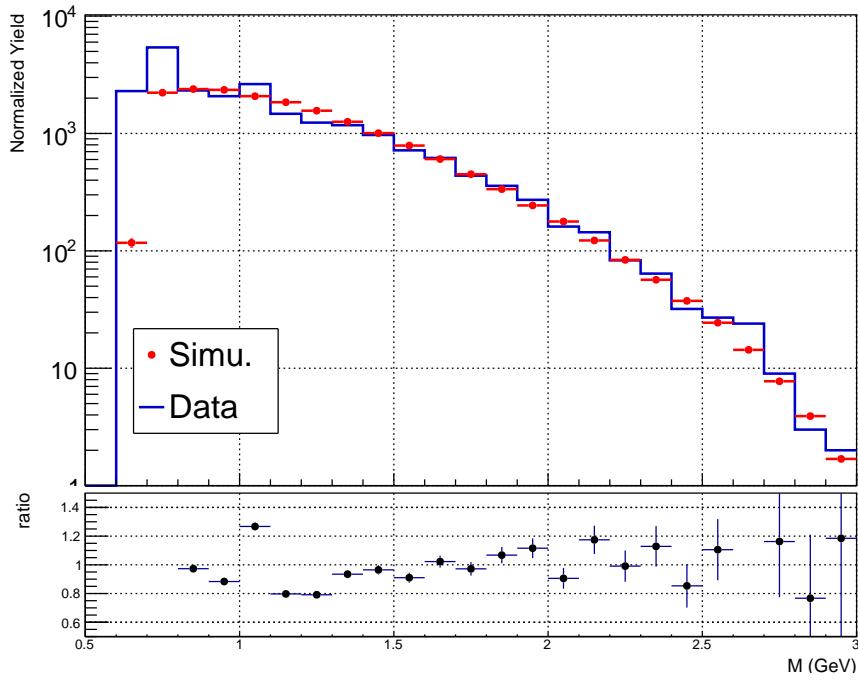


Figure 3.6: Comparison between data and simulations in the high-mass range. The ratio between the two plots is also shown. Cuts on the square of the transferred momentum ($0.15 \text{ GeV}^2 < -t < 0.8 \text{ GeV}^2$) and the photon energy ($2 \text{ GeV} < E_\gamma < 10 \text{ GeV}$) are applied.

933 integral as data distributions. The simulations-data agreement for all three variable is good and no
 934 large discrepancies are seen. The kinematics of the particles in the laboratory coordinates P and θ_{Lab}
 935 are displayed in Appendix C.

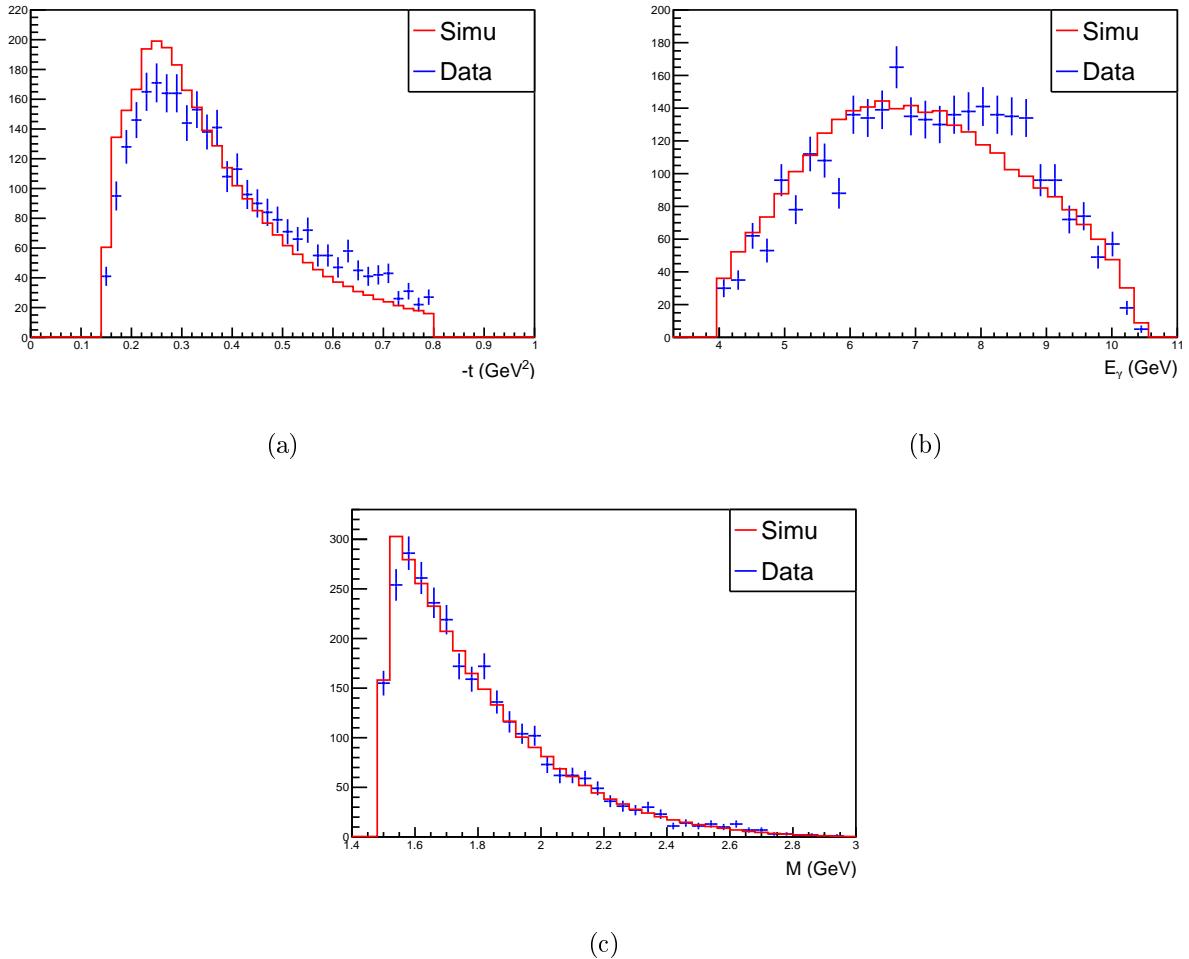


Figure 3.7: Simulations-data comparison for the distributions in (a) $-t$, (b) E_γ and (c) M . The simulation distributions are normalized to have equal integral as the data. Both the simulations and the data are in agreement, and no large discrepancies are seen.

3.4 Acceptance estimation

The large angular coverage of CLAS12 permits to detect most of the particles produced when an electron interacts with the target. However, due to holes between detector subsystems, some particles may escape without being detected. Furthermore, because of detection inefficiencies, the reconstructed particle kinematics may differ from their initial values at the vertex. In order to correct for these effects, the acceptance of CLAS12 for the $\gamma p \rightarrow e^- e^+ p'$ reaction is estimated using simulations including background merging. In the following, acceptance refers to the product of the geometrical acceptance of the detector convoluted with its efficiency. The acceptance is calculated in 5-dimensional bins. In a given bin \mathcal{B} , the acceptance is defined as the number of events reconstructed in this bin devided by the number of events generated in this bin:

$$Acc_{\mathcal{B}} = \frac{N_{\mathcal{B}}^{REC}}{N_{\mathcal{B}}^{GEN}}. \quad (3.14)$$

The number of reconstructed events in \mathcal{B} is:

$$N_{\mathcal{B}}^{REC} = \sum_{\{E_\gamma, -t, Q'^2, \theta, \phi\}_{REC} \in \mathcal{B}} w, \quad (3.15)$$

where w is the weight of the event given by *TCSGen* (Equation (3.1)) and the sum is performed over all events with reconstructed kinematics inside the considered bin \mathcal{B} . In the case where the data-driven

949 proton efficiency correction is used, the number of reconstructed events reads:

$$N_{\mathcal{B}}^{REC} = \sum_{\{E_\gamma, -t, Q'^2, \theta, \phi\}_{REC} \in \mathcal{B}} Eff_{Corr} w, \quad (3.16)$$

950 The number of generated events in \mathcal{B} is:

$$N_{\mathcal{B}}^{GEN} = \sum_{\{E_\gamma, -t, Q'^2, \theta, \phi\}_{GEN} \in \mathcal{B}} w, \quad (3.17)$$

951 with the same definition as before, except that the sum is now done over events with generated
 952 kinematics inside the bin \mathcal{B} . This definition encodes both acceptance and bin migration effects, provided
 953 that resolutions are well reproduced in simulations. Each event in the analysis is corrected by the
 954 acceptance factor corresponding to its reconstructed kinematics.

955 The binning used in the analysis is given in Table 3.1.

Variable	Bin limits
$-t$	0.15 - 0.25 - 0.34 - 0.48 - 0.8
E_γ	4.0 - 6.2 - 8.4 - 10.6
Q'^2	2.25 - 3.5 - 5.0 - 9.0
θ	from 30° to 160° , 10° bins
ϕ	from -180° to 180° , 10° bins

Table 3.1: Multi-dimensional binning used for the calculation of the acceptance.

956 For the results shown in Section 4.1, 36 million events were generated using *TCSGen*, passed
 957 through GEMC and reconstructed using the CLAS12 reconstruction software. In order to remove bins
 958 where there are too few events to perform the acceptance estimation, two additional cuts are used.
 959 The bins with acceptance below 5% are removed from the analysis. A cut on the relative error of the
 960 acceptance shown in Figure 3.8 is also applied:

$$\frac{\delta Acc}{Acc} < 50\%. \quad (3.18)$$

961 The resulting acceptance distributions are shown in Figures 3.9, 3.10, and 3.11, as a function of θ and
 962 ϕ for each $\{E_\gamma, -t, Q'^2\}$ bin.

963 Figures 3.9, 3.10 and 3.11 show the acceptance of CLAS12 for the TCS process. Each plot in the
 964 grids corresponds to a single $\{E_\gamma, -t, Q'^2\}$ bin, as defined in Table 3.1 in Chapter 3. The acceptance
 965 plots shown here are made prior to applying the acceptance quality cuts described in Section 3.4.

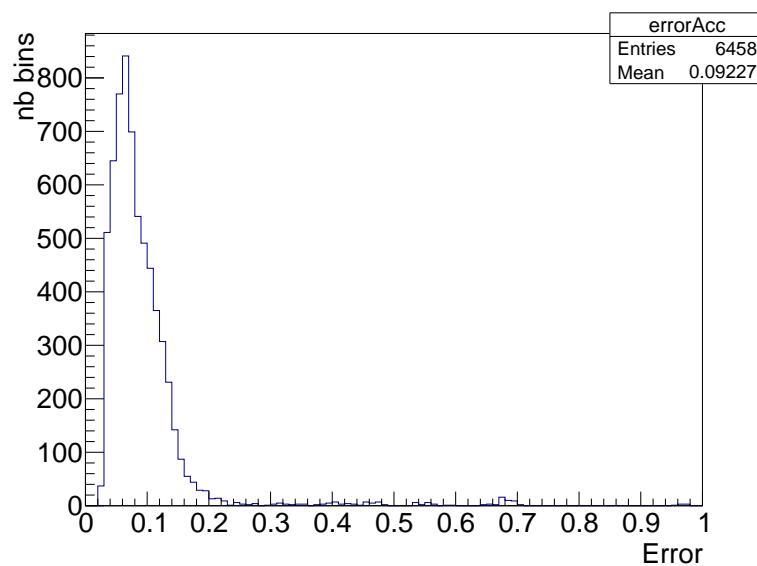
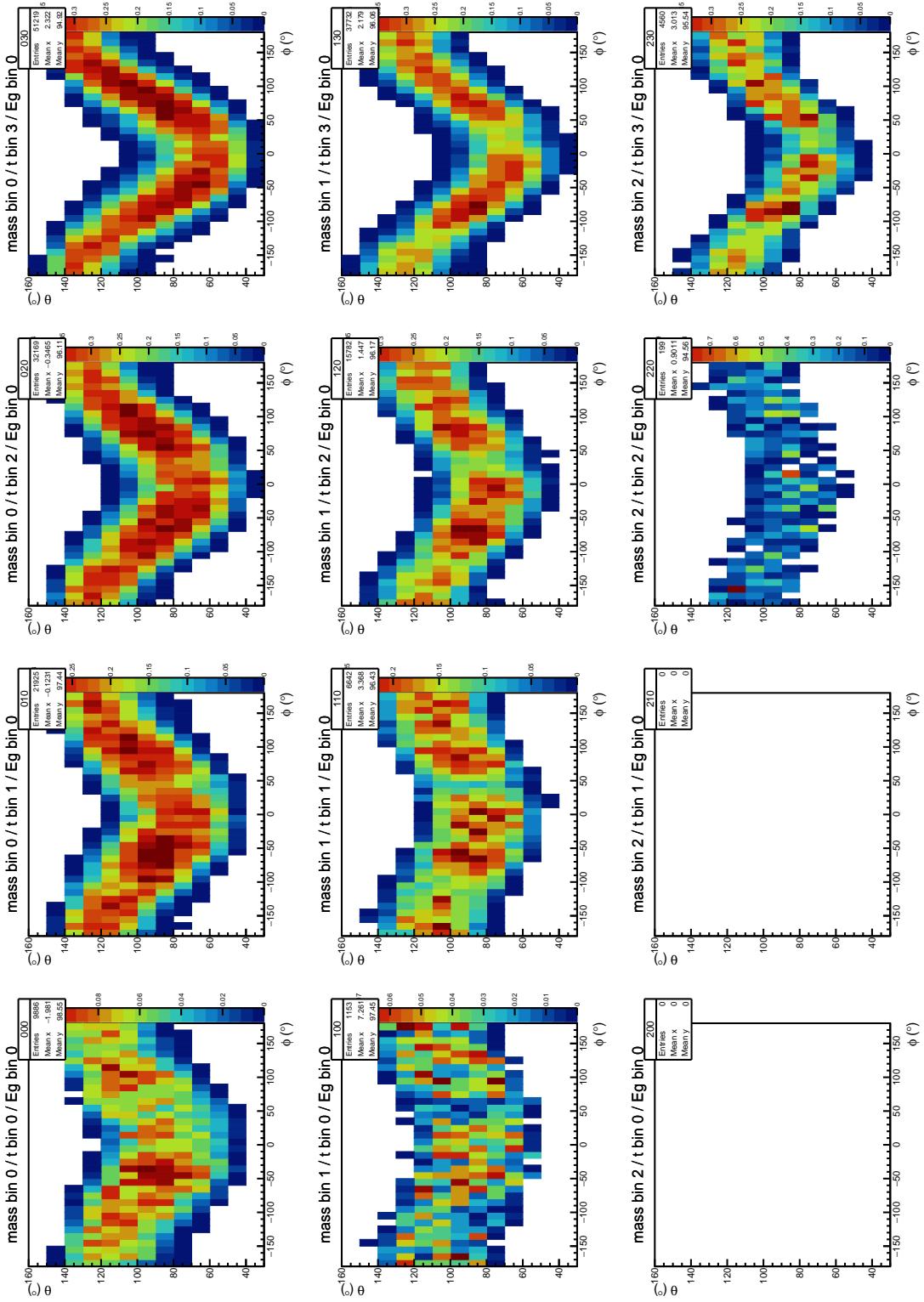


Figure 3.8: Relative error on the acceptance calculation, after bins with acceptance below 5% are removed. A cut is applied to remove bins with relative error higher than 50%.


 Figure 3.9: CLAS12 acceptance for TCS in the ϕ/θ plane for the first bin in E_γ .

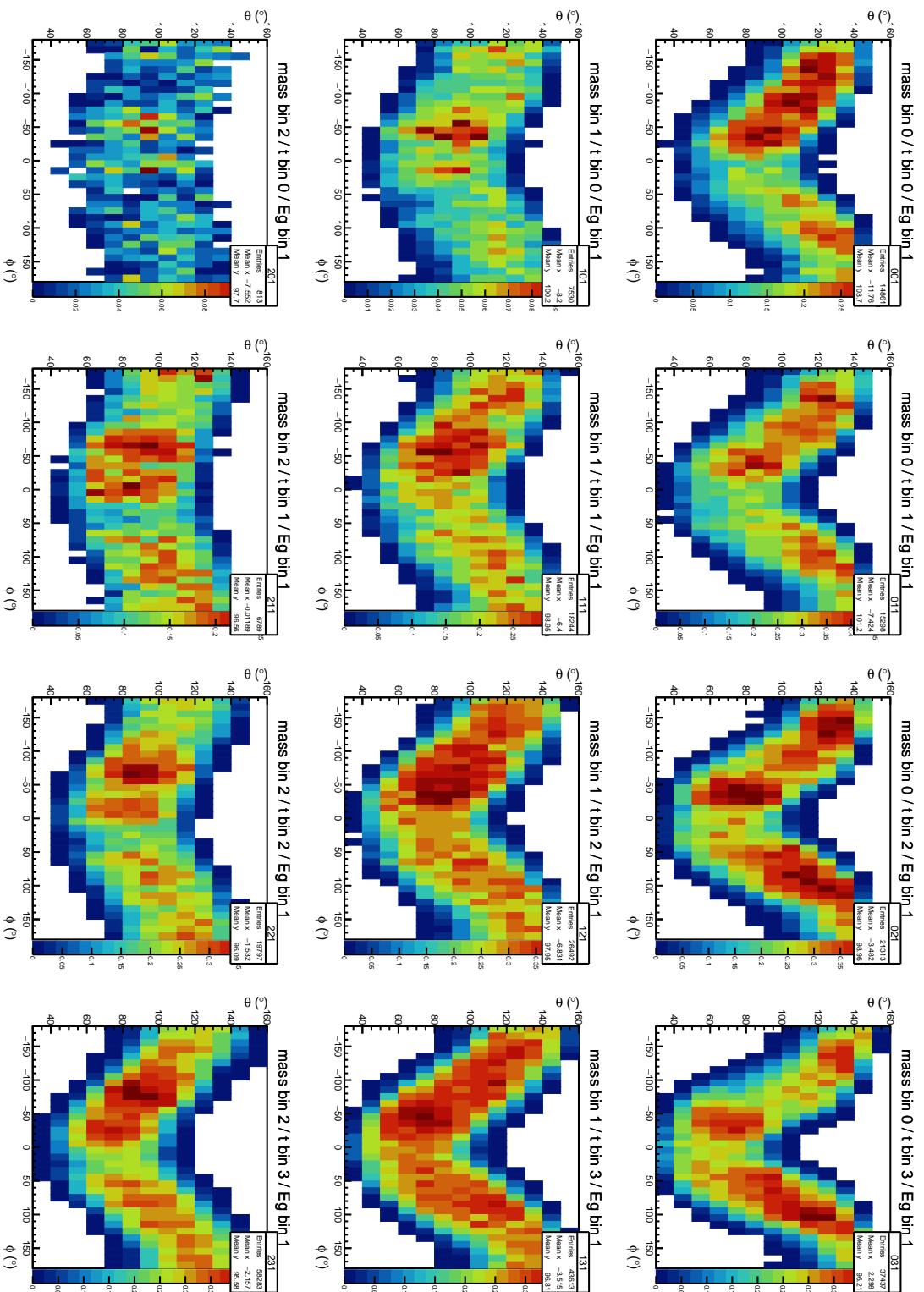


Figure 3.10: CLAS12 acceptance for TCS in the ϕ/θ plane for the second bin in E_γ .

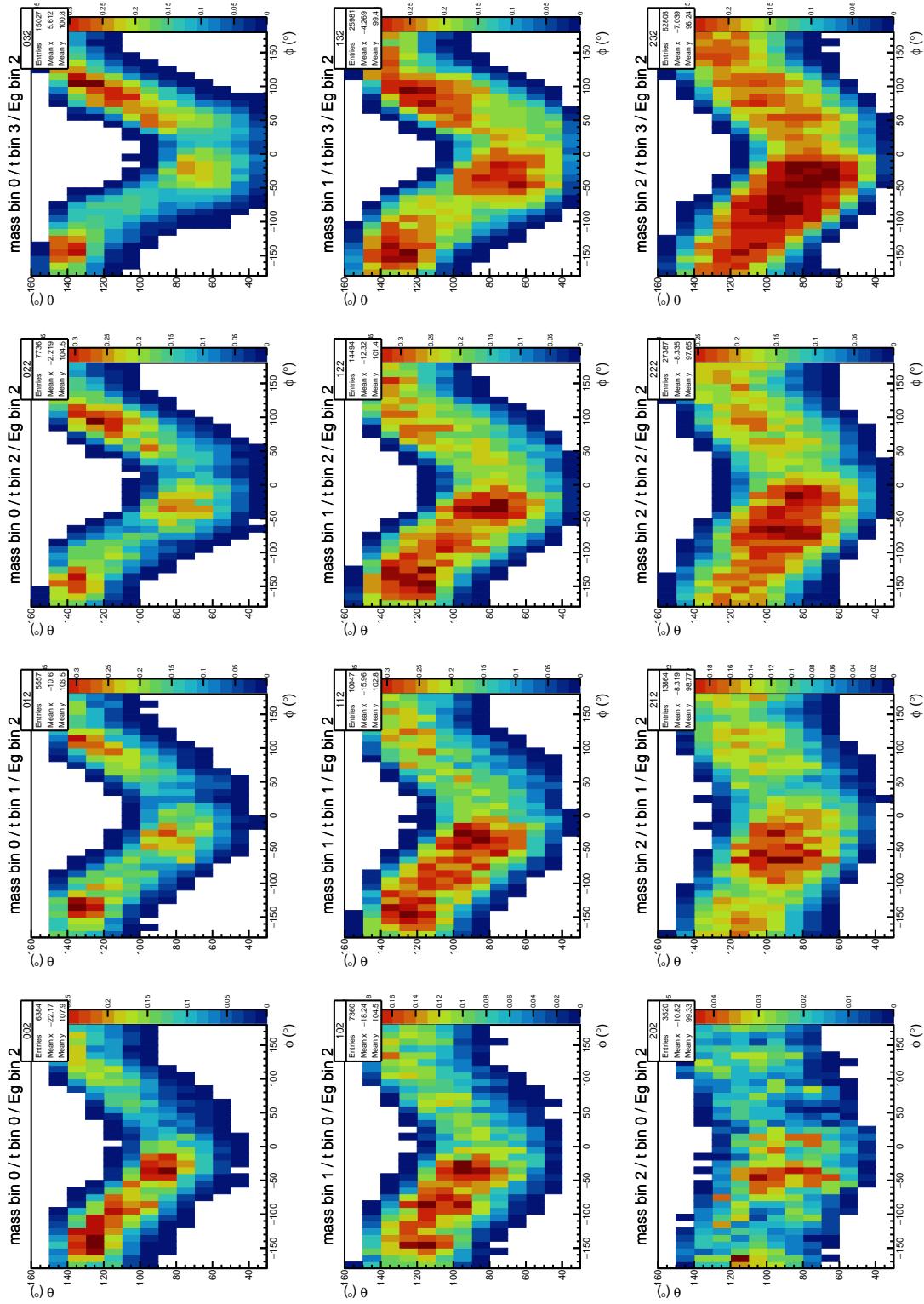


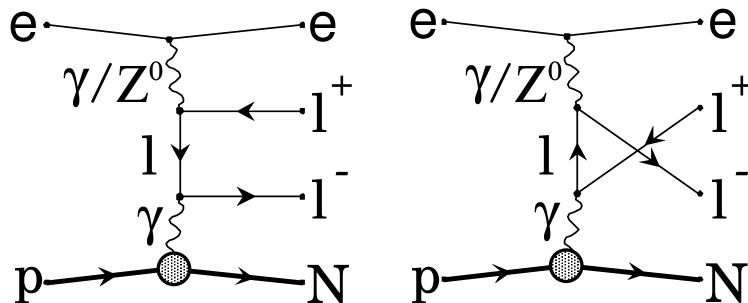
Figure 3.11: CLAS12 acceptance for TCS in the ϕ/θ plane for the third bin in E_γ .

3.5 Background estimations

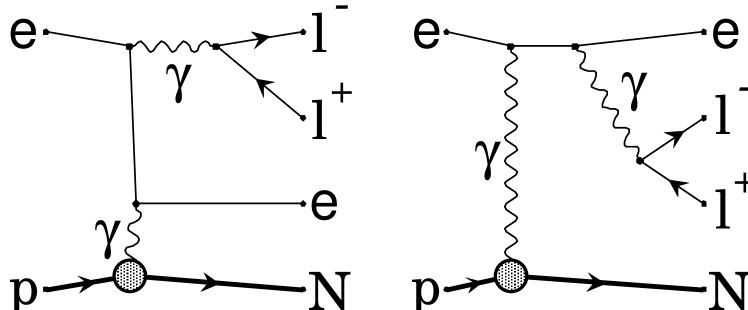
3.5.1 Electro-production of a lepton pair $ep \rightarrow pe^+(e^-)e_{scattered}$

From simulations

In Section 3.2, where the exclusivity cuts are presented, the scattered electron is assumed to stay undetected at low angles. However it could be deflected at high angle and the electron from the pair could stay undetected at low angles, mimicking the TCS reaction. Also, because the final state of the reaction has effectively two electrons, interferences between them have to be investigated. These effects are included in the *Grape* generator which allows to investigate the effects of using quasi-real photons for the TCS measurement. Two samples of events were generated. The first sample contains events in which two electrons, the scattered one and the one from the pair, are generated. All possible pair-production channels are added: the BH channel (top plots in Figure 3.12), the pair production from radiated photons (bottom plots in Figure 3.12), as well as the interference between the two electrons involved in the reaction. In the second sample the primary electron is generated at very low angles and only the BH process is considered.



(a) Bethe-Heitler type diagrams



(b) QED-Compton type diagrams

Figure 3.12: Diagrams of the processes included in the *GRAPE* event generator. The top two plots (a) correspond to the BH processes associated with TCS where the incoming photon is radiated from an electron. The bottom two plots (b) represent the processes where a radiated photon emits a photon that then decays in a lepton pair.

In the first sample including all possible effects, the two electrons can be distinguished by their transverse momentum. The electron with a high transverse momentum is likely to be detected and identified as a "pair" electron, while the low transverse momentum electron is likely to stay undetected. Following this observation, two configurations for the measurement are possible (a positron and a low or high transverse momentum electron). Depending on which electron is detected, the reconstructed invariant mass of the lepton pair, as well as the calculated photon kinematics, are different. Figure 3.13a shows the distribution of generated events in the plane defined by the two possible invariant masses, Q'^2 calculated from the high transverse momentum electron and Q''_2 from the low transverse

momentum electron. All the mass configurations that could lead to a reconstructed invariant mass in the range $2 \text{ GeV}^2 - 9 \text{ GeV}^2$ are generated. Finally, Figure 3.13b shows the distribution of the photon energy calculated assuming that the high transverse momentum electron is the one from the lepton pair. The blue histogram corresponds to the BH-only case, the red one to the full simulation. Both simulations are normalized to have equal integrals between 2 GeV and 7 GeV.

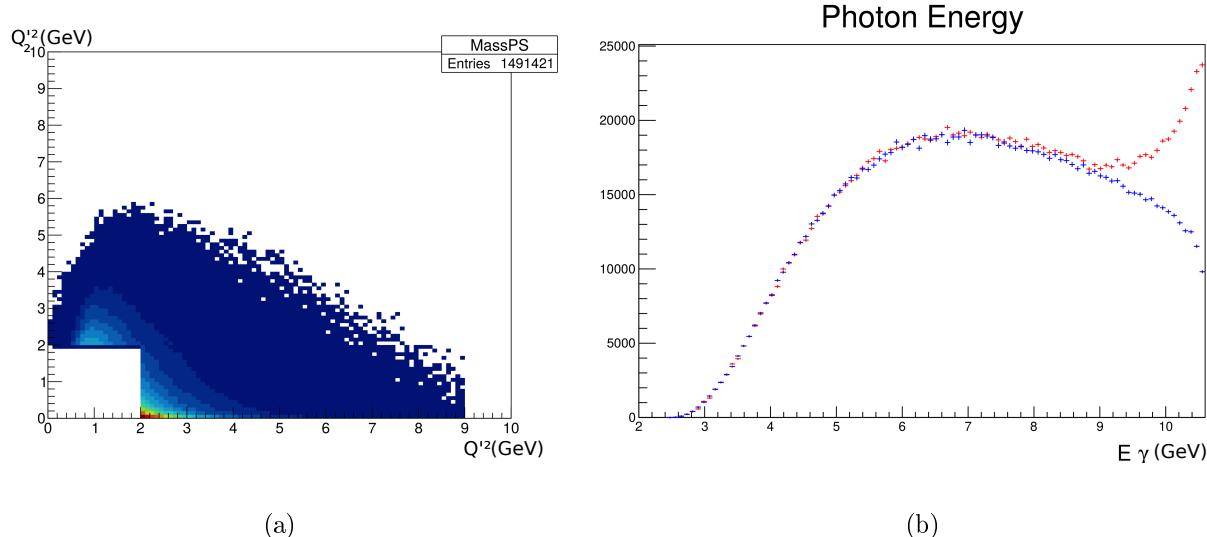


Figure 3.13: (a) Generated invariant mass phase space in the case where all possible processes leading to the final state $pe^+(e^-)e_{\text{scattered}}$ are considered. (b) Generated photon energy distributions for BH-only events (blue) and including all the possible processes (red).

The two generated samples are then passed through the CLAS12 simulation chain and exclusivity cuts are applied. The resulting measured distributions are shown in Figure 3.14. No large effect coming from the use of quasi-real photo-production is noticed.

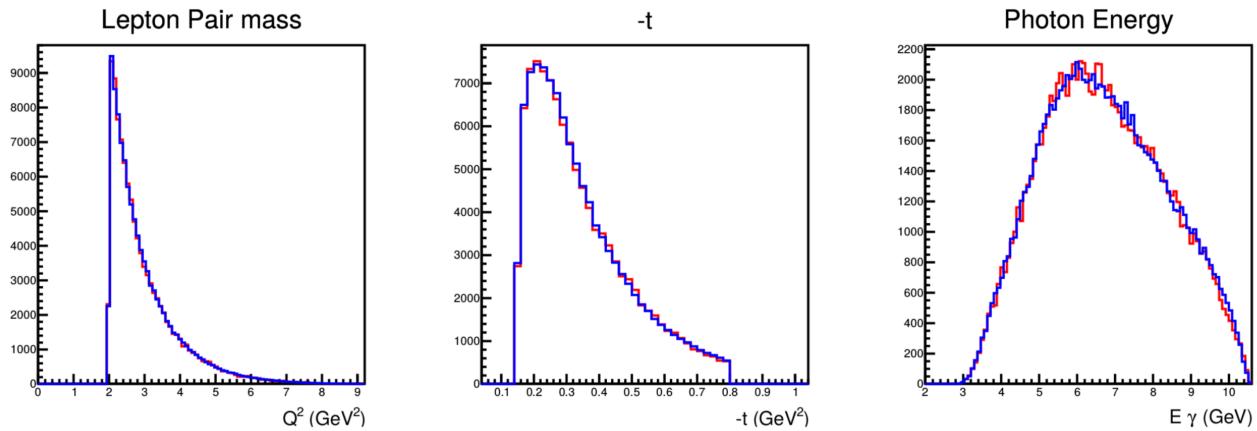
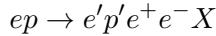


Figure 3.14: Measured invariant mass Q^2 (left), proton transferred momentum $-t$ (middle) and photon energy E_γ (right) distributions, for the generated BH-only events (blue) and with all the processes included (red).

996 From data

Fraction of events with scattered electron Events in electroproduction of (e^+e^-) pair contain two electrons in the final state. Final states with lepton pairs can come not only from a time-like photon or a vector meson decay but also bremsstrahlung photon conversions in the target or the detector material, from Dalitz decays, and of course, there could be also accidentals. The reaction

1001 below has been used to study Time-like Compton Scattering with kinematic cuts to isolate events from
 1002 the so-called quasi-real photoproduction.



1003 Here the scattered electron denoted as e' , the electron from the lepton pair is e^- . It is possible, even
 1004 with cuts to select a quasi-real photoproduction of lepton pairs when the missing particle is identified
 1005 as an electron, $M_X^2 \sim 0$ GeV 2 , with momentum constrained to $p_\perp \sim 0$, that the detected electron is
 1006 e' . To estimate the fraction of such events, one can study the final state with two electrons, $(e^- e^-)$,
 1007 in the same kinematics. Physics processes that will produce $(e' e^+)$ and the kinematics of such events
 1008 are the same as for $(e^- e^-)$ production. The difference in detecting these pairs is the CLAS12 efficiency
 1009 (acceptance) for *in-bending* and *out-bending* tracks in the torus field of the forward detector. Using
 1010 data from the "inbending" and "outbending" runs, one can estimate the efficiency differences and the
 1011 fraction of $e' e^+$ events in the $e^- e^+$ final state.

1012 **Analysis** Several runs from both polarity of the torus setting, "inbending" (runs 5361, 5366, 5367,
 1013 and 5368) and "outbending" (runs 5430 and 5499), have been used to study background from final
 1014 states with detected scattered electron and a positron from secondary processes (photon conversion,
 1015 Dalitz decay, accidentals). Recon DSTs from RG-A Fall2018 Pass1 processed data were converted to
 1016 *hbook ntuples*. Events were selected using the exclusivity cuts in the reaction:



1017 Final states were identified using the event builder particle IDs with simple cuts developed for TCS
 1018 and J/ψ analysis:

- 1019 • electron: in forward detector, PID=11, $p > 1$ GeV/c, $|\chi^2 PID| < 3$, $SF_{IN} > 0.2 - SF_{PCAL}$
- 1020 • positron: in forward detector, PID=-11, $p > 1$ GeV/c, $|\chi^2 PID| < 3$, $SF_{IN} > 0.2 - SF_{PCAL}$
- 1021 • proton: in forward or central detector, PID=2212, $p_{FD} > 0.3$ GeV/c, $|\chi^2 PID| < 3$

1022 Then for both, $e^+ e^- p$ and $e^- e^- p$ final states, the missing mass and the missing transverse momen-
 1023 tum cuts are applied to select "quasi-real" photoproduction events:

- 1024 • $M_X^2 < 0.4$ GeV 2 (missing mass squared)
- 1025 • $|p_X^x/p_X| < 0.05$ (fraction of the x-component of the missing 3-momentum)
- 1026 • $|p_X^y/p_X| < 0.05$ (fraction of the y-component of the missing 3-momentum)

1027 In Figures 3.15 and 3.16, the invariant mass distributions of lepton pairs for the "inbending" and
 1028 "outbending" data sets are shown, respectively. In each figure, top graphs are the invariant mass
 1029 distributions of $e^+ e^-$ for $e^+ e^- p$ final state. The lower graphs are the invariant mass distributions of
 1030 two electrons in $e^- e^- p$ final state. On the graphs, the total number events in each case and the number
 1031 of events with $M(ee) > 1.5$ GeV are shown as well.

1032 Lets define the efficiency of the CLAS12 forward detector for the *in-bending* lepton as a and for the
 1033 *out-bending* as b . It is reasonable to assume that the detection of $e^+ e^-$ pairs has the same efficiency
 1034 for the "inbending" and "outbending" settings, independent of the detected electron is the scattered
 1035 one or from the decay. Then on can write the following relations:

$$(In) \frac{N^{in}(e'e^-)}{N(e'e^+) + N(e^-e^+)} = \frac{\sigma(e'e^+) \cdot a \cdot a}{N(e'e^+) + N(e^-e^+)} = \frac{48}{1598} = 0.03$$

$$(Out) \frac{N^{out}(e^-e^-)}{N(e'e^+) + N(e^-e^+)} = \frac{\sigma(e'e^+) \cdot b \cdot b}{N(e'e^+) + N(e^-e^+)} = \frac{424}{821} = 0.516$$

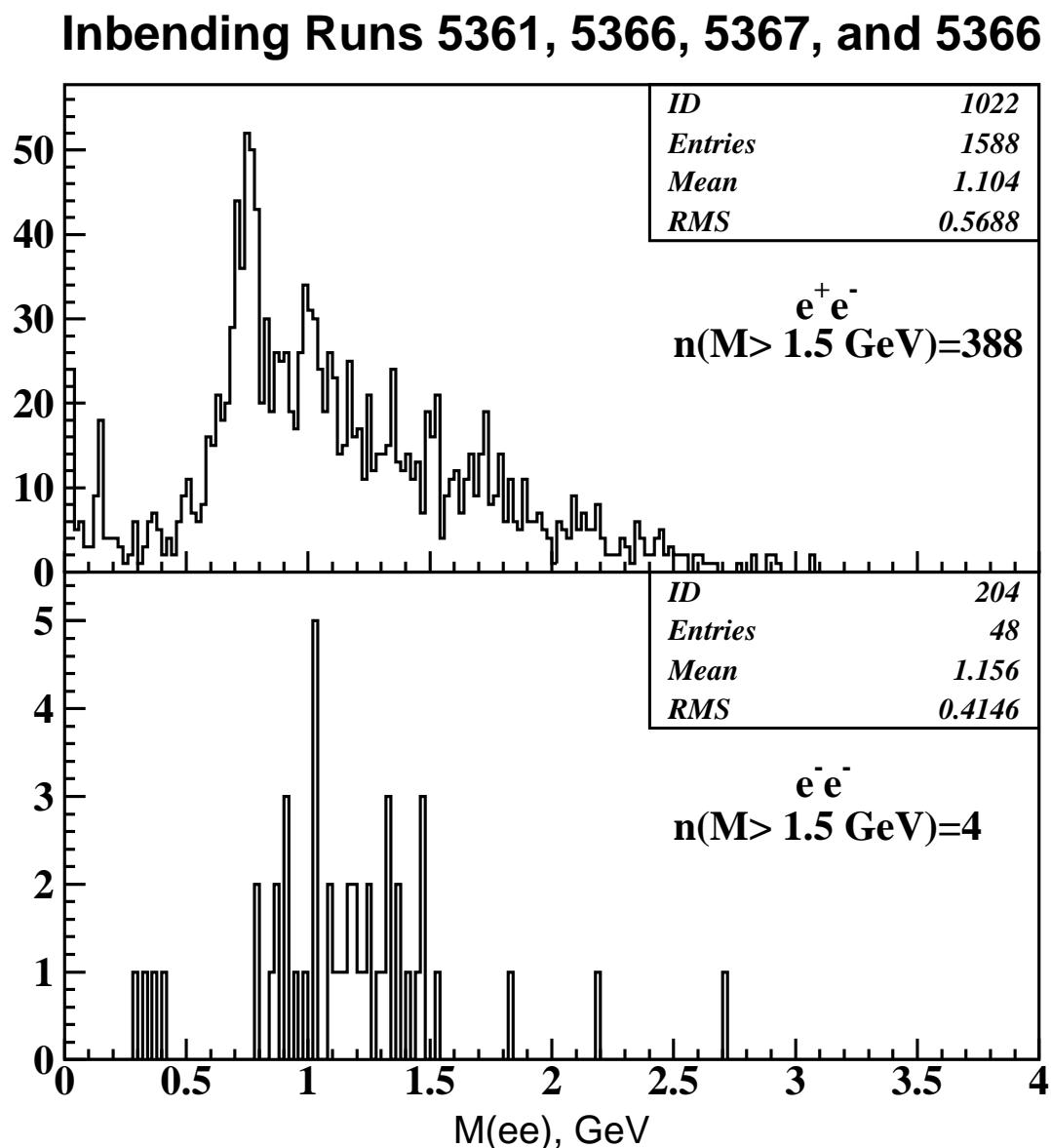


Figure 3.15: The invariant mass distributions for (e^+e^-) , top, and (e^-e^-) , bottom, pairs from "in-bending" data. Events in both samples were selected with the same conditions, final state eep , and exclusivity cuts.

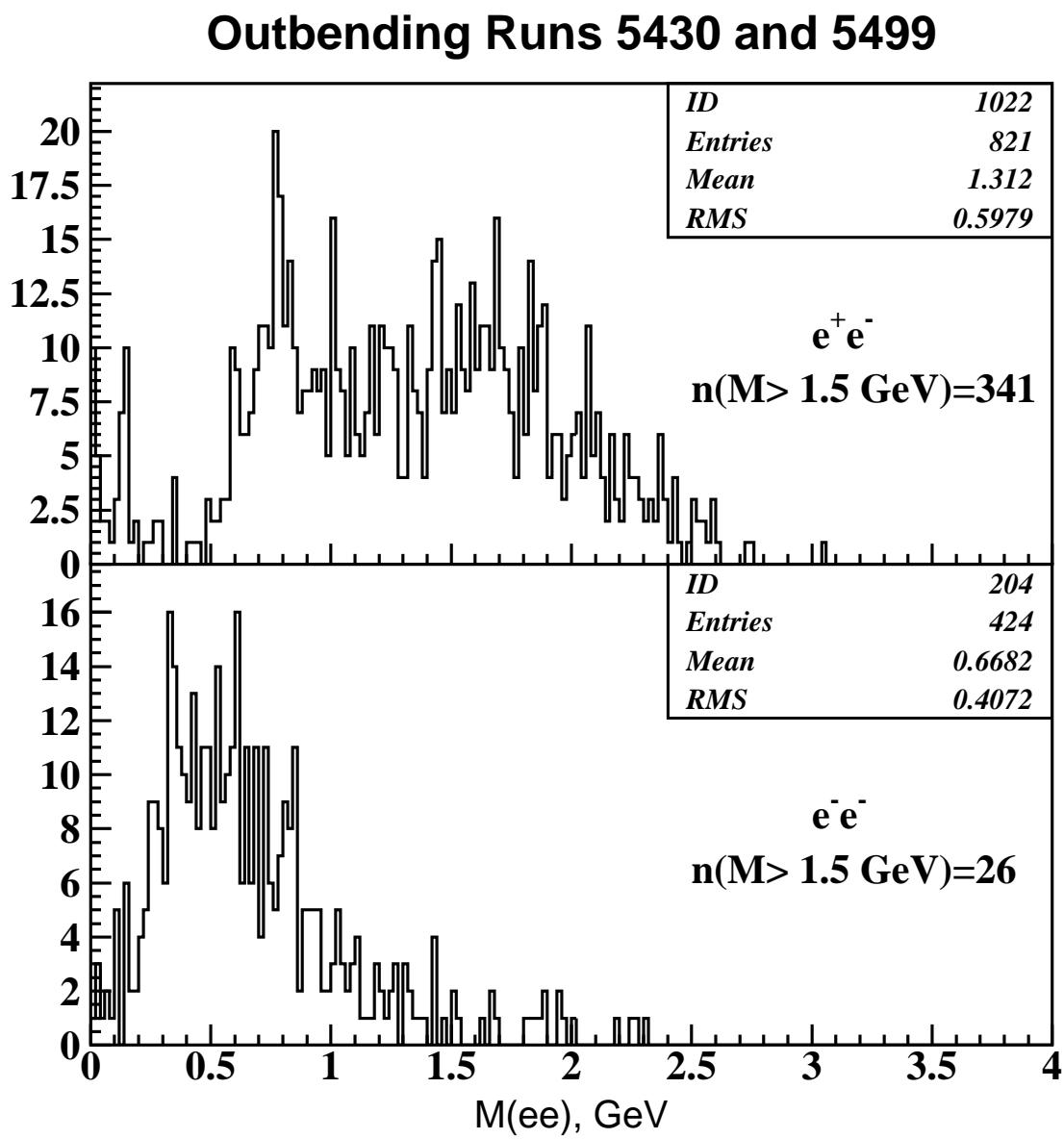


Figure 3.16: The invariant mass distributions for (e^+e^-) , top, and (e^-e^-) , bottom, pairs from "outbending" data. Events in both samples were selected with the same conditions, final state eep , and exclusivity cuts.

1036 Then the super ratio will be -

$$\frac{In}{Out} = \frac{a^2}{b^2} = 0.058$$
$$\frac{a}{b} = 0.24$$

1037 Now if we look the number events in each case with the invariant mass of letons $M_{ee} > 1.5$ GeV, and
1038 correct for the differences of detection efficiency for *in-bending* and *out-bending* tracks, the fraction of
1039 e^+e^- where $e^- \equiv e'$ in "inbending" sample is $1.7\% \pm 0.5\%$, for the "outbending" data it is $1.8\% \pm 0.3\%$.
1040 Within the measurements statistical errors two results are consistent, and we safely can conclude that
1041 the fraction of such background is $< 2\%$.

1042 3.5.2 Pion contamination

1043 Pions can be mistakenly identified as leptons, especially at high momenta (above ~ 4.5 GeV) where
1044 the HTCC produces signals for leptons and pions. The pion background is addressed in Chapter 2.
1045 The neural network PID is implemented in the TCS analysis with a cut at 0.5. We show in Chapter
1046 2 that this method reduces the background ratio by a factor 10, to less than 5%. To account for
1047 the remaining background and to estimate its impact on the measured observables, the background
1048 rejection cut is varied. This latter point is addressed in the Section 3.12.

1049 3.5.3 J/ψ contamination

1050 Another potential background of this analysis is J/ψ photo production events, where the invariant
1051 mass of the J/ψ is falsely reconstructed below 3 GeV due to radiative effects. To assess this background,
1052 Monte-Carlo simulations are used. JThese simulations were done by Joseph Newton and show that 8%
1053 of the total number of generated J/ψ are reconstructed in the radiative tail below 3 GeV. With the
1054 exclusivity cuts of this analysis, there are 82 events above 3GeV (all coming from Jpsi). Thus there
1055 are potentially $82 \times 8/92 = 7J/\psi$ in the $1.5 \text{ GeV} < M < 3 \text{ GeV}$ region (0.2% of the total number of
1056 events used in the analysis). This is much lower than the potential contamination from π^+ and can be
1057 ignored.

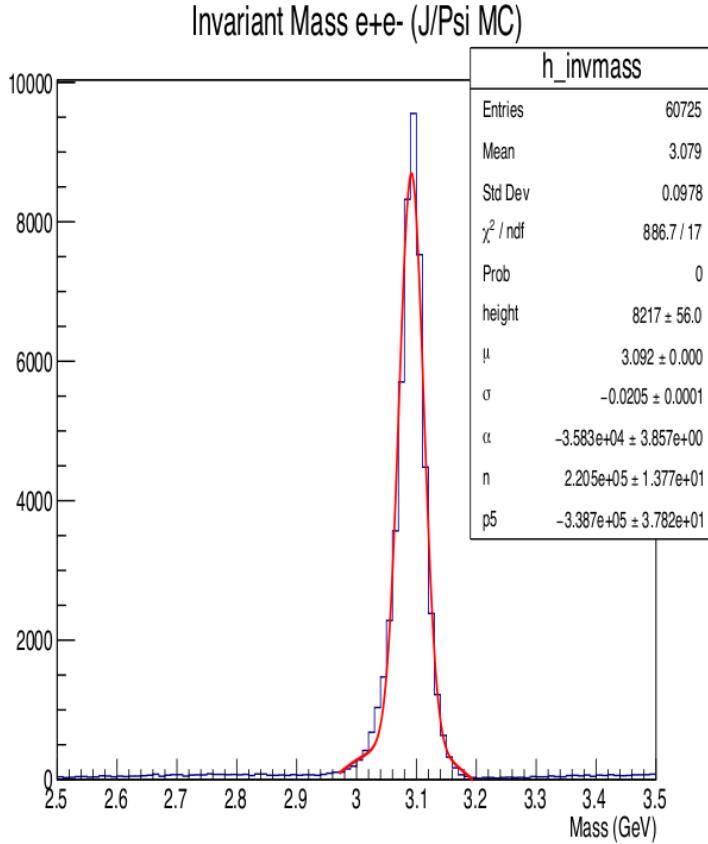


Figure 3.17: J/ψ simulation (done by J.Newton). 8% of the initially generated events are reconstructed below 3 GeV.

1058 3.6 Experimental cross-section ratio

1059 The theoretical R ratio is calculated over the full angular phase space of TCS. However CLAS12
 1060 acceptance is limited and only covers a fraction of the angular phase space. The theoretical R ratio
 1061 is thus inaccessible, but it can be calculated over the CLAS12 acceptance. Following the notations in
 1062 [3], we call R' the ratio integrated over the CLAS12 acceptance. It is calculated as:

$$R' = \frac{\sum_{\phi} Y_{\phi} \cos(\phi)}{\sum_{\phi} Y_{\phi}}, \quad (3.20)$$

1063 where the sum over ϕ is done in 10° bins and the $\cos(\phi)$ factor is calculated at the center of each bin.
 1064 The Y_{ϕ} quantity is calculated for each ϕ bin as:

$$Y_{\phi} = \sum_{\text{events in } \phi \text{ bin}} \left(\frac{L}{L_0} \right) \left(\frac{1}{Acc} \right), \quad (3.21)$$

1065 where the ratio $\frac{L}{L_0}$ is calculated event-by-event (the factors L and L_0 are given in [1]), respectively),
 1066 and Acc is the acceptance in the kinematic bin of each event (given in Equation (3.14)). The statistical
 1067 error of Y_{ϕ} is calculated as:

$$E^2(Y_{\phi}) = \sum_{\text{events in } \phi \text{ bin}} \left(\frac{L}{L_0} \right)^2 \left(\frac{1}{Acc} \right)^2, \quad (3.22)$$

1068 Statistical uncertainty using Monte Carlo method

1069 As the numerator and the denominator used in the computation of R' are correlated, a Monte Carlo
 1070 approach is used to compute the statistical uncertainty. It works as follows:

- Y_ϕ and $E(Y_\phi)$ are computed as described above,
- for each ϕ bin, a value Y_ϕ^R is randomly generated following a gaussian probability distribution of mean Y_ϕ and sigma $E(Y_\phi)$,
- the ratio is calculated using the randomly generated values: $R^R = \frac{\sum_\phi Y_\phi^R \cos(\phi)}{\sum_\phi Y_\phi^R}$.
- The previous two steps are repeated 10000 times and the resulting distribution of R^R is fitted with a gaussian,
- the σ of the fitted gaussian is defined as the statistical uncertainty of the measurement. An example of the fitted distribution is shown in Figure 3.18.

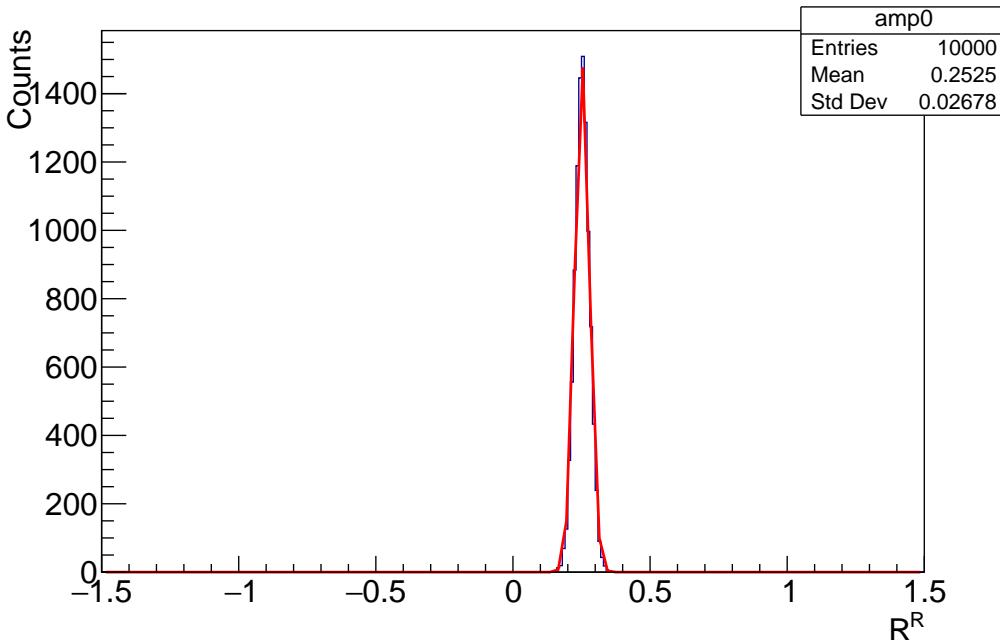


Figure 3.18: Distribution of R^R for the determination of the statistical error for first bin in ξ . The full results are shown in Figure 4.5. The gaussian fit is overlaid.

3.7 Phenomenological study of the TCS Forward-Backward asymmetry

No predictions have yet been published for the Forward-Backward asymmetry (A_{FB}) in TCS. In this section the phenomenological work performed on the TCS A_{FB} is presented. The main goal of this analysis is to estimate the size of the asymmetry as well as to gain an insight on its kinematic dependencies in order to determine if CLAS12 can measure it.

This section is divided in three parts. In Subsection 3.7.1, the limitations of CLAS12 acceptance and the consequences on the measurement of the A_{FB} are presented. Initial results are compared to other predictions provided by independent groups and models. The results obtained according to the conclusion of the first section are then displayed in subsequent Subsections 3.7.2 and 3.7.3.

3.7.1 Early considerations and comparison with other models

To study the TCS A_{FB} , the TCS and BH processes and their interference cross sections have been calculated using the VGG code [13] provided by Michel Guidal. The formulas used in this code are described in [2]. The cross section is estimated in the forward and backward directions and the obtained

1093 results are then combined to produce the asymmetry curves shown in the following. The GPD model
 1094 used is the VGG model described in [14]. The D-term contribution to the GPD H is included (unless
 1095 mentioned otherwise). The α' coefficient fixing the t -dependence of the GPDs is set to 1.098, the b_{val}
 1096 and b_{sea} parameters fixing the ξ dependence are set to 1 (unless specified otherwise).

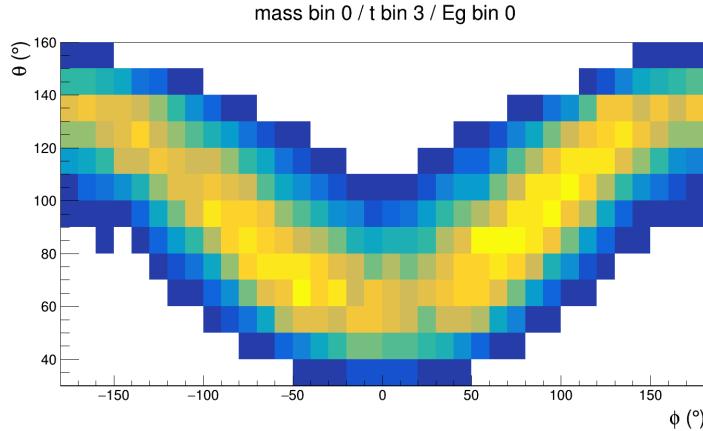


Figure 3.19: Acceptance of CLAS12 in the θ/ϕ plane. The region around $\phi = 0^\circ$ and low polar angle, as well as $\phi = \pm 180^\circ$ and high polar angle are well covered by CLAS12.

1097 The CLAS12 acceptance does not cover the full $\phi - \theta$ plane of TCS. An example of the acceptance
 1098 coverage, obtained from the CLAS12 acceptance study presented in Section 3.4, is shown in Figure
 1099 3.19. The regions around $\phi = 0^\circ$ and low θ , and around $\phi = \pm 180^\circ$ and high θ are well covered by
 1100 CLAS12. Following this observation, it is decided to study the FB asymmetry at $\phi_0 = 0^\circ$.

1101 A prediction for the asymmetry is shown in Figure 3.20a, where the t -dependence of the asymmetry
 1102 is plotted for various values of θ_0 . These results were cross checked independently by M. Vanderhaeghen
 1103 also using the VGG code. Figure 3.20b shows the results of this independent analysis. Both results
 1104 are pointing toward an asymmetry with a positive value. Note that these results were obtained after
 1105 correcting a missing minus sign in Equation (17) of [2]. The consequence of adding this minus sign is
 1106 to flip the sign of the asymmetry.

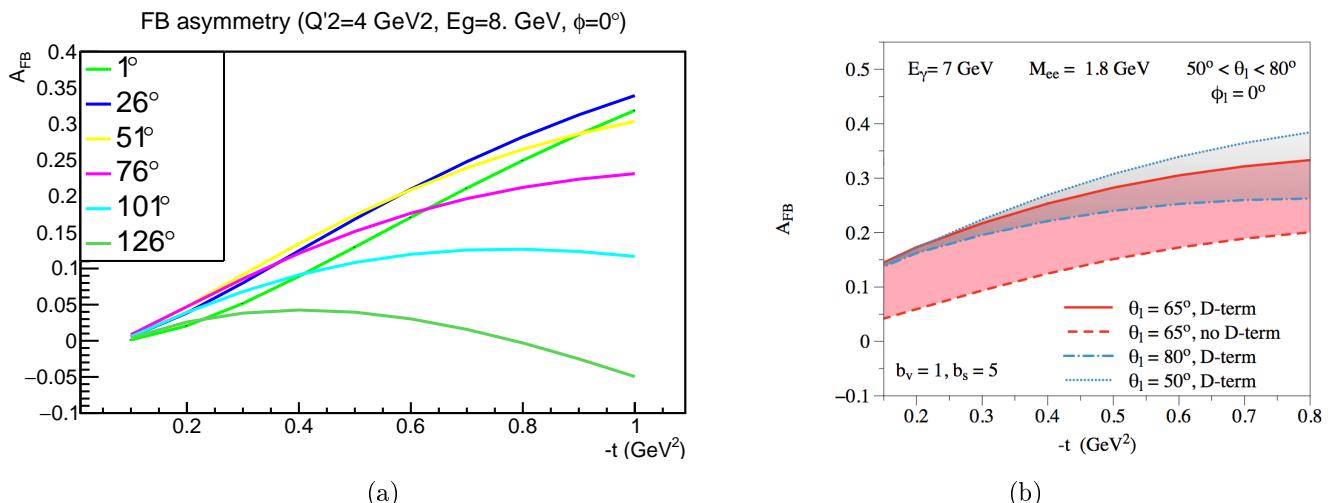


Figure 3.20: (a) FB asymmetry as a function of $-t$ at fixed $\phi_0 = 0^\circ$ for different values of θ_0 . (b) Asymmetry as a function of $-t$ at fixed $\phi_0 = 0^\circ$ for different values of θ_0 and different models: the impact of the D-term on the asymmetry is shown. Figure courtesy of M. Vanderhaeghen.

1107 Later predictions realized by Paweł Sznajder [15], using the *PARTONS* software [16], are shown in
 1108 Figure 3.21. These predictions also point to a positive asymmetry, independently of the model used

1109 for the CFFs as well as high order corrections.

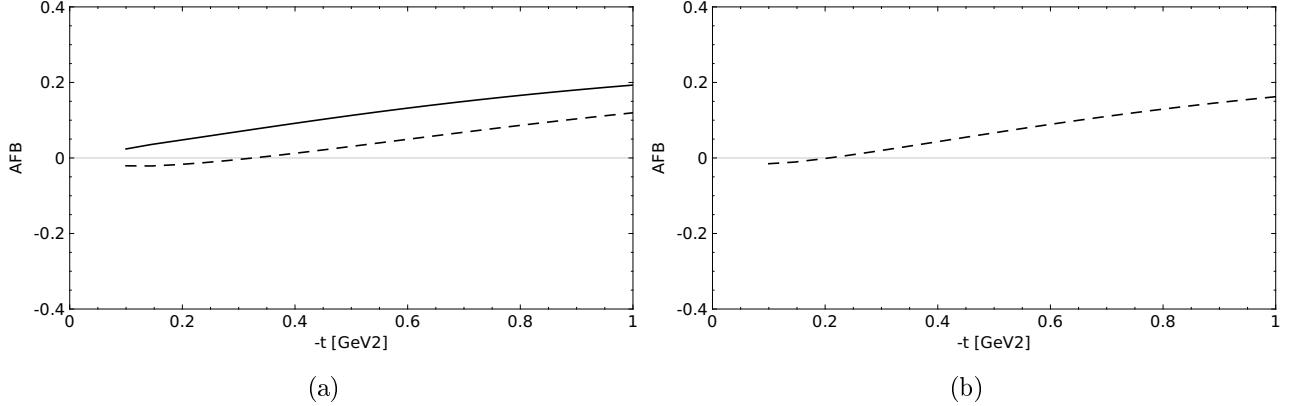


Figure 3.21: FB asymmetry as a function of $-t$ at fixed $\phi_0 = 0^\circ$, $\theta_0 = 60^\circ$, $E_\gamma = 7$ GeV and $Q'^2 = 1.8$ GeV 2 (a) using the GK model at LO (dashed) and NLO (plain), (b) using the VGG model at LO. Figures courtesy of Paweł Sznajder.

1100 Furthermore, the analytical formulae for TCS derived in [1] were used to cross-check the asymmetry
1101 in VGG. This was done after making sure the CFF conventions are consistent. The results obtained
1102 for the analytical formulae using VGG CFFs are shown in Figure 3.22. The sign of the asymmetry
1103 obtained is consistent with the *PARTONS* results and with the corrected VGG asymmetries.

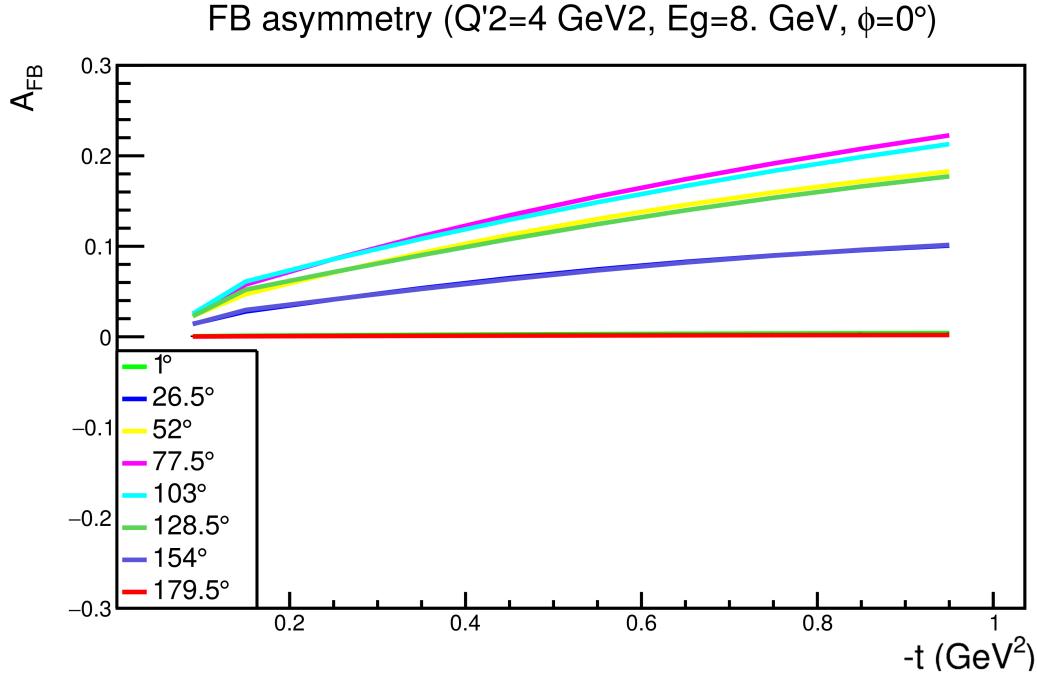


Figure 3.22: FB asymmetry as a function of $-t$ at fixed $\phi_0 = 0^\circ$ for different values of θ_0 using formulas derived in [1].

1114 A final consistency check was performed by calculating the asymmetry from the TCS only and BH
1115 only cross sections. The absence of asymmetry in both cases is well verified.

1116 3.7.2 TCS A_{FB} kinematic dependencies

1117 A main feature of the FB asymmetry is that it can be measured over a small portion of the TCS
1118 angular phase space. This prevents any large detector-induced false asymmetry caused by holes in
1119 the acceptance, which instead occurs in the case of the R ratio (see Section 3.6). On the contrary,

the statistics dramatically falls if the angular phase space of the measurement is too narrow. The main goal of this analysis is to determine the angular range that can be used for the measurement in CLAS12, in order to maximize statistics. One has to make sure that the FB asymmetry conserves its sign over the angular integration domain to maximize the size of the measured asymmetry. Therefore, the angular dependence of the FB asymmetry is studied. From the explicit expression of the A_{FB} derived following [1]:

$$A_{FB}(\theta_0, \phi_0) = \frac{-\frac{\alpha_{em}^3}{4\pi s^2} \frac{1}{-t} \frac{m_p}{Q'} \frac{1}{\tau\sqrt{1-\tau}} \frac{L_0}{L} \cos \phi_0 \frac{(1+\cos^2 \theta_0)}{\sin(\theta_0)} \text{Re}\tilde{M}^{--}}{d\sigma_{BH}}, \quad (3.23)$$

one can see that the ϕ_0 dependence is driven by the $\cos(\phi_0)$ factor. The ϕ_0 behavior is shown in Figure 3.23 for fixed $\theta = 80^\circ$. The asymmetry changes sign at $\phi_0 \approx \pm 90^\circ$. This behavior is reproduced by both the VGG model and the analytical model. The differences are attributed to the terms that are neglected in the analytical model (see [1] for full details). Following this investigation, one can see that the FB asymmetry in the region around $\phi_0 = 0^\circ$ is maximal and that the sign is constant in its vicinity. In Figure 3.20a, showing the asymmetry as a function of $-t$ at $\phi_0 = 0^\circ$ and for different values of θ_0 , the asymmetry has constant sign over the θ_0 range accessible in the CLAS12 acceptance (from $\sim 40^\circ$ to $\sim 100^\circ$).

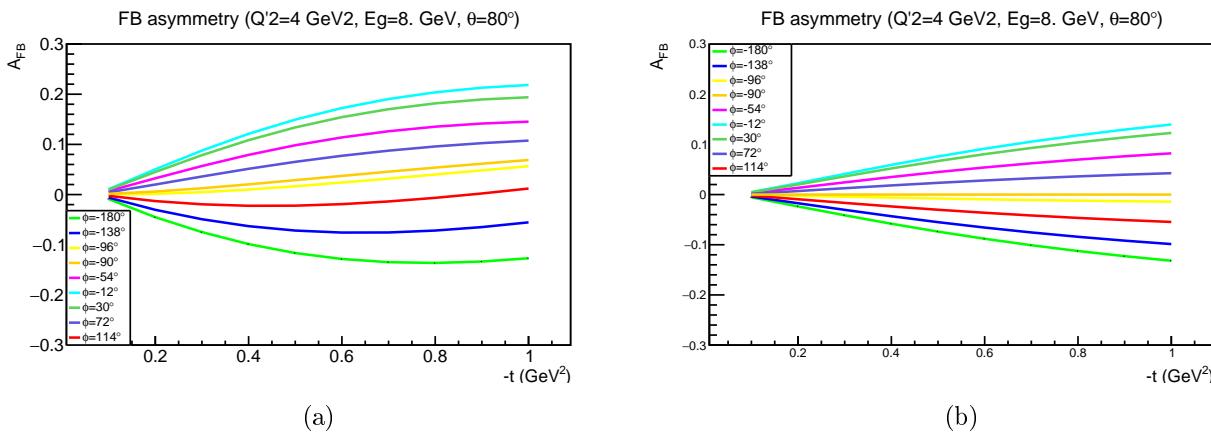


Figure 3.23: FB asymmetry as a function of $-t$ at fixed $\theta_0 = 80^\circ$ for different values of ϕ_0 computed (a) with the VGG code, (b) using the formulas derived in [1].

Finally, the asymmetry was studied as a function of the incoming photon energy E_γ and the square of the invariant mass of the lepton pair Q'^2 , aiming at increasing the statistics available for the measurement. Figure 3.24 shows the Q'^2 and E_γ dependencies of the FB asymmetry at $\phi_0 = 0^\circ$ and for various values of θ_0 . These plots indicate that the sign of the asymmetry is constant over the kinematic range accessible by CLAS12. This implies that it is possible to integrate the measurement over a large portion of the phase space and still measure an asymmetry.

3.7.3 TCS A_{FB} model dependencies

After exploring the kinematic dependencies of the FB asymmetry, its dependencies on GPD model parameters are tested.

D-term dependence

The first model dependence investigated is the D-term. The FB asymmetry is sensitive to the real part of the CFFs and thus to the D-term. This makes this observable a good candidate to investigate its strength. The FB asymmetry is calculated with and without the D-term contribution. The D-term used is the one presented in [17]. The effect of adding the D-term to the GPDs parametrization is shown in Figure 3.25. One can see a sizable effect on the asymmetry induced by the D-term. It is also important to notice that the amplitude of the asymmetry increases with the addition of the D-term, and that the sign of the asymmetry does not change when adding it.

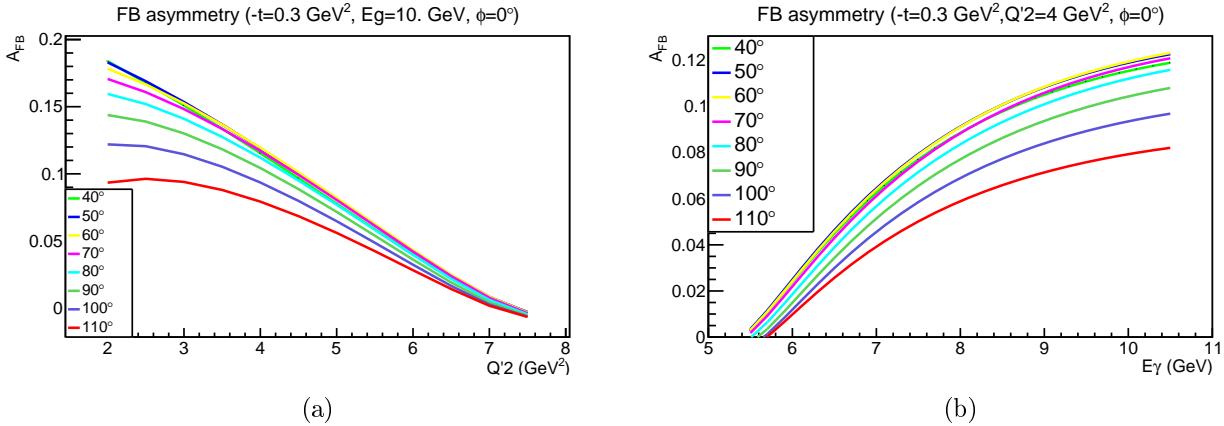


Figure 3.24: FB asymmetry calculated using the VGG code (a) as a function of Q'^2 and (b) as a function of E_γ at fixed $\phi_0 = 0^\circ$ and for different values of θ_0 .

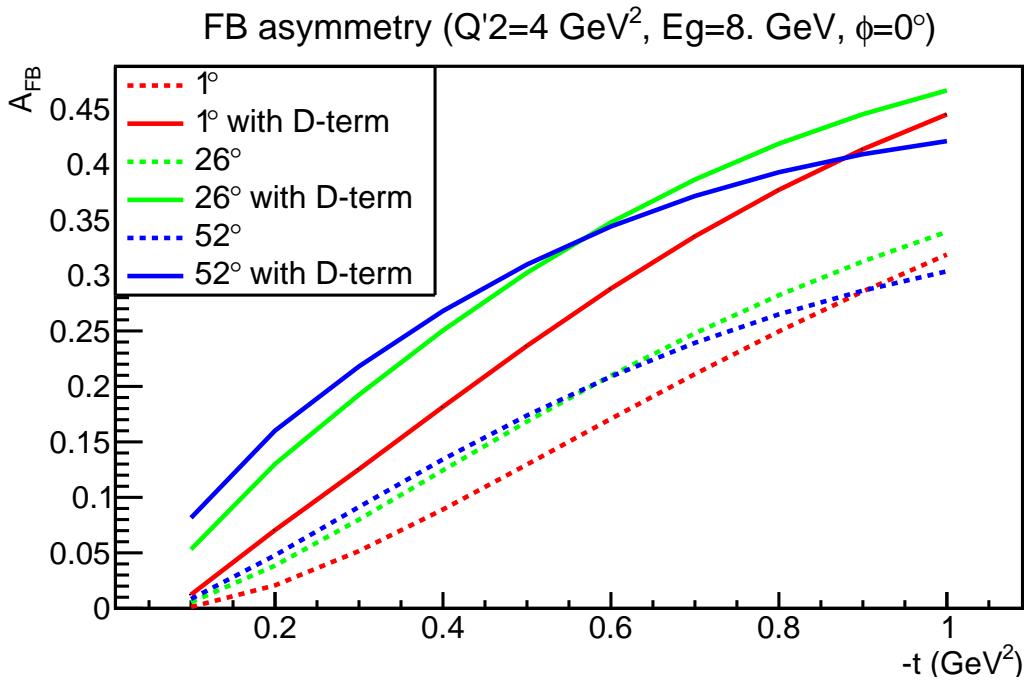


Figure 3.25: FB asymmetry as a function of $-t$ at $\phi_0 = 0^\circ$ for different values of θ ; and with (solid lines) or without (dashed lines) the D-term contributions to the GPD H.

1151 b_{sea} -dependence

1152 The dependence of the asymmetry to the skewness parameter b_{sea} is also explored. Indeed in [18] it is
 1153 suggested that DVCS data, sensitive to the imaginary part of the \mathcal{H} , are better described with skewness
 1154 parameters $b_{val} = 1$ and $b_{sea} = 5$. Following this observation, the value of the sea parameter is varied
 1155 in the asymmetry calculation, from 1 (strong skewness dependence) to 8 (low skewness dependence),
 1156 to verify if the FB asymmetry can help in the determination of this parameter. The results are shown
 1157 in Figure 3.26, where the asymmetry dependencies on $-t$ and Q'^2 are plotted for various values of the
 1158 sea parameter. No large effect is seen when varying the parameter b_{sea} .

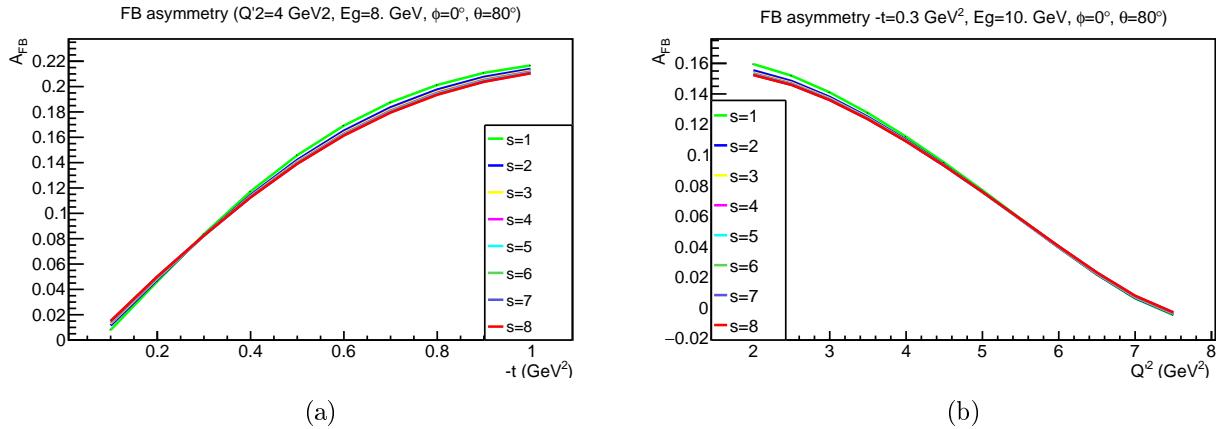


Figure 3.26: (a) $-t$ dependence of A_{FB} for different values of the b_{sea} parameter. (b) Q'^2 dependence of A_{FB} for different values of the b_{sea} parameter.

1159 3.8 Experimental Forward-Backward asymmetry

1160 According to the conclusions of the phenomenological study in Section 3.7, the FB asymmetry is
 1161 calculated at $\phi_0 = 0^\circ$. Furthermore, by looking at the angular coverage of CLAS12, it is decided to
 1162 integrate over the following angular bin in the forward direction:

1163 • $-40^\circ < \phi_0 < 40^\circ$

1164 • $50^\circ < \theta_0 < 80^\circ$.

1165 The corresponding backward bin limits are:

1166 • $140^\circ < 180^\circ + \phi_0 < 220^\circ$

1167 • $100^\circ < 180^\circ - \theta_0 < 130^\circ$.

1168 Bin volume correction

1169 Figure 3.27 shows the acceptance of CLAS12 in the θ - ϕ plane for a given $-t$, E_γ and Q'^2 bin. The
 1170 limits of the forward and backward bins are shown in green and red, respectively. Although the limits
 1171 of the angular bins have been chosen to be covered by the CLAS12 acceptance, some $\{E_\gamma, -t, Q'^2\}$
 1172 acceptance bins do not fully cover the forward or backward angular bins. The difference in coverage
 1173 between the forward and the backward direction can be the source of fake asymmetries. To correct
 1174 for this, a bin volume correction is applied during the calculation of the FB asymmetry. For each
 1175 acceptance bin $\{E_\gamma, -t, Q'^2\}$, the fraction of the angular bins covered by the acceptance, $CorrVol_{F/B}$,
 1176 is calculated. Each event detected in the forward (resp. backward) bin is then assigned a correction
 1177 weight equal to the inverse of the fraction of the volume covered by the acceptance in this bin. This
 1178 correction assumes that the cross section of the TCS reaction is relatively constant within the volume
 1179 of the forward (resp. backward) bin and that it can be estimated only by measuring it in the volume
 1180 covered by the acceptance of CLAS12. This hypothesis is verified by extracting the FB asymmetry with
 1181 BH-weighted simulation events and the difference between the expected value for BH (null asymmetry)
 1182 and the extracted value is assigned as a systematic uncertainty (see Section 3.12). Two sets of volume
 1183 correction coefficients are obtained, one for the forward and one for the backward angular bins.

1184 For a given bin, the value of the FB asymmetry is calculated as the ratio:

$$A_{FB} = \frac{N_F - N_B}{N_F + N_B} \quad (3.24)$$

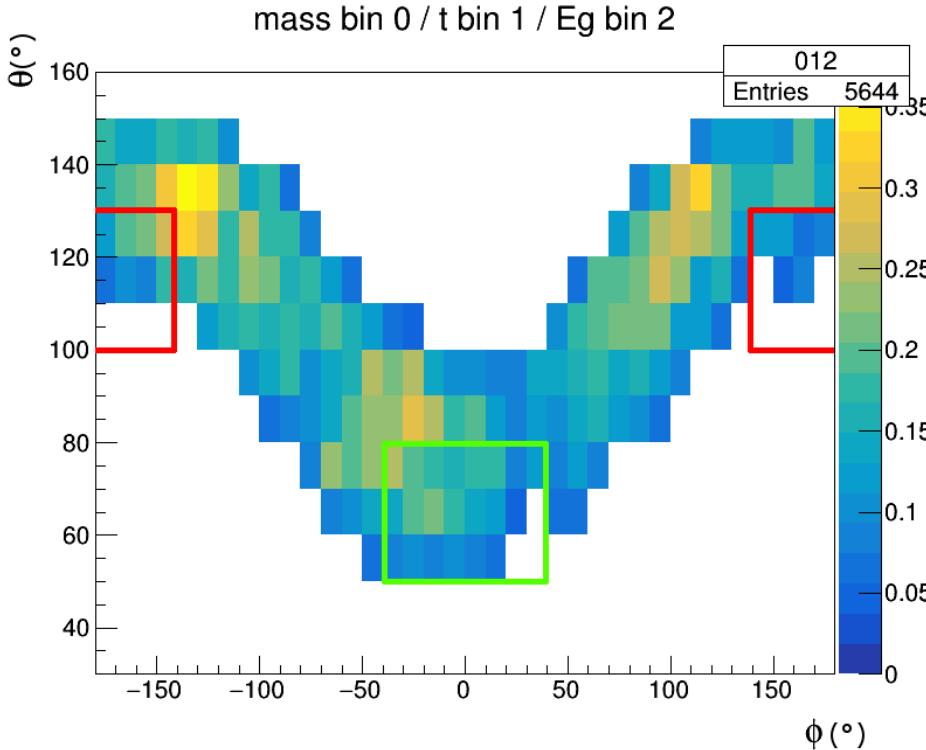


Figure 3.27: Volume correction for the A_{FB} calculation. The forward bin is represented by the green square. The red square delimits the backward bin. One can see that CLAS12 does not fully cover the angular bins for some $\{E_\gamma, -t, Q^2\}$ bins.

1185 where $N_{F/B}$ is the number of events in the forward (resp. backward) angular bin, corrected by the
1186 acceptance and the bin volume, as:

$$N_{F/B} = \sum \frac{1}{Acc \times CorrVol_{F/B}}. \quad (3.25)$$

1187 The reported statistical error bars are calculated by propagating the weighted error on $N_{F/B}$:

$$E^2(N_{F/B}) = \sum \left(\frac{1}{Acc \times CorrVol_{F/B}} \right)^2. \quad (3.26)$$

1188 3.9 Experimental beam-spin asymmetry

1189 As mentioned previously, this analysis is done on quasi-real photoproduction events, where a real
1190 photon is radiated by the initial electron beam. In this configuration, the circular polarization of the
1191 photon can be inferred from the initial longitudinal polarization of the electron beam. An electron
1192 polarized in the direction (resp. opposite) of the beam emits a right-(resp. left-) handed circularly
1193 polarized photon, with a polarization transfer Pol_{transf} , fully calculable analytically in QED (see
1194 [19, 20] for details and in Section 3.10 for the detailed formulas used to calculate the polarization
1195 transfer).

1196 The photon polarization asymmetry is calculated as a function of the azimuthal angle ϕ as:

$$BSA(-t, E\gamma, M; \phi) = \frac{1}{Pol_{eff}} \frac{N^+ - N^-}{N^+ + N^-}, \quad (3.27)$$

1197 where the number of events with reported positive N^+ (resp. negative N^-) electron helicity in each
1198 bin is corrected for the acceptance and the polarization transfer as:

$$N^\pm = \sum \frac{1}{Acc} Pol_{transf}. \quad (3.28)$$

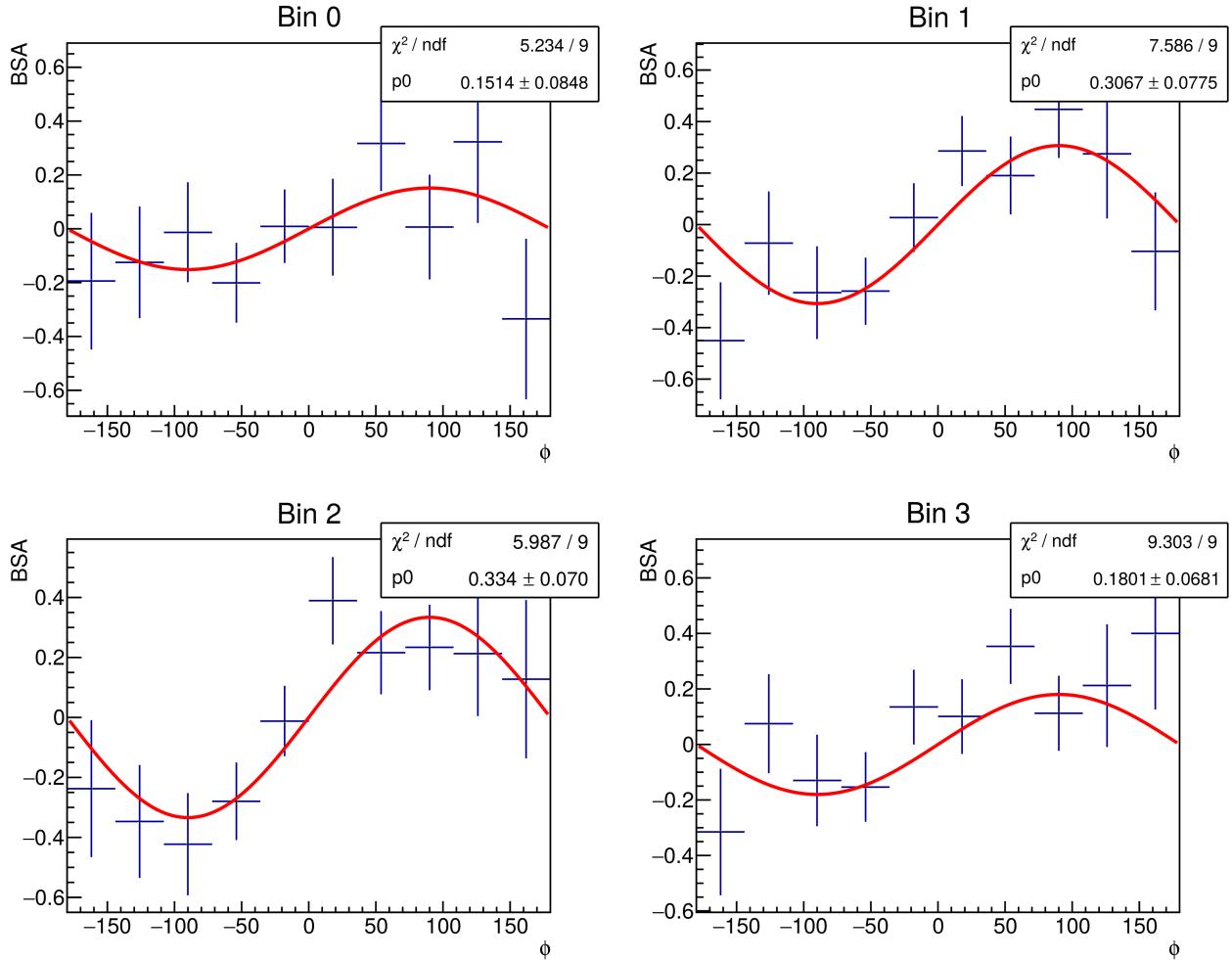


Figure 3.28: Distributions of the photon helicity asymmetry as a function of ϕ for four bins in $-t$ (as described in Section 3.11). The fit used to extract the asymmetry is overlaid. The fit amplitudes are displayed in Figure 4.14.

1199 P_{eff} is the effective polarization of the CEBAF electron beam, which is estimated to be 85% on
1200 average. The ϕ -distribution is then fitted with a sinus function:

$$BSA(-t, E\gamma, M; \phi) = BSA(-t, E\gamma, M) \sin(\phi), \quad (3.29)$$

1201 as shown in Figure 3.28. Following the definition of [2], the amplitude of the sinus function given by
1202 the fit (BSA at $\phi = \pi/2$) is extracted. The reported statistical error bars are given by the fit error on
1203 the amplitude of the function.

1204 3.10 Photon polarization transfer

1205 The circular polarization P of the incoming quasi-real photon in the $ep \rightarrow (e')\gamma p \rightarrow (e')e^-e^+p'$ reaction
1206 can be fully calculated in QED. It is given by the helicity S of the electron from the beam that emitted
1207 the photon multiplied by a polarization transfer coefficient L . If the electron has positive helicity
1208 the photon has right-handed polarization, while a negative helicity electron will emit a left-handed
1209 polarized photon. The following formulas are developed in [19], and the notation of [20] is used.

The photon circular polarization P is given as a function of the incoming electron beam helicity S as:

$$P = S L, \quad (3.30)$$

$$L = k [(E_1 + E_2)(3 + 2\Gamma) - 2E_2(1 + 4u^2\xi^2\Gamma)] / I_0, \quad (3.31)$$

$$I_0 = (E_1^2 + E_2^2)(3 + 2\Gamma) - 2E_1E_2(1 + 4u^2\xi^2\Gamma). \quad (3.32)$$

where L is the polarization transfer factor, E_1 and E_2 are respectively the energy of the incident and the scattered electrons and $k = (E_1 - E_2)$ is the energy of the photon. The Coulomb screening factor Γ is given by:

$$\Gamma = \mathcal{F}\left(\frac{\delta}{\xi}\right) - \ln(\delta) - 2 - f(Z), \quad (3.33)$$

where $\delta = k/2E_1E_2$. The factor ξ is calculated as $\xi = 1/(1 + u^2)$, with $u = E_1 \sin \theta_\gamma$, where θ_γ is the angle between the incoming electron and the radiated photon. The function \mathcal{F} accounts for the screening effects. Tabulated values for \mathcal{F} are given in Table 3.2. The screening function is also plotted in Figure 3.29. Finally the Coulomb correction function $f(Z)$ is given by:

$$f(a) = a^2 \sum_{n=1}^{\infty} \frac{1}{n(n^2 + a^2)}, \quad (3.34)$$

where $a = \alpha Z$, α is the electromagnetic coupling constant and Z is the atomic number of the target material where the radiation of the photon occurs.

Δ	0.5	1	2	4	8	15	20	25	30	35
$\mathcal{F}\left(\frac{\delta}{\xi}\right)$	0.0145	0.0490	0.14	0.3312	0.6758	1.126	1.367	1.564	1.731	1.875
Δ	40	45	50	60	70	80	90	100	120	
$\mathcal{F}\left(\frac{\delta}{\xi}\right)$	2.001	2.114	2.216	2.393	2.545	2.676	2.793	2.897	3.078	

Table 3.2: Tabulated values of the screening function \mathcal{F} as a function of $\Delta = (6Z^{-1/3}/121)(\xi/\delta)$.

The polarization transfer function L used for the TCS analysis is plotted as a function of the ratio between the energy of the photon and the energy of the beam in Figure 3.30.

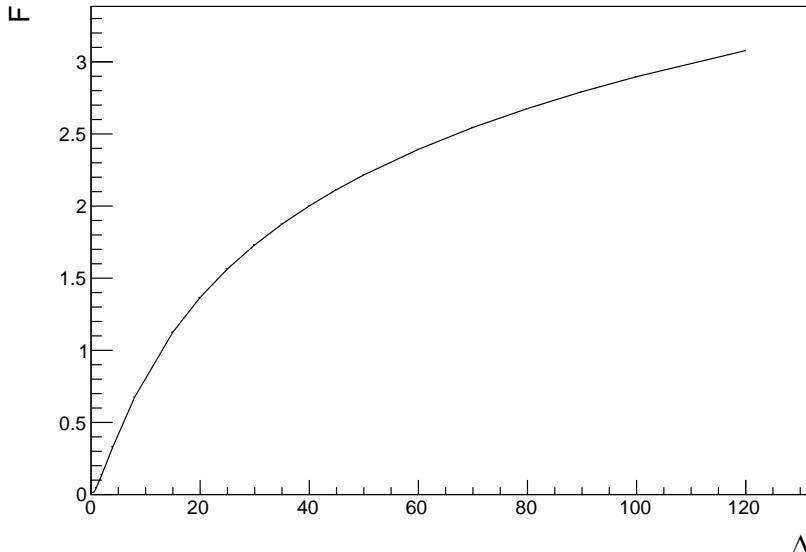


Figure 3.29: Screening function \mathcal{F} as a function of Δ .

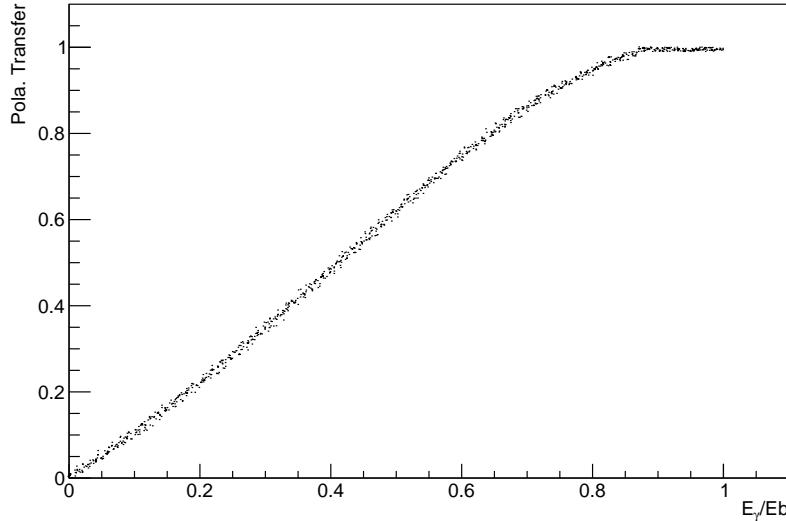


Figure 3.30: Polarization transfer function used for the TCS analysis plotted as a function of the ratio between the energy of the radiated photon and the energy of the electron beam. At high energy (for $E_\gamma/E_b > 0.86$), the calculated polarization transfer is bigger than one. The polarization transfer is set to one when this happens.

1221 3.11 Binning of the data

1222 The phase space used for the analysis is binned to have similar number of events in each bin for the
 1223 t , E_γ and ξ variables. This aims at achieving similar error bars in each bins. There are four bins in
 1224 $-t$, three in E_γ and ξ . For the lepton invariant mass, the ranges from 1.5 GeV to 2 GeV and from
 1225 2 GeV to 3 GeV are divided in two bins each, in order to study the possible effect of vector-meson
 1226 contamination in the low-mass bin. Besides, the binning in $-t$ is different in the two mass ranges. The
 1227 bin limits are summarized in Table 3.3 and superimposed on the kinematic distributions of the data
 1228 in Figure 3.31.

Variable	Bin limits
$-t$ (1.5 GeV < M < 2 GeV)	0.15 - 0.25 - 0.34 - 0.48 - 0.8
$-t$ (2 GeV < M < 3 GeV)	0.15 - 0.35 - 0.45 - 0.55 - 0.8
E_γ	4.0 - 6.4 - 8.0 - 10.6
$M = \sqrt{Q^2}$	1.5 - 1.7 - 2.0 - 2.5 - 3.0
ξ	0.0 - 0.12 - 0.15 - 0.4

Table 3.3: The binning grid used for the data in this analysis. Two different binnings for $-t$ are used, one for each invariant mass range.

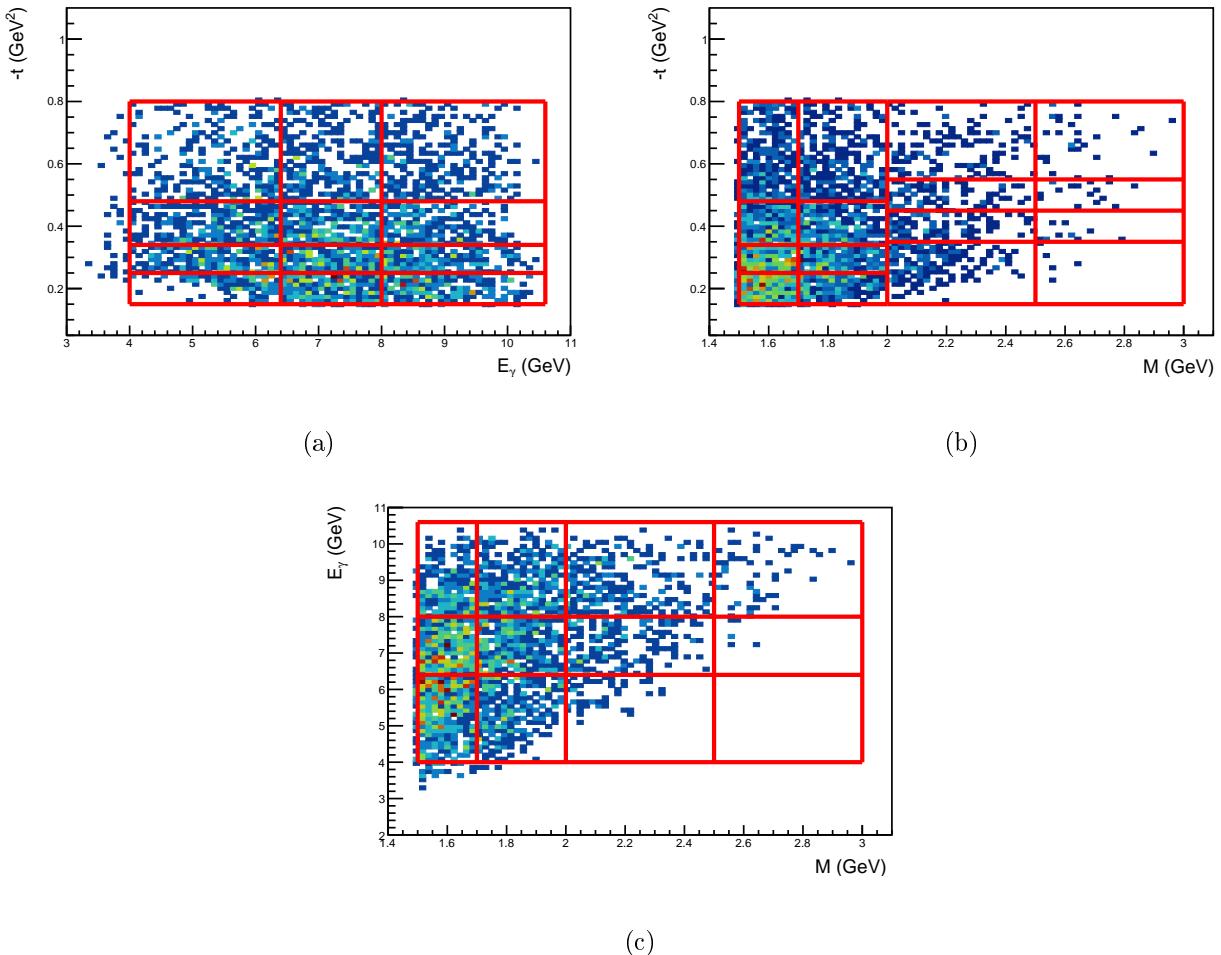


Figure 3.31: Binning grid for the TCS observables in (a) the E_γ - $(-t)$ plane (the binning in $-t$ used in this figure is the one used for the [1.5 GeV – 2 GeV] mass range), (b) the M - $(-t)$ plane and (c) the M - E_γ plane.

1229 3.12 Systematic errors estimation

1230 In this section the methods used to estimate the systematic errors are described. Four different
 1231 sources of systematic uncertainties have been studied. For each source of uncertainty, a value of
 1232 systematic shift is calculated for every observable and for each bin. Table 3.4 summarizes the sources
 1233 of systematic uncertainties tested in this analysis.

1234 Combining systematic uncertainties

1235 An issue with the determination of systematic uncertainty in this study is its large correlation with
 1236 the statistic in each bin. Indeed, especially for A_{FB} , the number of events in each bin can be quite low
 1237 (few tens) and changing parameters of the analysis can impact the value of the observable extracted,
 1238 leading to large fluctuations of the uncertainties between neighboring bins. However we expect the
 1239 systematic uncertainties to be smoothly varying functions. To mitigate these variations, a floating
 1240 average of each uncertainty is made for each bin. In other word, for a given source of systematic
 1241 uncertainty, the value assigned to a bin will be the arithmetic average between the uncertainty in this
 1242 bin and its non-empty neighbors. This method allow to smooth the bin-to-bin variation of each source
 1243 of systematic uncertainty. The difference source of uncertainties are then added in quadrature. Note
 1244 that this is done during the combining of the uncertainties and that in the following the uncertainties
 1245 for each source are shown unaveraged. Only the reported total uncertainty includes the bin-to-bin
 1246 average.

Source of uncertainty	Standard value	Alternative values
Method	Generated events	Full chain simulations
Positron ID	0.5	0.2/0.8
χ^2 proton	No cuts	3- σ cuts on FD and CD
Acceptance model	BH-weight from <i>TCSGen</i>	Weights set to 1
Proton efficiency	BG merging + eff. corr.	BG merging only
Excl. cuts	$\left \frac{P_{tX}}{P_X} \right < 0.05, M_X^2 < 0.4 \text{ GeV}^2$	$\left \frac{P_{tX}}{P_X} \right < 0.05 \pm 0.01, M_X^2 < 0.4 \text{ GeV}^2 \pm 0.1$

Table 3.4: List of the sources of systematic uncertainties that have been tested. For each source, the standard value used in the analysis is quoted. The alternative values used in the systematic error estimation are also listed.

1247 Method

1248 The method used to calculate the observable involves binned acceptance corrections and bin volume
 1249 corrections for the FB asymmetry. To estimate the impact of these corrections on the extracted values,
 1250 the method systematic error is computed using simulations. First the observables are calculated using a
 1251 sample of generated BH-weighted events, without passing them through GEMC and with no corrections
 1252 applied. This corresponds to the green point in Figure 3.33a. In the case of the R' ratio, only events
 1253 within a non-zero acceptance bin are kept. In the case of the FB asymmetry, the observable is calculated
 1254 within the full angular bin defined in Section 3.8. Finally, in the case of the BSA, a random beam
 1255 helicity is associated to each event. Then the observables are calculated a second time with the full-
 1256 chain simulation events, applying all the corrections, except that the acceptance is not corrected for
 1257 efficiency and data-driven momentum corrections are not applied. Two samples of simulation events
 1258 are used, one using weighted events from *TCSGen* and one generated with *GRAPE* (blue points and
 1259 red points on Figure 3.33a). The systematic uncertainty associated with the extraction method is
 1260 the difference between the ideal case (no corrections) and the value extracted after the full analysis
 1261 procedure. The systematic error is set between 0 and the difference between the ideal case and the
 1262 extracted case for the two simulation samples. This procedure can thus result in an asymmetric error
 1263 in order to reflect the shift induced by the measurement method. This systematics mostly affects
 1264 observables binned in ξ (for example the R' ratio in Figure 4.5), as the acceptance is not binned in this
 1265 variable. The maximum error induced on the extracted observables is at most 0.1 for the A_{FB} binned
 1266 in ξ in Figure 4.10.

1267 Efficiency / Background merging

1268 The CLAS12 background merging procedure, aiming at reproducing the detection efficiencies, is
 1269 used in this analysis. A second method, a method to estimate the proton efficiency is described in
 1270 Section 2.8. The systematic error associated with these corrections is estimated by the difference of the
 1271 values of the observables obtained with both methods, Δ_{Eff} . The systematic error bar is then defined
 1272 as $\pm \Delta_{Eff}/2$. Figure 3.33d illustrates the determination of the efficiency systematic error, with blue
 1273 points produced using efficiency corrected acceptance and red points without the correction. As shown
 1274 in Section 4.1, the efficiency systematics is most of the time dominated by other sources of systematic
 1275 errors for most of the observables. The maximum observed shift on the extracted observables is 0.1
 1276 for the A_{FB} as a function of $-t$ (Figure 4.12) and the BSA as a function of M (Figure 4.13).

1277 Positron identification

1278 In this analysis the positron identification algorithm plays an important role. To estimate the impact
 1279 of the remaining pion contamination, the cut applied on the output of the neural network is varied
 1280 around the chosen value, 0.5 ± 0.3 . The values are chosen according to considerations described in
 1281 Section 2.3.10. The acceptance is also recalculated accordingly. The difference between the observables
 1282 extracted with the standard and the shifted cuts is assigned as the positron identification systematics.
 1283 This systematics can be asymmetric as the variation of the extracted observables with the shifted

1284 positron cuts can be different in each case. As illustrated in Figure 3.33c, the variation is small for
 1285 most of the observables. The induced shift is at most 0.1 for the A_{FB} as a function of $-t$ in the
 1286 high-mass region in Figure 4.12, but usually limited to 0.03 for most of the observables.

1287 **χ^2 proton**

1288 In the analysis, no cut on the χ^2 of the proton is applied. This quantity provided by the CLAS12
 1289 event builder is related to the match of the measured time-of-flight of the proton to the expected one
 1290 calculated from momentum. In order to assess the quality of the proton identification and its impact
 1291 on the results of the analysis, the latter is done applying a 3σ cut on the proton as:

$$|\chi^2 - \mu| < 3 \cdot \sigma, \quad (3.35)$$

1292 where the mean μ and width σ are obtained from data and simulation via the fits shown in Figure 3.32.
 1293 Their values are given in Table 3.5. The *simulation* values are applied on simulation when computing
 acceptance, the *data* values are applied on data.

Sample	Region	μ	σ
Simulation	CD	1.413	1.897
Simulation	FD	-0.29	1.812
Data	CD	0.81	1.972
Data	FD	0.26	1.207

Table 3.5: Values of the means μ and widths σ of the χ^2 of the proton in the CD and the FD for data and simulation.

1294
 1295 Figure 3.33b illustrates how this systematics is calculated. This systematic uncertainty can produce
 1296 a shift up to 0.05, but is most of the time dominated by the other sources of uncertainties.

1297 **Acceptance model**

1298 The dependence of the extracted results on the model used in the acceptance is also studied. The
 1299 acceptance is calculated with BH-weighted events and events without weights (i.e. phase-space gener-
 1300 ator). The difference between the two methods Δ_{Acc} is defined as the acceptance model systematics,
 1301 and the associated error bar is set to $\pm \Delta_{Acc}/2$. The method is illustrated in Figure 3.33f. This source
 1302 of systematics becomes large for observables with low statistics, such as the A_{FB} , for which the induced
 1303 absolute shift can reach values up to 0.05 (as in Figure 4.7).

1304 **Exclusivity cuts**

1305 Finally, the last source of systematic uncertainty studied is the values chosen for the exclusivity cuts.
 1306 To estimate this uncertainty, the analysis is performed with tighter/wider cuts than those presented
 1307 in Section 3.2:

$$\left| \frac{Pt_X}{P_X} \right| < 0.04 \text{ and } |M_X^2| < 0.3 \text{ GeV}^2, \quad (3.36)$$

1308 OR

$$\left| \frac{Pt_X}{P_X} \right| < 0.06 \text{ and } |M_X^2| < 0.5 \text{ GeV}^2. \quad (3.37)$$

1309 In each case the acceptance is recalculated with the given set of cuts. The systematic uncertainty
 1310 is determined as for the positron ID. The maximum deviation above and under the standard value is
 1311 assigned as the uncertainty, thus this systematic can be asymmetric. This systematics dominates in
 1312 bins where the statistics is limited, especially for the A_{FB} and the BSA at high mass in Figure 4.13.
 1313 This shows that the exclusivity cut error is largely correlated with the statistical error in these bins.
 1314 However, the observed systematic shift always remains within the statistical error bars.

1315 Statistics of the simulation sample

1316 The impact of the number of events in the simulation sample used to produce the acceptance is
1317 also verified. The results obtained with a smaller simulation set (16M events, instead of 36M events)
1318 are compared to the results obtained with the standard procedure. Figure 3.34 shows the difference
1319 between these results. One can see that the size of the simulation sample has a very small impact on
1320 the final results. For this reason, we do not include this difference in the systematic uncertainties.

1321 Lepton neural network PID: further checks

1322 The analysis was performed using the neural network classifier for both electrons and positrons and
1323 for all momenta. The acceptance is recalculated accordingly. The difference in the final observables
1324 obtained with the standard procedure (neural network only applied to positrons and only for momenta
1325 bigger than 4 GeV) is shown in Figure 3.35. One can see that the impact of such a change is small
1326 compared to the other systematic errors studied so far. For this reason, we only consider these results
1327 as cross checks and do not include this error in the final systematics.

1328 Background merging check

1329 A method to estimate the proton efficiency is described in Section 2.8. The background merging, used
1330 to reproduce the detection efficiencies is presented in Section 2.7. In the thesis from which this note
1331 is extracted, both techniques were compared one to one. Although one should use both background
1332 merging and proton efficiency corrections, it is worth comparing the two in order to assess the effect
1333 of efficiency correction on the results of this analysis. The maximum observed shift on the extracted
1334 observables is 0.1 for the A_{FB} as a function of $-t$ (Figure 3.36a) and the BSA as a function of
1335 M (Figure 3.36b). The size of this uncertainty is comparable with the one obtained in the above
1336 subsection *Efficiency / Background merging*. Thus one can only consider the later in the calculation
1337 of the final uncertainty.

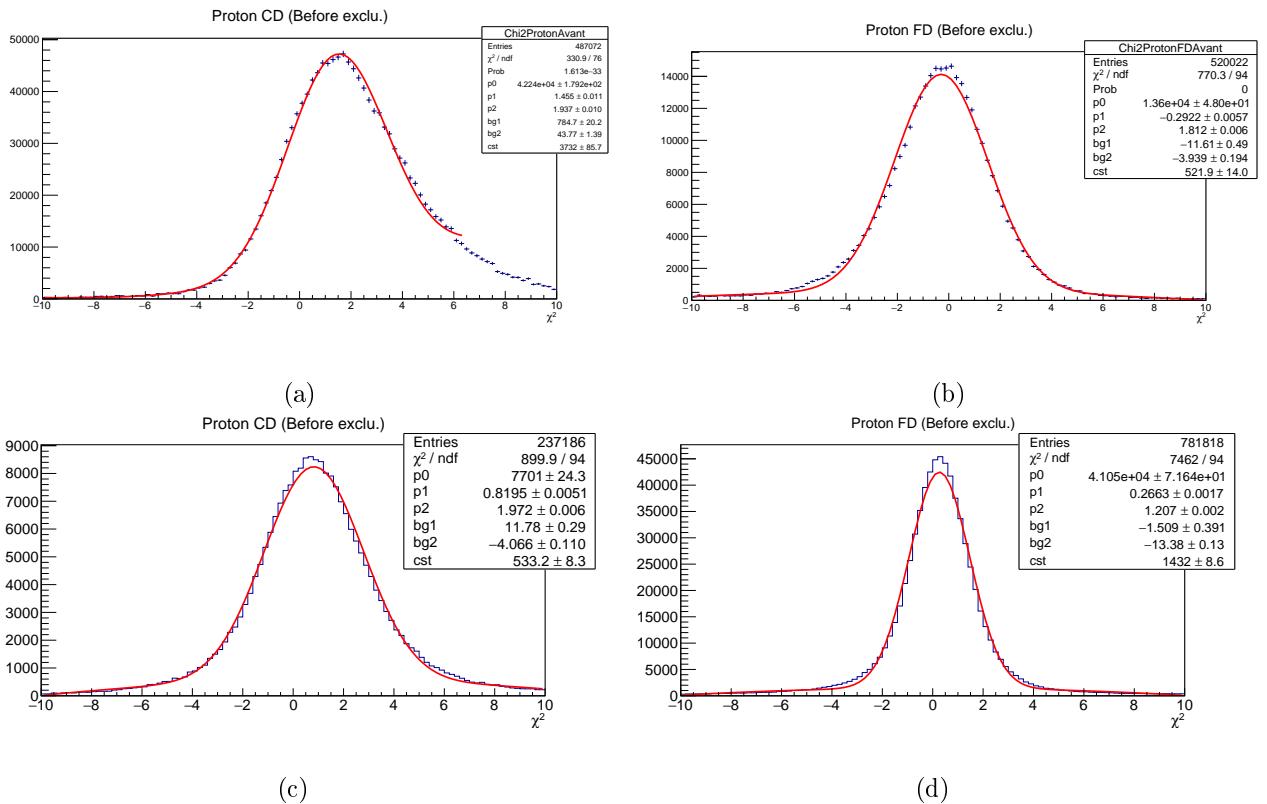


Figure 3.32: Fits used to determine the values shown in Table 3.5. A gaussian plus second order polynomial is used. (a) and (b) show the fits from simulation for the CD and FD respectively. (c) and (d) show the fits from data for the CD and FD respectively.

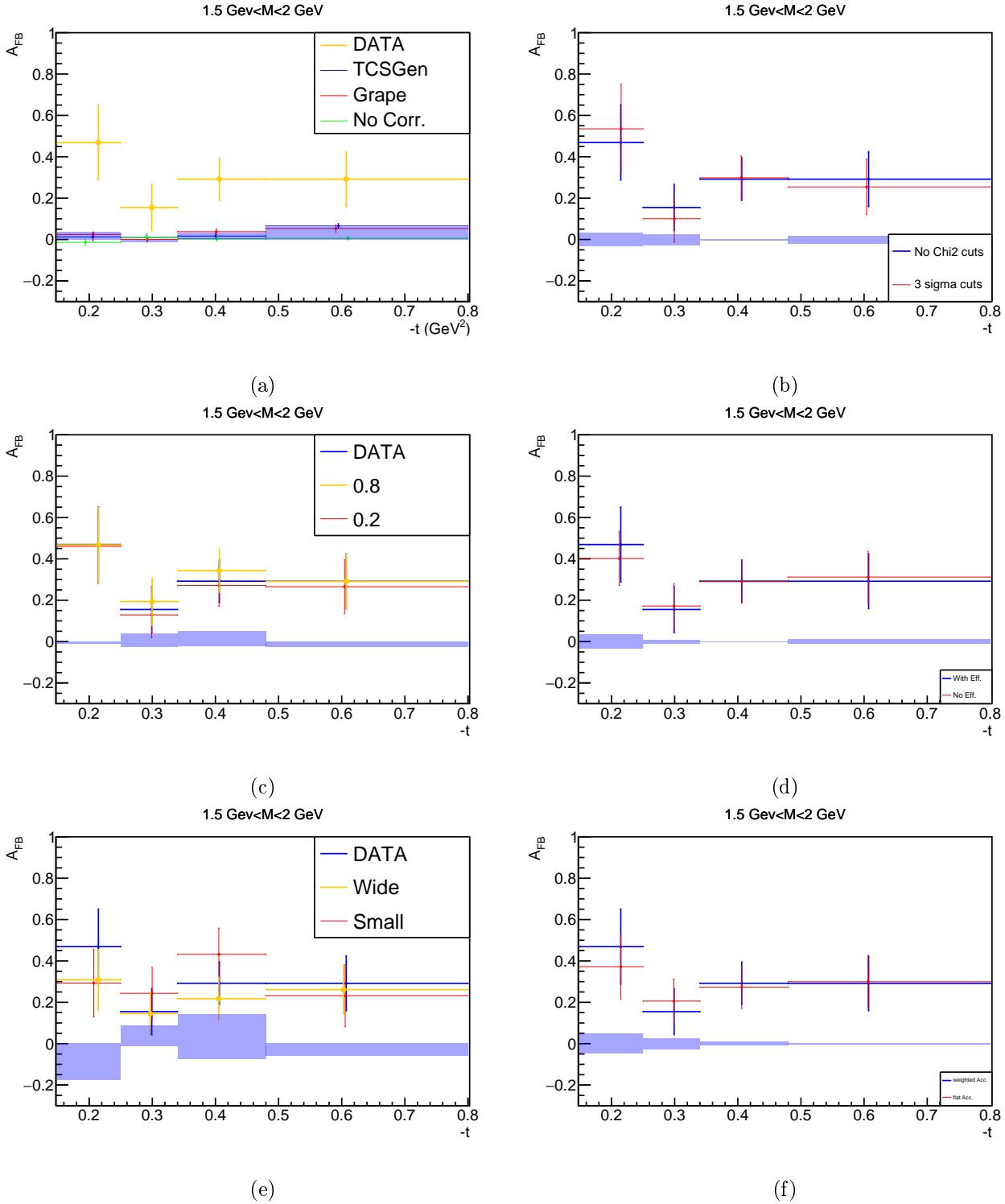


Figure 3.33: Systematic errors for the FB asymmetry in the $1.5 \text{ GeV} < M < 2 \text{ GeV}$ mass range, as blue bands with respect to a reference value of 0. The vertical error bars correspond to the statistical errors and horizontal error bars expand along the bin size. The same study is done for all the observables extracted in this analysis. (a) Systematic error from the extraction method. The green points are calculated from generated events, the blue and the red points are obtained after the full chain simulation and analysis. The data points are also displayed in yellow in order to compare this systematics with the statistical error bars. (b) Systematics associated with the χ^2 cut applied on the proton calculated as the difference of the observable calculated without cut (blue) and with a 3σ cut (red). (c) Positron ID systematic error. The cut on the output of the neural network is varied from 0.5 (blue) by ± 0.3 (red and yellow) to obtain the value of this error. (d) Systematic error associated with efficiency correction. (e) The error associated to the exclusivity cuts is estimated by tightening the transverse momentum and the missing mass cuts (red) with respect to the standard cut (blue). (f) Acceptance model systematics obtained by calculating the observable with a flat acceptance (red).

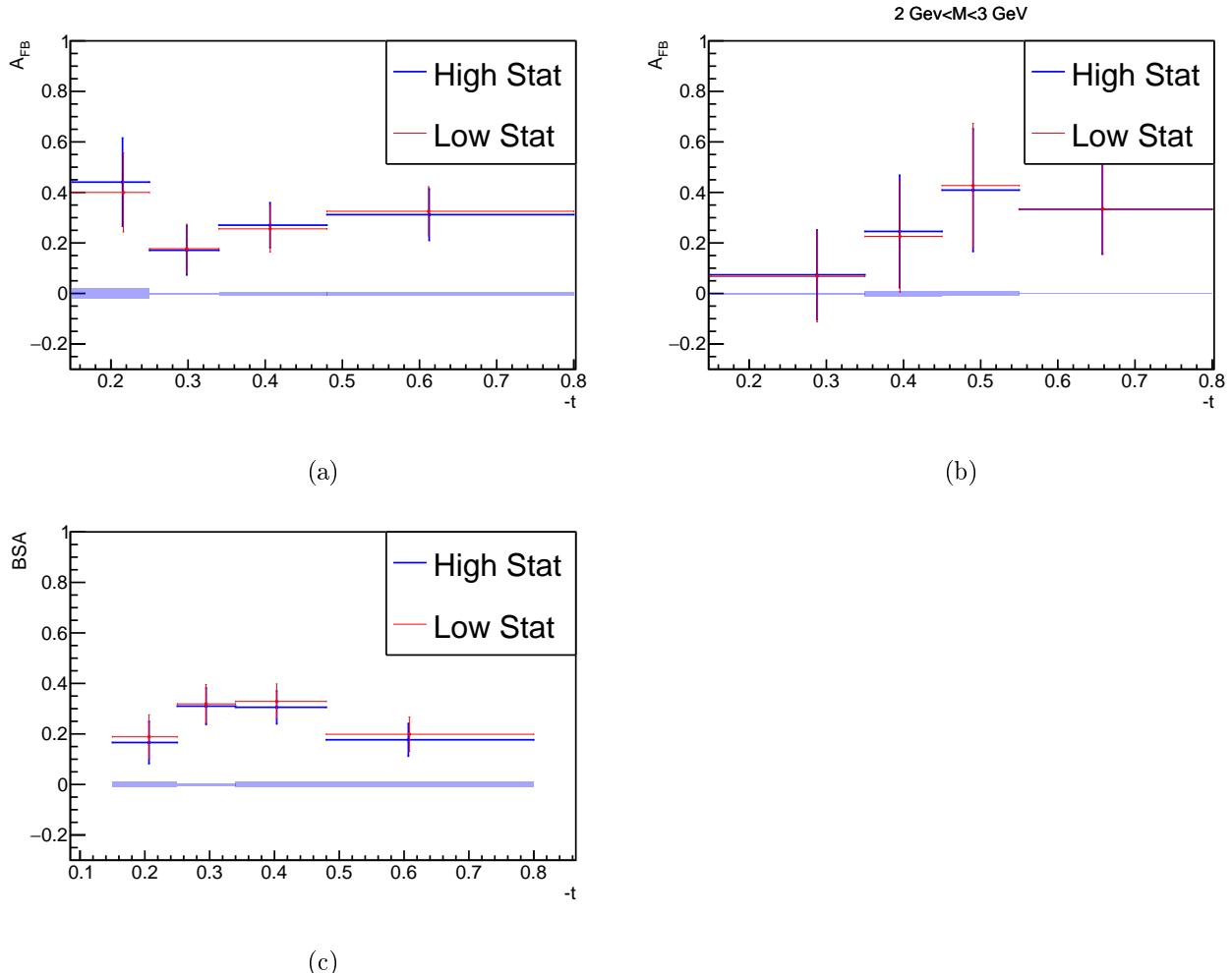


Figure 3.34: Difference of the values of the observables obtained with simulation sample of 16M events (in red) and 36M events (in blue) (a) for the integrated A_{FB} , (b) for the A_{FB} at high mass only, (c) for the integrated BSA

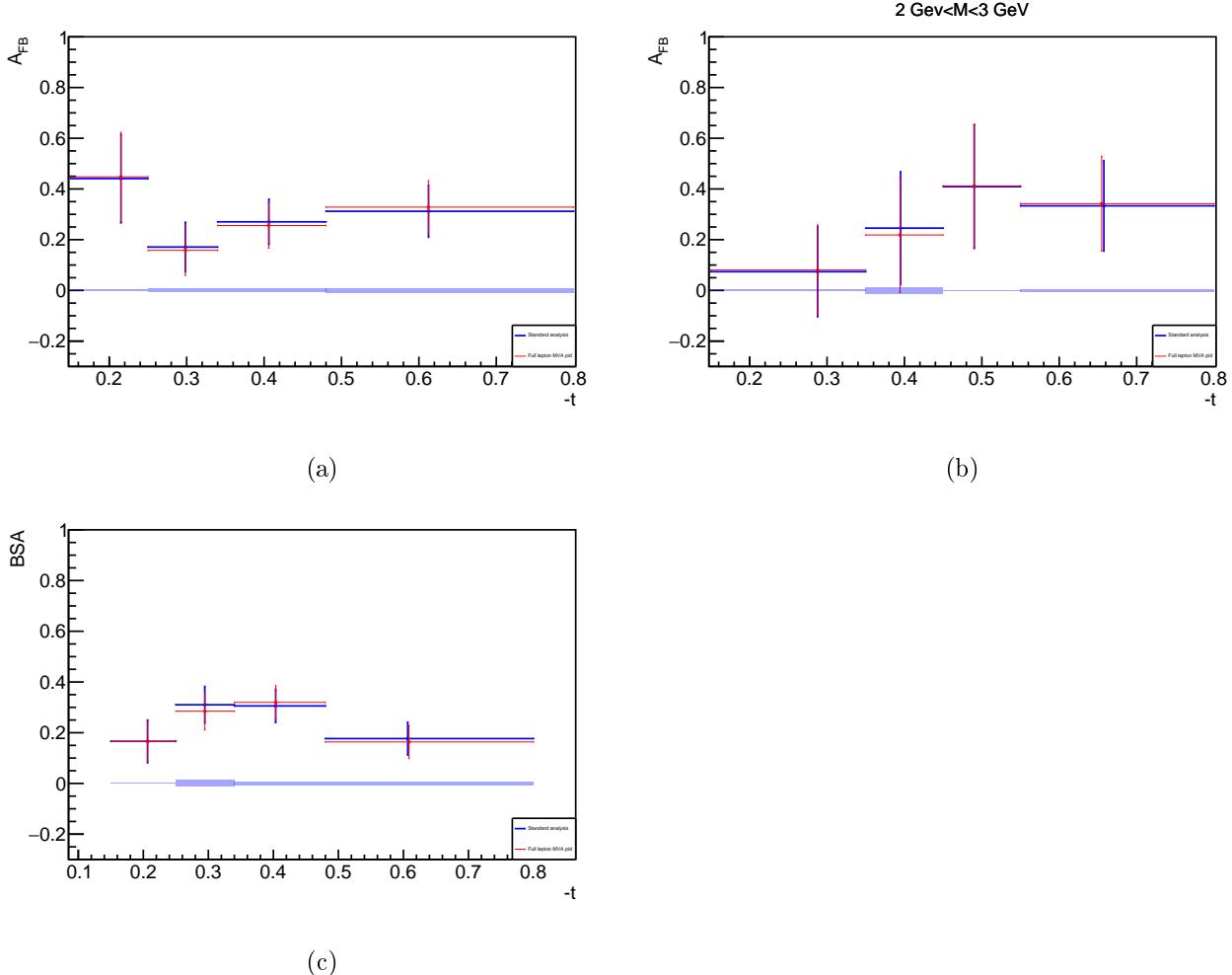


Figure 3.35: Difference of the values of the observables obtained with positron neural network applied to both electrons and positrons for all momenta (in red) and for only positrons with momentum higher than 4 GeV (standard method) (in blue) (a) for the integrated A_{FB} , (b) for the A_{FB} at high mass only, (c) for the integrated BSA

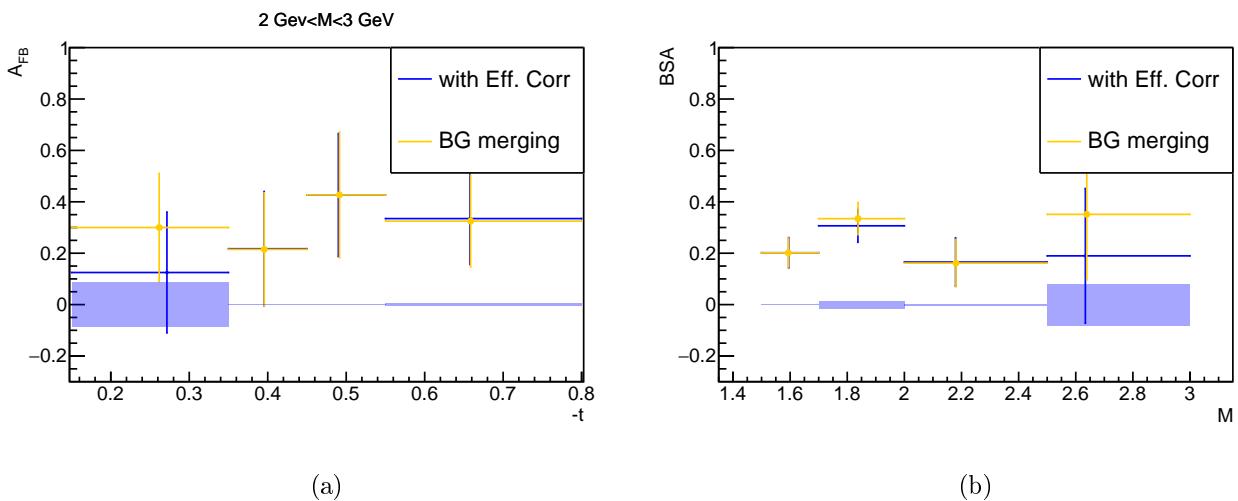


Figure 3.36: Comparison between results obtained with the proton efficiency only and with the background merging only, (a) for the A_{FB} at high mass and (b) for the BSA as a function of the mass.

1338 Chapter 4

1339 **Results and comparison with model
1340 predictions**

1341 This chapter presents the results obtained with the Fall 2018 CLAS12 dataset in the inbending con-
1342 figuration. The three TCS observables (R' ratio, A_{FB} and BSA) are extracted for different kinematic
1343 bins as a function of different variables. First, all the extracted values of the TCS observables and their
1344 systematic errors are shown in Section 4.1. A comparison with the CLAS data is shown in Section 4.4.
1345 Finally some results are compared with model predictions and discussed in Section 4.5.

1346 **4.1 Complete CLAS12 results for the TCS observables**

1347 In this section all the data points obtained in the analysis are displayed. For each observable and for
1348 each bin, the statistical error bar is shown as a vertical error bar. The horizontal error bar corresponds
1349 to the size of the bin. The horizontal position of the data points is the average value of the variable in
1350 the bin, corrected by the acceptance (and in the case of the A_{FB} , also by the bin volume correction).
1351 The grey bands show the total systematic uncertainty defined as the quadratic sum of all of the
1352 systematic contributions described in Section 3.12. The decomposition of the systematic uncertainty
1353 is shown under each plot. The red points reported on each plots correspond to the expected values
1354 of the observables in each bin for BH-only events. These points are calculated using BH-weighted
1355 simulation events and, within the acceptance of CLAS12 for the R' ratio and the BSA, and within
1356 the experimental forward and backward bin for the FB asymmetry. Finally, the mean values of the
1357 integrated kinematic variables, calculated using the same simulations, are given above each plot.

1358 **4.1.1 R' ratio**

1359 The R' ratio is extracted as a function of $-t$ (Figure 4.4) and ξ (Figure 4.5), with all the other
1360 variables integrated. The dependence on $-t$ is further explored by plotting the ratio as a function of
1361 $-t$ in the invariant mass bin [2 GeV – 3 GeV] (Figure 4.6).

1362 The measured values of the R' ratio are always bigger than the values expected if only the BH
1363 process was contributing to the $\gamma p \rightarrow p' e^+ e^-$ cross section (red points in the following figures). This
1364 behavior is also seen in the high-mass region [2 GeV – 3 GeV] in Figure 4.6, where no vector-meson
1365 contamination is expected. This observation validates the fact that the CLAS12 data are sensitive to
1366 the BH-TCS interference cross section.

1367 **4.1.2 A_{FB}**

1368 The FB asymmetry is extracted as a function of E_γ (Figure 4.7), M (Figure 4.8), $-t$ (Figure 4.9)
1369 and ξ (Figure 4.10), with all the other variables integrated. In order to explore the invariant mass
1370 dependence further, and especially to investigate the effect of low-mass vector-meson resonances, it is
1371 also measured as a function of $-t$ for the two additional mass bins, in the low-mass bin [1.5 GeV – 2 GeV]
1372 (Figure 4.11) and in the high-mass bin [2 GeV – 3 GeV] (Figure 4.12).

1373 As in the case of the R' ratio discussed above, the extracted values of the A_{FB} shown in Figures
 1374 4.7, 4.8 and 4.9 are not compatible with zero, confirming that the BH process does not contributes
 1375 alone to the $\gamma p \rightarrow p' e^+ e^-$ cross section. A none-zero asymmetry is seen in both the low-mass and
 1376 high-mass bins. The signal seen in the high-mass bin, in Figure 4.12, can be attributed to the BH-TCS
 1377 interference cross section, as there is no known vector-meson resonance in this mass range.

1378 4.1.3 BSA

1379 The BSA is extracted in CLAS12 acceptance, as a function of M (Figure 4.13), $-t$ (Figure 4.14)
 1380 and ξ (Figure 4.15), with all the other variables integrated.

1381 In each of these figures a clear photon polarization asymmetry is reported. This is a further confir-
 1382 mation that we observe the BH-TCS interference in the CLAS12 dataset, as the expected asymmetry
 1383 for the BH process only is zero.

1384 4.2 Note on the theoretical comparison

1385 In the following, we compare some of our results with theoretical predictions. In order to do so, the
 1386 data points are compared with theoretical curves calculated at the mean kinematic of the integrated
 1387 variables as (referred as Method 1):

$$\bar{x} = \frac{\int x \frac{d\sigma}{dx} dx}{\int \frac{d\sigma}{dx} dx}, \quad (4.1)$$

1388 where x is the integrated variable and σ the cross section of our process. Experimentally this quantity
 1389 is obtained as:

$$\bar{x} = \frac{\sum_{events} x \frac{1}{Acc}}{\sum_{events} \frac{1}{Acc}}, \quad (4.2)$$

1390 where the sum runs over events in the data, the calculated values are given above each figures in the
 1391 following.

1392 This method is however approximate and one should instead use the following formula to find the
 1393 point \bar{x} at which to evaluate the model prediction and where the cross section is equal to its integral
 1394 over the Δx range (Method 2):

$$\frac{d\sigma}{dx}(\bar{x}) = \frac{\bar{\sigma}}{\Delta x}. \quad (4.3)$$

1395 However the later formula requires to know the functional form of the cross section $\frac{d\sigma}{dx}$. Simulations
 1396 events are used to extract \bar{x} using Method 2. The average yield $\frac{\bar{\sigma}}{\Delta x}$ is estimated as:

$$\frac{\bar{\sigma}}{\Delta x} = \frac{\sum_{non\ empty\ bins} \sum_{events} \frac{w}{Acc}}{\text{Number of non empty bins}}, \quad (4.4)$$

1397 the (M, E_γ) bin which content is equal (within a 20% margin) to the average yield are selected. They
 1398 correspond to the points \bar{x} satisfying Equation 4.3 and are represented in yellow in Figure 4.1. One
 1399 can see that these points do not correspond to the mean kinematic point in blue on Figure 4.1. One
 1400 can calculate the standard deviation between the mean kinematic points (Method 1) and the average
 1401 yield points (Method 2) as :

$$\sigma_X = \sqrt{\frac{\sum_{y \in Yellow\ points} (X_{Mean} - X_y)^2}{\text{Nb of yellow points}}}, \quad (4.5)$$

1402 where the sum runs over the point where the average yield is reached, and X denotes either the mass
 1403 or the energy of the photon. We have shown that in our case this standard deviation is close to the
 1404 standard deviation of both M and E_γ obtained from 1-D distributions of data. The average kinematic
 1405 points and the corresponding standard deviation (evaluated using data) are provided in Table 4.1.

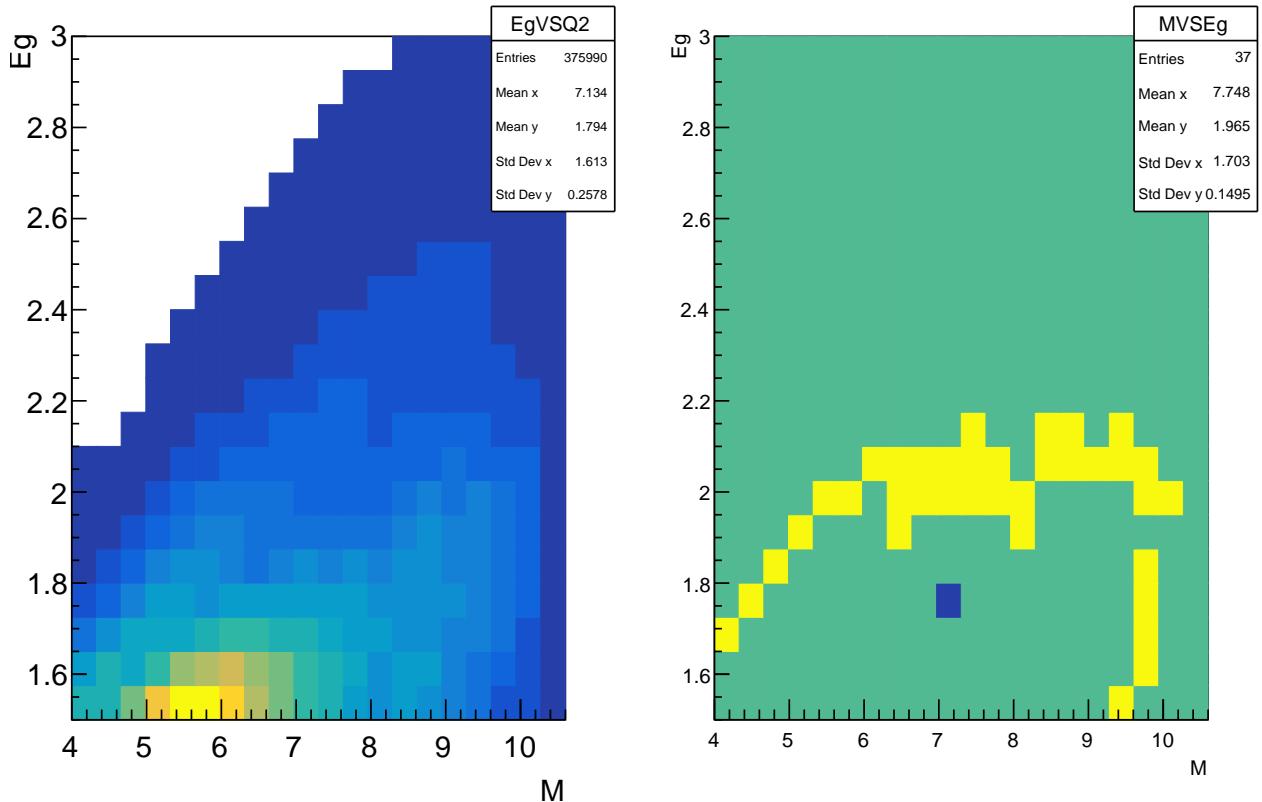


Figure 4.1: (left) Energy of the photon E_γ as a function of M for simulated events. (right) Distribution of the bins (in yellow) where the yield equates (within 20%) the average yield, and mean kinematic point in blue. Note that these points were obtained with simulated events to have enough statistics.

1406 In order to assess the effect of evaluating the model with Method 1 or Method 2, theoretical
 1407 prediction at the mean kinematic point (Method 1) and 1-sigma away from it are produced and
 1408 compared. Figures 4.2 and 4.3 show the data points and the different theoretical predictions. One can
 1409 see that all the curves are within a range of the size of the statistical error bars of the data. Thus
 1410 we have decided to compare the data points with the theoretical predictions at the mean kinematic
 1411 point and to provide 1-sigma range for these values in Table 4.1.

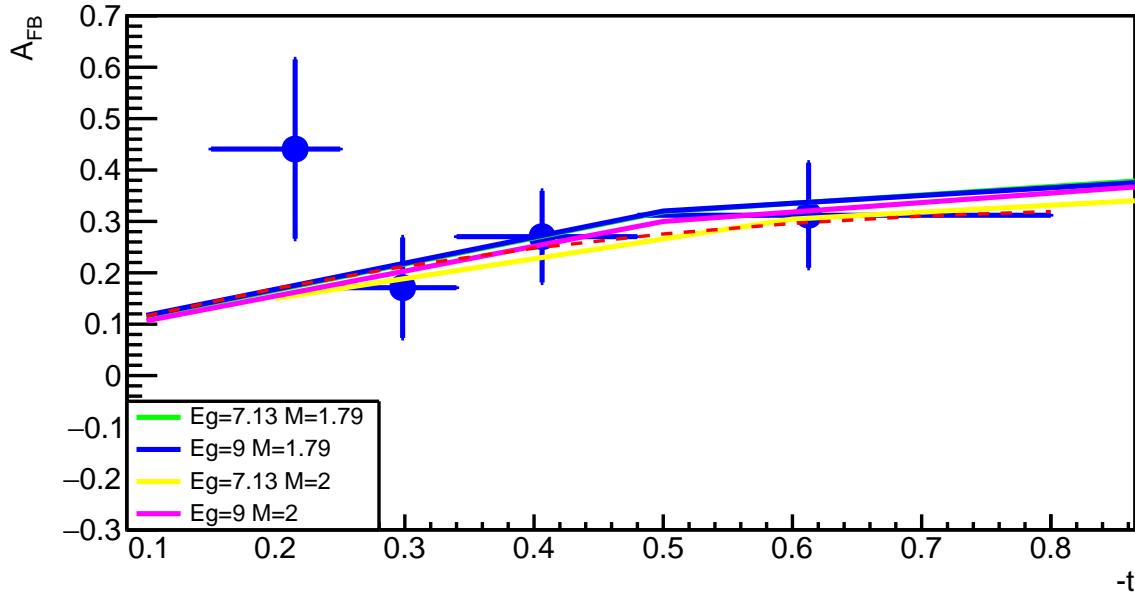


Figure 4.2: A_{FB} as a function of $-t$. The data points are compared with model prediction at the mean kinematic points and 1-sigma away from this point. The obtained theoretical curves are all within the same range.

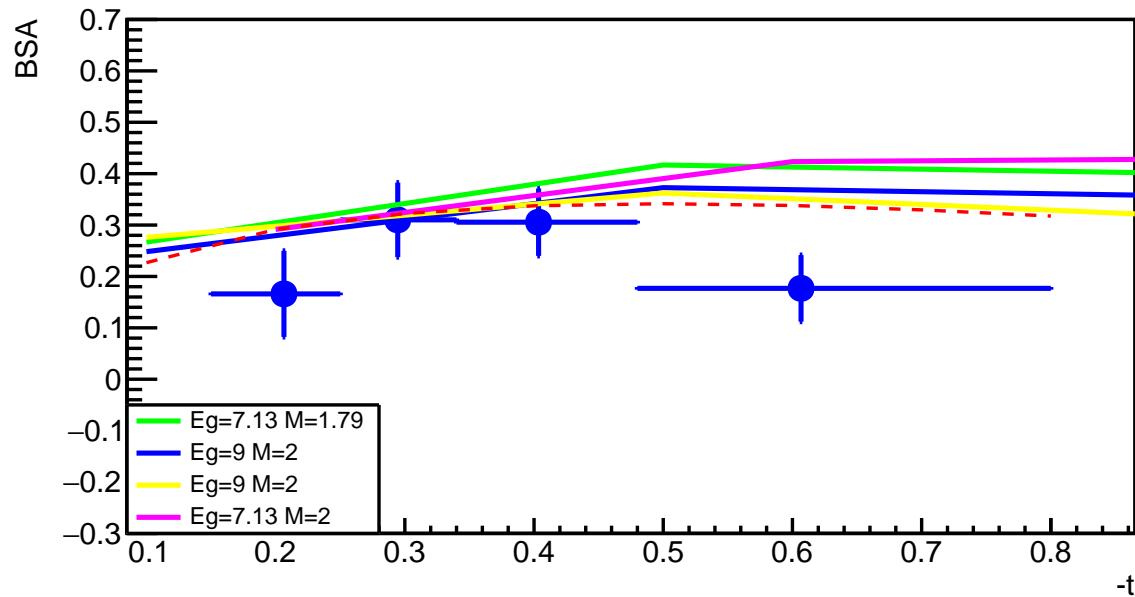


Figure 4.3: BSA as a function of $-t$. The data points are compared with model prediction at the mean kinematic points and 1-sigma away from this point. The obtained theoretical curves are all within the same range.

Observable	E_γ	M
BSA	7.29 ± 1.55 GeV	1.80 ± 0.26 GeV
A_{FB}	7.23 ± 1.61 GeV	1.81 ± 0.26 GeV
A_{FB} (2 GeV $< M < 3$ GeV)	8.13 ± 1.23 GeV	2.25 ± 0.20 GeV

Table 4.1: Values of the mean kinematics and there errors for the A_{FB} and BSA as a function of $-t$.

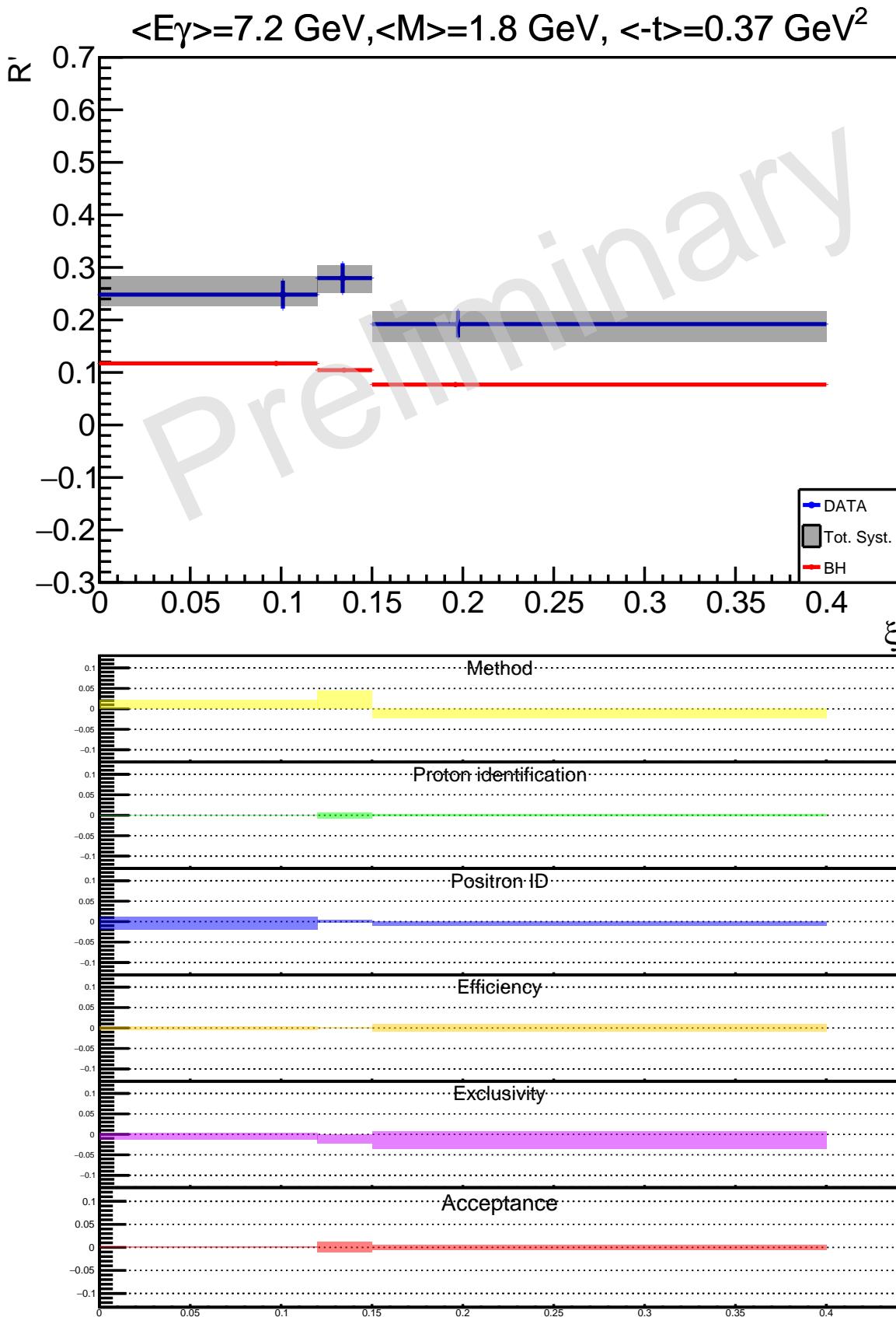


Figure 4.5: R' ratio as a function of ξ , integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.3 in Appendix 4.3.

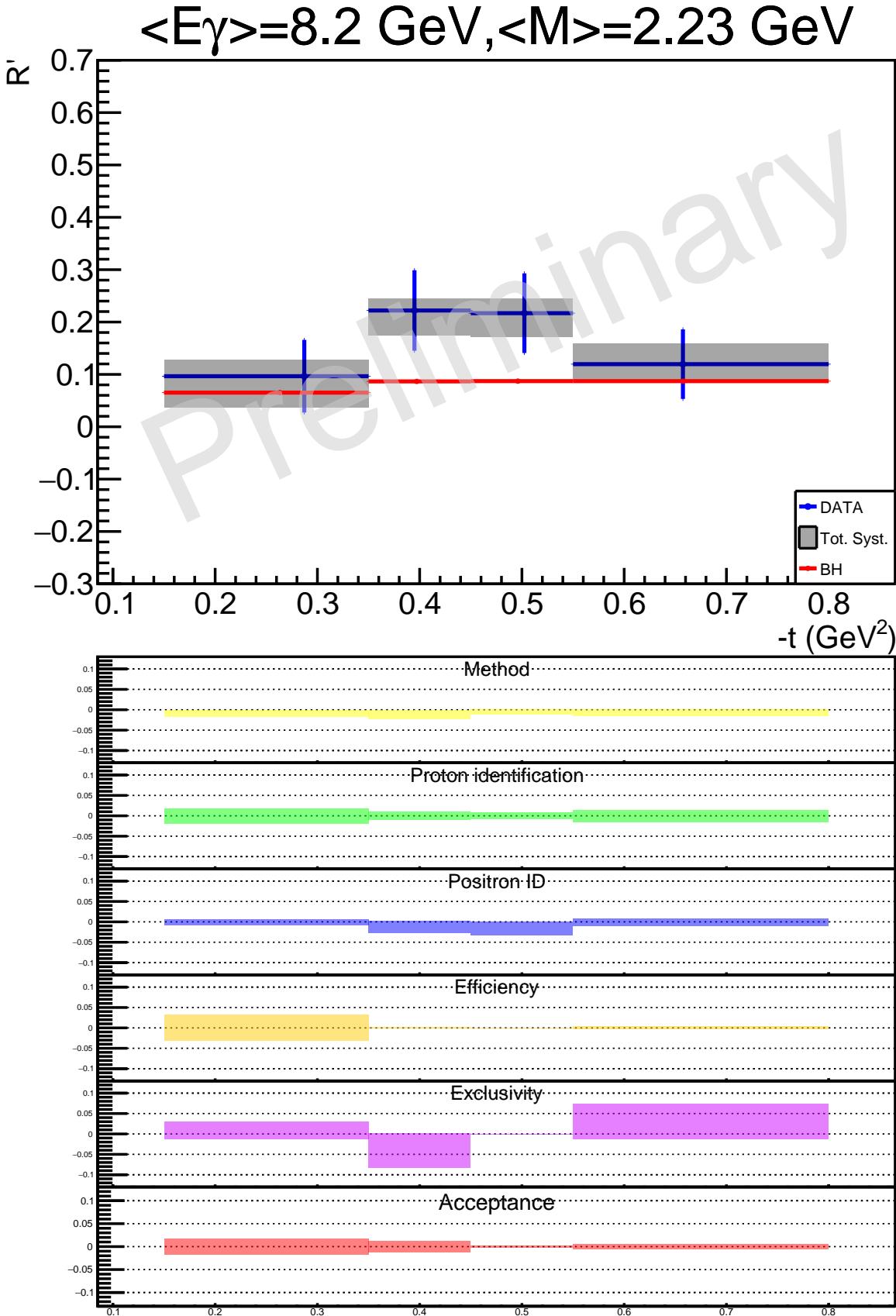


Figure 4.6: R' ratio as a function of $-t$, for the mass range [2 GeV – 3 GeV], integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.4 in Appendix 4.3.

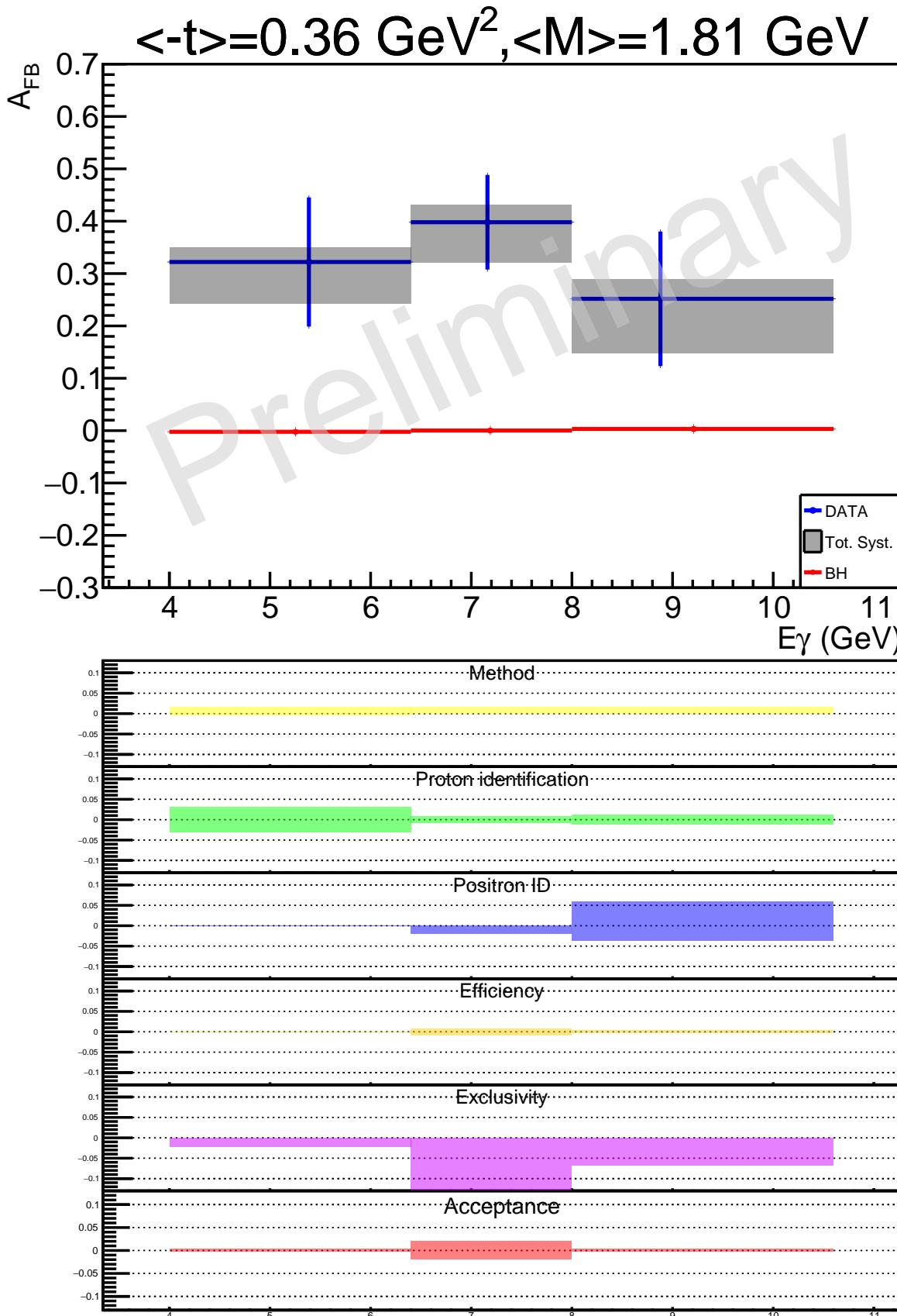


Figure 4.7: A_{FB} as a function of E_γ , integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.5 in Appendix 4.3.

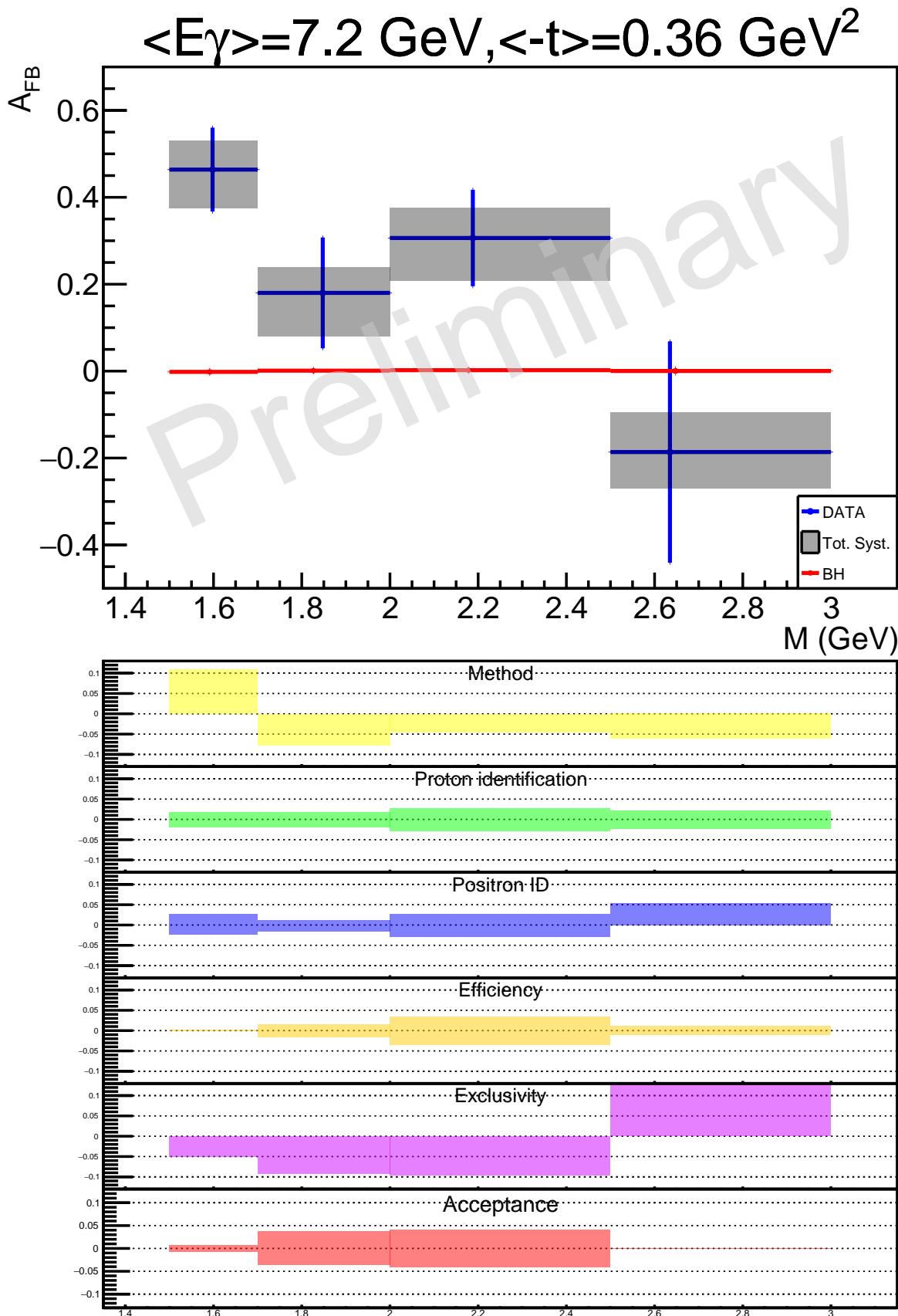


Figure 4.8: A_{FB} as a function of M , integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.6 in Appendix 4.3.

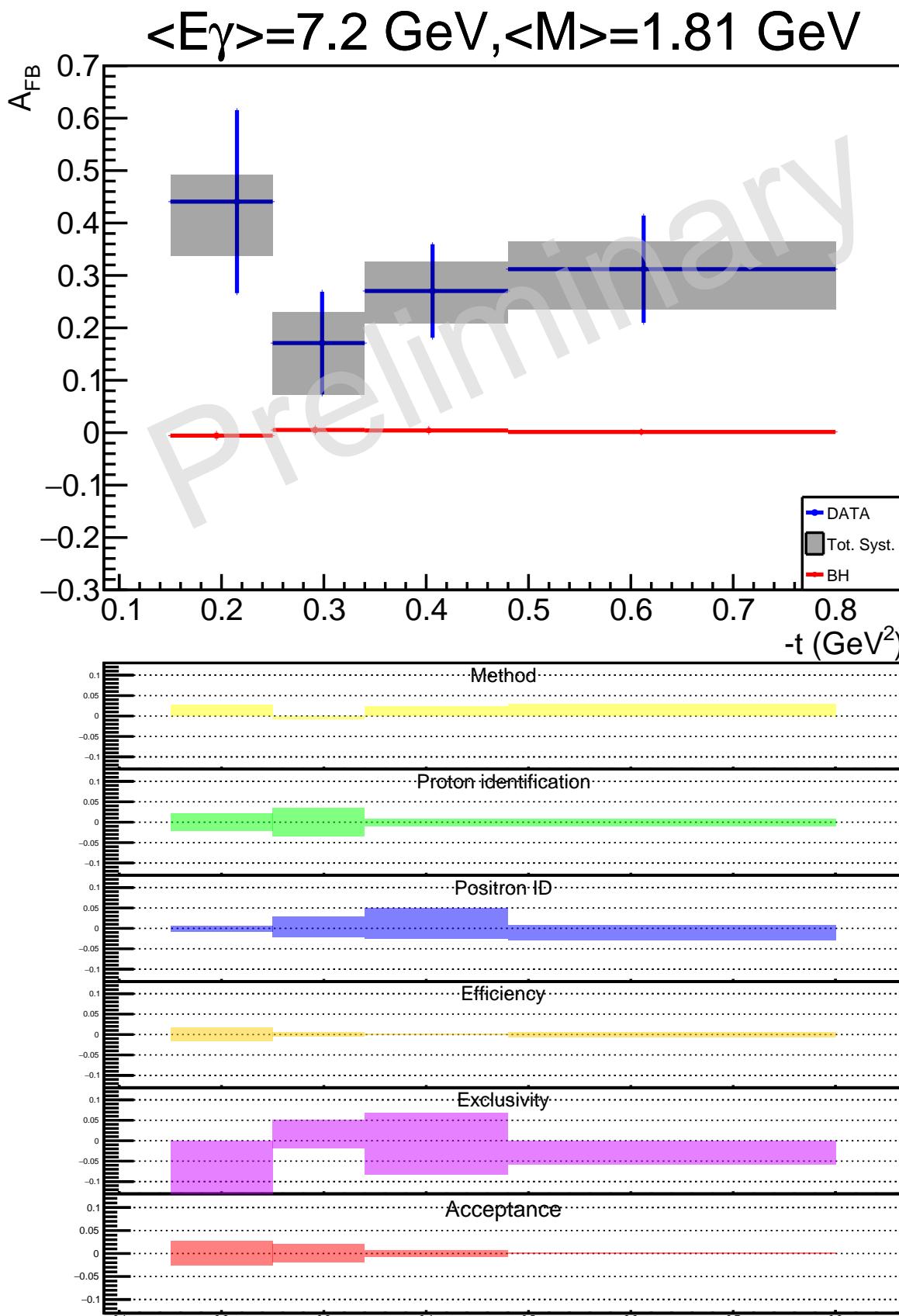


Figure 4.9: A_{FB} as a function of $-t$, integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.7 in Appendix 4.3.

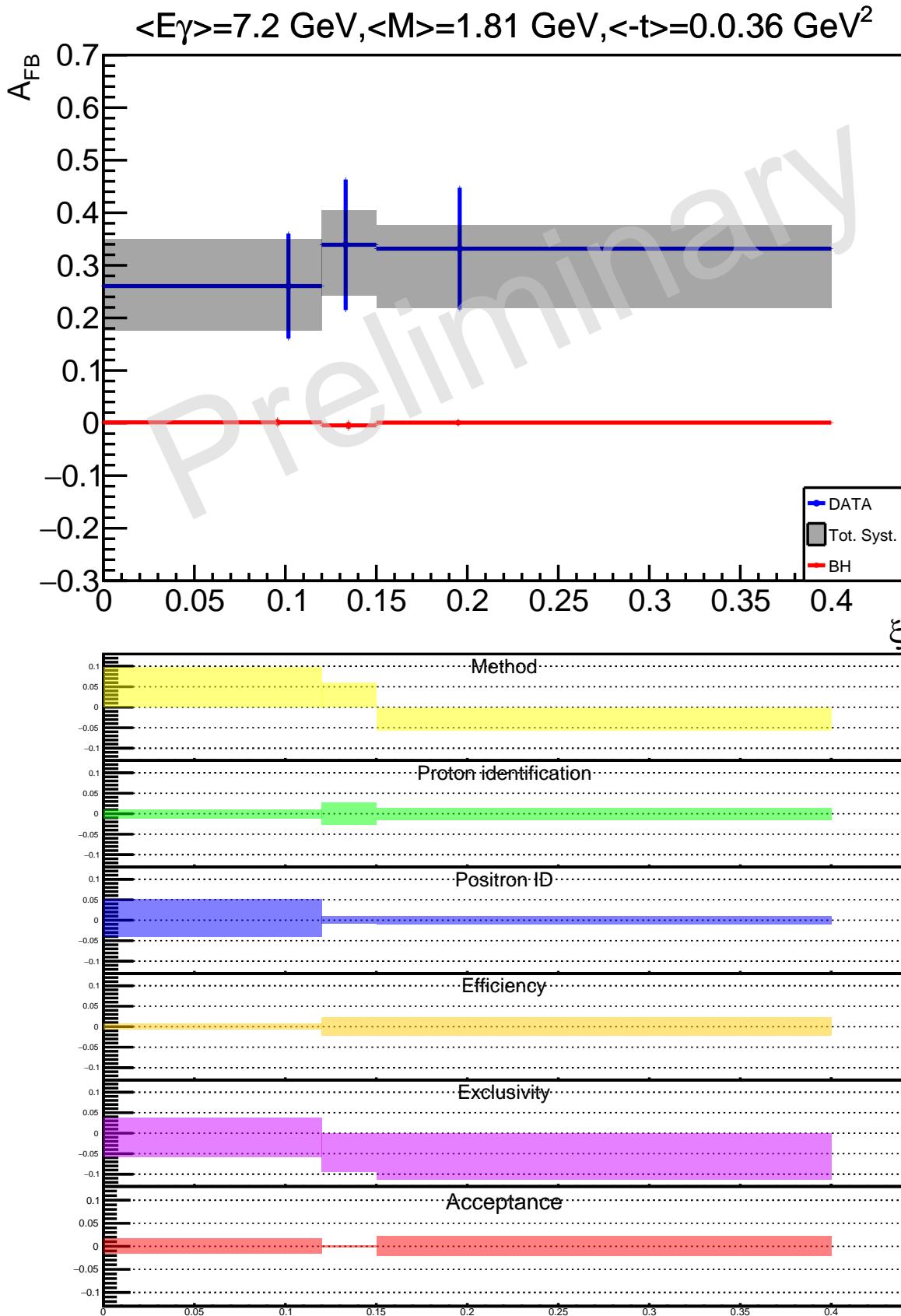


Figure 4.10: A_{FB} as a function of ξ , integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.8 in Appendix 4.3.

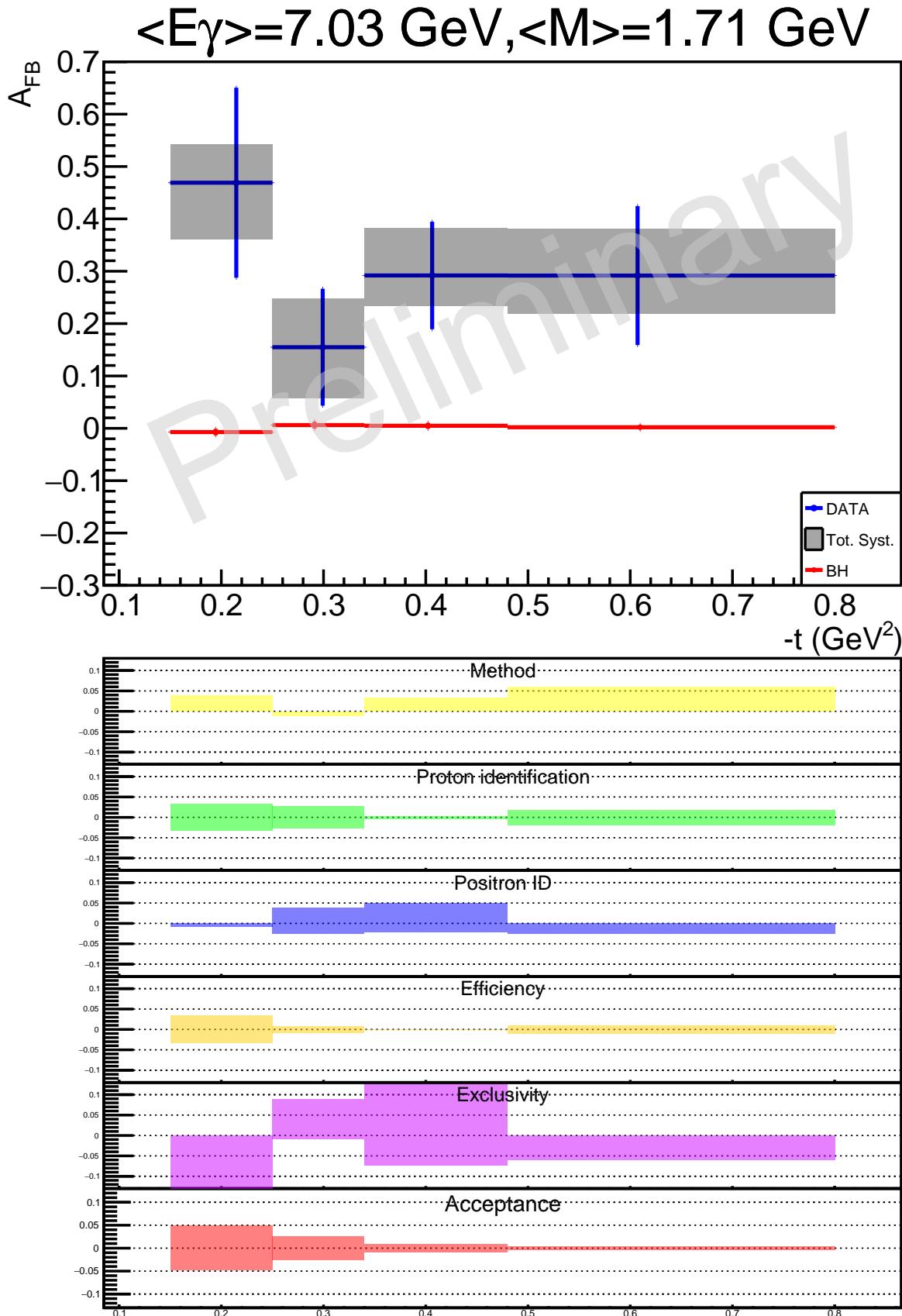


Figure 4.11: A_{FB} as a function of $-t$, in the $[1.5 \text{ GeV} - 2 \text{ GeV}]$ mass range, integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.9 in Appendix 4.3.

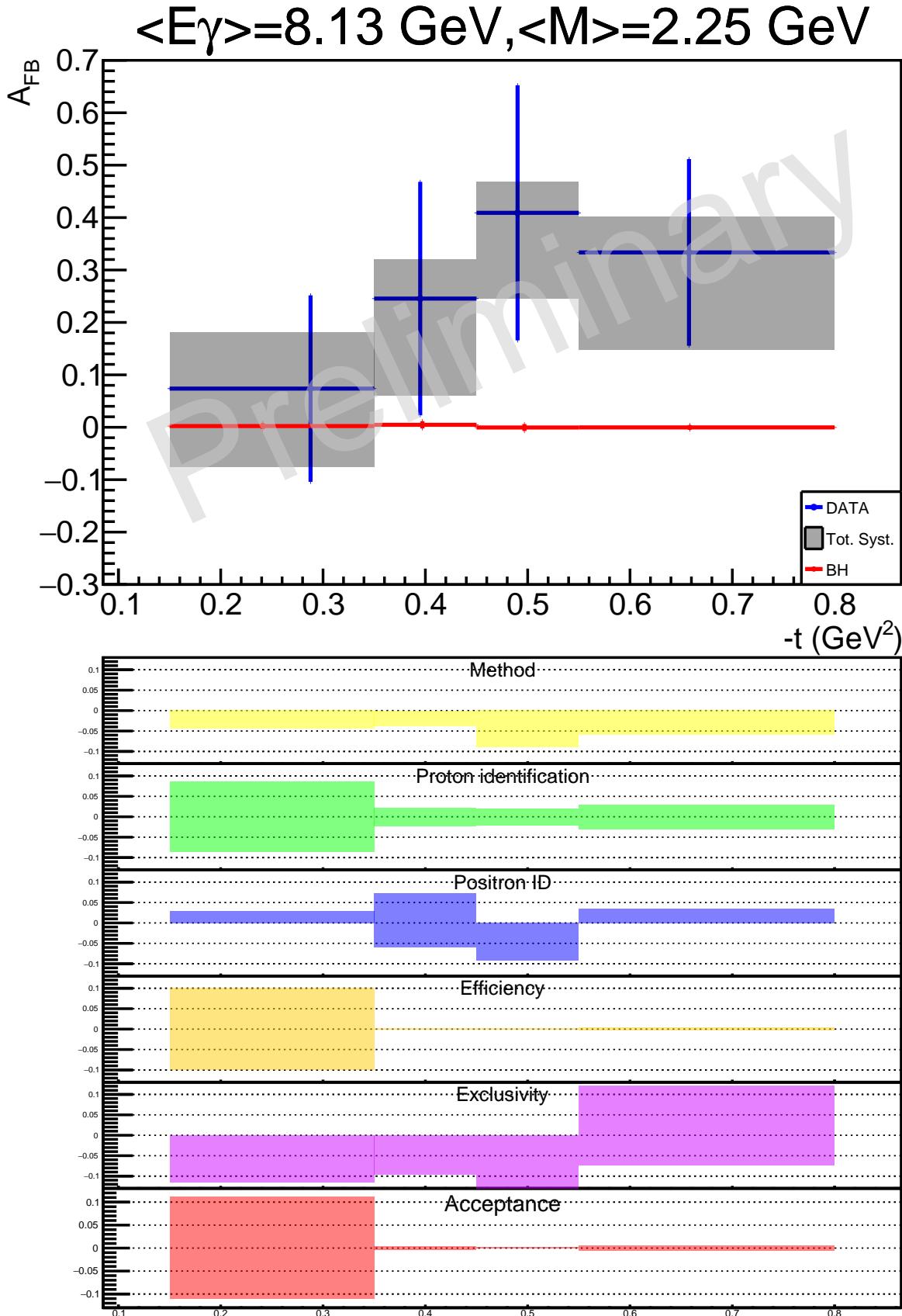


Figure 4.12: A_{FB} as a function of $-t$, in the [2 GeV – 3 GeV] mass range, integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.10 in Appendix 4.3.

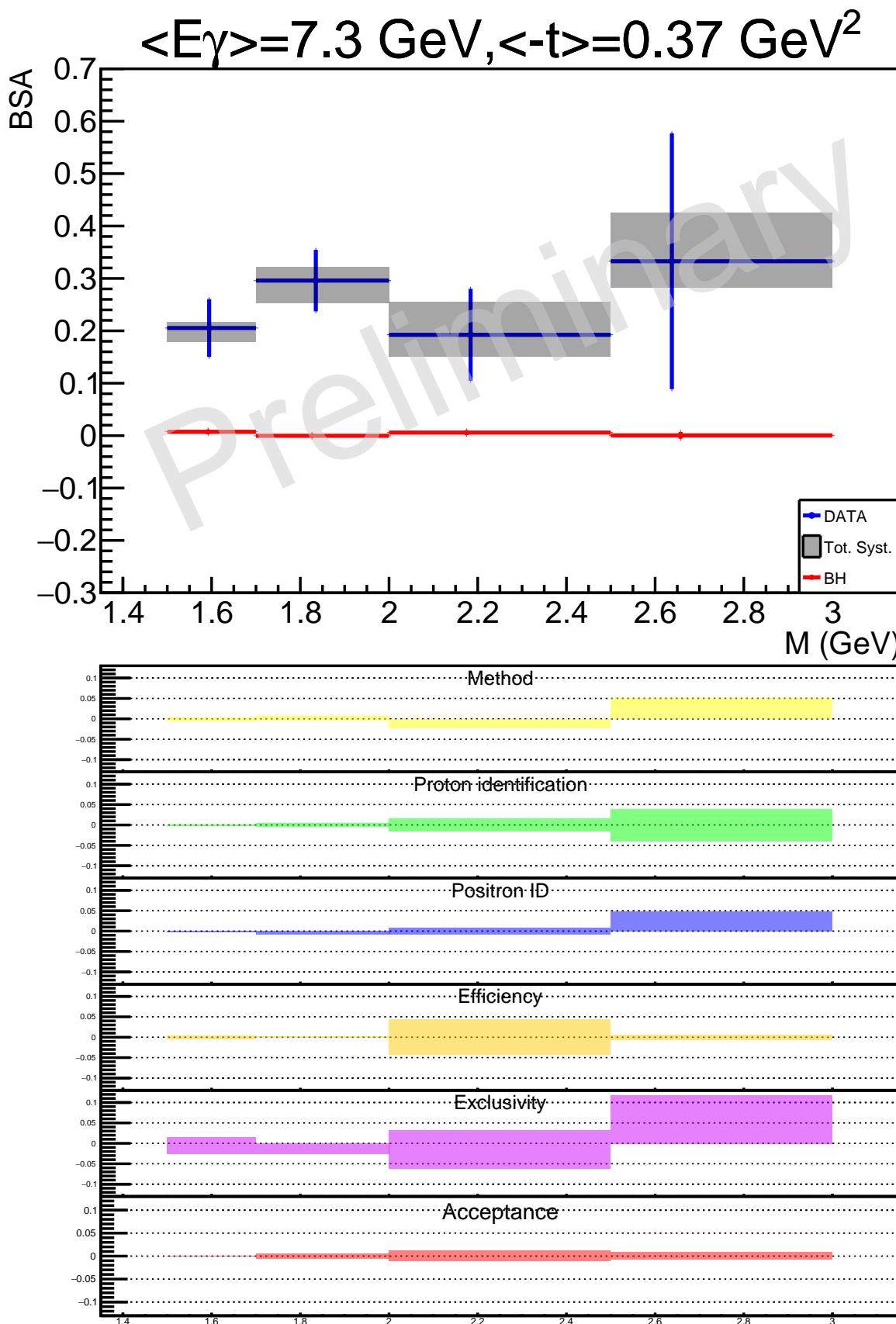


Figure 4.13: BSA as a function of M , integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.11 in Appendix 4.3.

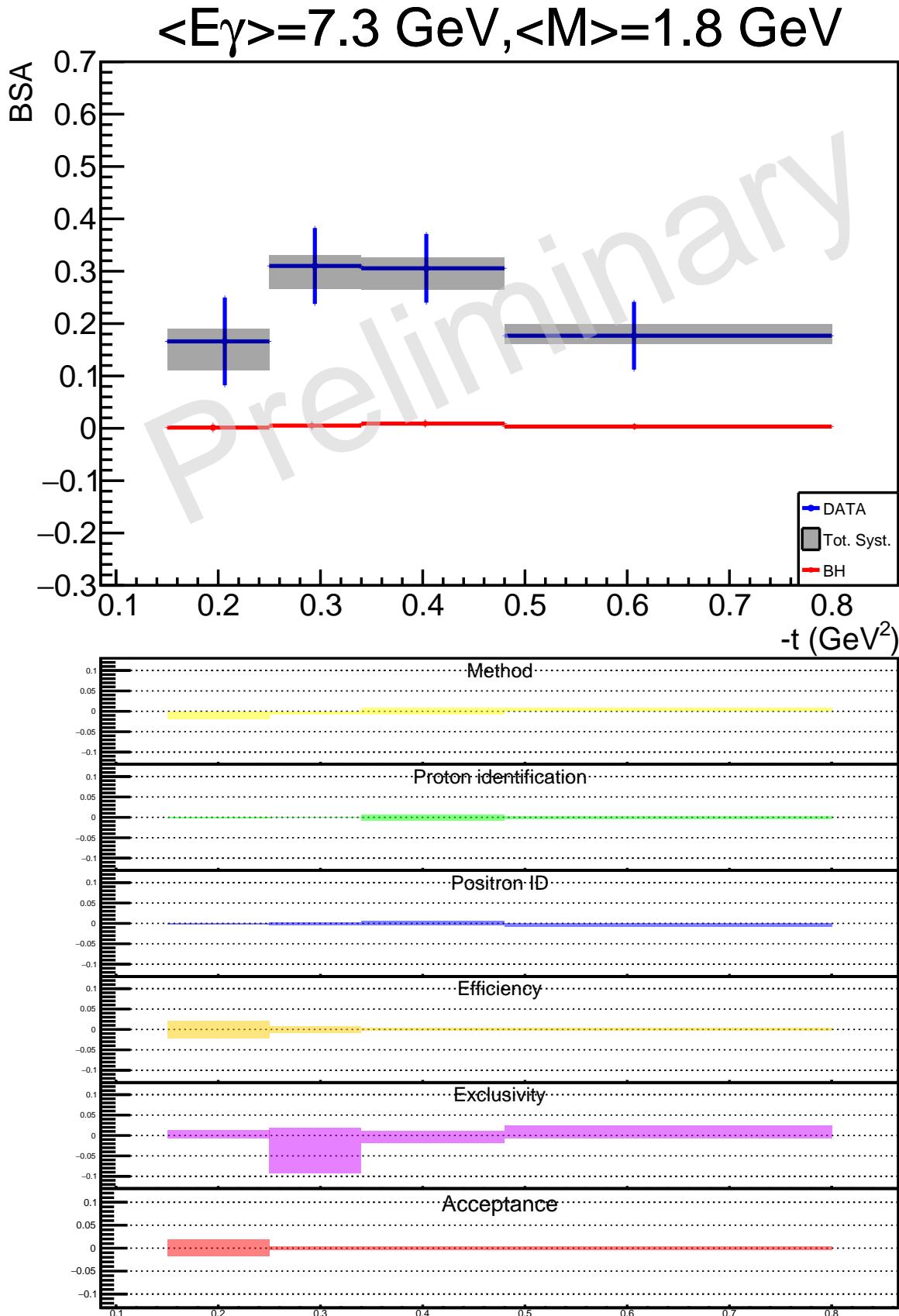


Figure 4.14: BSA as a function of $-t$, integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.12 in Appendix 4.3.

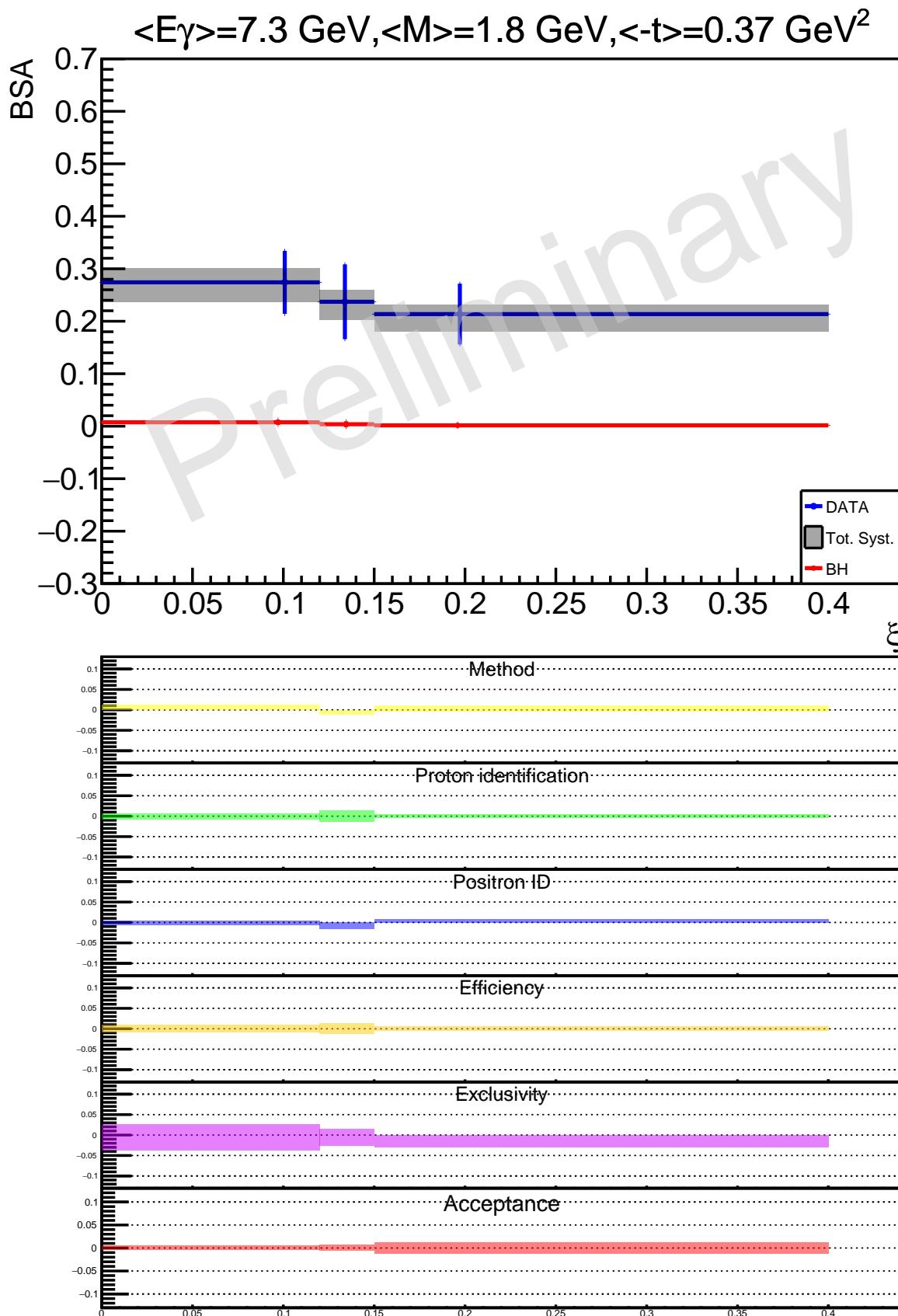


Figure 4.15: BSA as a function of ξ , integrated over all the other variables, using the same plotting conventions as in Figure 4.4. Tabulated values in Table 4.13 in Appendix 4.3.

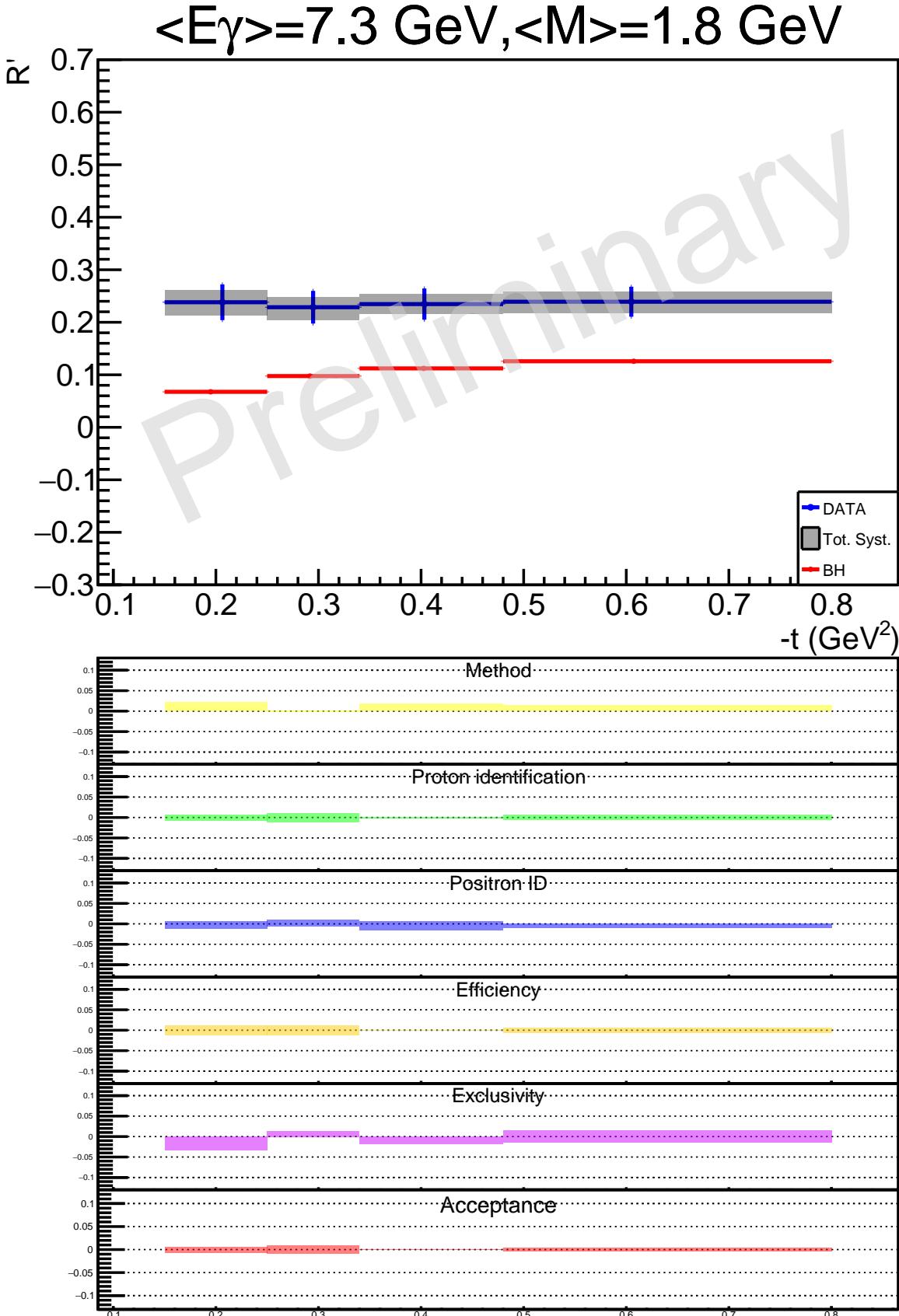


Figure 4.4: R' ratio as a function of $-t$, integrated over all the other variables. The horizontal error bars denote the bin size, the vertical error bars are statistical errors and the grey bands display the total systematic uncertainty. The red points are expected values for BH-only events, obtained with simulations. Tabulated values in Table 4.2 in Appendix 4.3.

4.3 Tabulated results

The results for the three TCS observables obtained from Fall 2018 CLAS12 dataset are tabulated in this section, with the corresponding statistical and systematic uncertainties. Tables 4.2, 4.3 and 4.4 contain results for the R' ratio, Tables 4.5 to 4.10 display the results for the FB asymmetry and Tables 4.11, 4.12 and 4.13 show the BSA tabulated results.

$-t(GeV^2)$	R'	Stat. error	Low Syst. Error	High Syst. Error
0.206	0.238	0.034	0.0242	0.0215
0.295	0.229	0.0311	0.0213	0.0185
0.403	0.235	0.0298	0.0156	0.0177
0.605	0.239	0.0286	0.0179	0.019

Table 4.2: R' ratio as a function of $-t$ (see Figure 4.4).

ξ	R'	Stat. error	Low Syst. Error	High Syst. Error
0.101	0.248	0.0266	0.019	0.0341
0.134	0.28	0.0282	0.026	0.0239
0.197	0.192	0.0255	0.033	0.0248

Table 4.3: R' ratio as a function of ξ (see Figure 4.5).

$-t(GeV^2)$	R'	Stat. error	Low Syst. Error	High Syst. Error
0.287	0.0965	0.0694	0.0576	0.0311
0.395	0.222	0.0769	0.042	0.0225
0.502	0.217	0.076	0.0393	0.0285
0.657	0.12	0.0664	0.0229	0.0395

Table 4.4: R' ratio as a function of $-t$ in the mass range [2 GeV – 3 GeV] (see Figure 4.6).

$E\gamma(GeV)$	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
5.39	0.322	0.123	0.0784	0.0282
7.16	0.398	0.0903	0.0755	0.0259
8.88	0.252	0.128	0.1	0.0239

 Table 4.5: A_{FB} as a function of E_γ (see 4.7).

$M(GeV)$	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
1.6	0.464	0.0965	0.0874	0.0627
1.85	0.18	0.128	0.0988	0.0543
2.19	0.306	0.111	0.0969	0.0628
2.63	-0.186	0.255	0.0823	0.0824

 Table 4.6: A_{FB} as a function of M (see 4.8).

$-t(GeV^2)$	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
0.215	0.441	0.175	0.103	0.048
0.298	0.171	0.0978	0.0962	0.0522
0.406	0.27	0.0892	0.057	0.0486
0.613	0.312	0.102	0.0719	0.045

 Table 4.7: A_{FB} as a function of $-t$ (see 4.9).

ξ	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
0.102	0.261	0.0998	0.0809	0.0849
0.133	0.339	0.124	0.095	0.0608
0.196	0.332	0.116	0.113	0.0442

 Table 4.8: A_{FB} as a function of ξ (see Figure 4.10).

$-t(GeV^2)$	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
0.215	0.469	0.181	0.107	0.0714
0.299	0.155	0.112	0.0949	0.0885
0.406	0.292	0.103	0.0533	0.0853
0.607	0.292	0.133	0.0693	0.0852

 Table 4.9: A_{FB} as a function of $-t$ in the mass range [1.5 GeV – 2 GeV] (see 4.11).

$-t(GeV^2)$	A_{FB}	Stat. error	Low Syst. Error	High Syst. Error
0.288	0.0738	0.178	0.147	0.0937
0.395	0.245	0.223	0.18	0.0671
0.49	0.409	0.243	0.159	0.0488
0.658	0.334	0.178	0.187	0.0681

 Table 4.10: A_{FB} as a function of $-t$ in the mass range [2 GeV – 3 GeV] (see 4.12).

$M(GeV)$	BSA	Stat. error	Low Syst. Error	High Syst. Error
1.59	0.205	0.0551	0.0265	0.0107
1.83	0.296	0.0587	0.0431	0.0251
2.18	0.192	0.0878	0.0408	0.0634
2.64	0.333	0.244	0.0505	0.0927

 Table 4.11: BSA as a function of M (see Figure 4.13).

$-t(GeV^2)$	<i>BSA</i>	Stat. error	Low Syst. Error	High Syst. Error
0.206	0.166	0.0838	0.055	0.0243
0.295	0.31	0.0725	0.0433	0.0206
0.404	0.306	0.0656	0.0401	0.0206
0.607	0.177	0.0647	0.015	0.0217

 Table 4.12: BSA as a function of $-t$ (see Figure 4.14).

ξ	<i>BSA</i>	Stat. error	Low Syst. Error	High Syst. Error
0.101	0.274	0.06	0.0359	0.0279
0.134	0.237	0.0713	0.0341	0.0222
0.197	0.214	0.058	0.0324	0.0183

 Table 4.13: BSA as a function of ξ (see Figure 4.15).

4.4 Comparison with CLAS results

In the exploratory study performed on CLAS data, a first extraction of the cross-section ratio R' was performed. Because of the low energy of the electron beam delivered to CLAS (5.48 GeV), the lepton invariant mass range was limited between 0 and 2 GeV. The analysis was thus performed in the mass region above the $\phi(1020)$, and the ratio was extracted as a function of the squared proton transferred momentum [3].

For the comparison presented in this section, the same kinematic region as for the CLAS analysis is selected from the CLAS12 dataset:

- $2 \text{ GeV} < E_\gamma < 5 \text{ GeV}$
- $0.15 \text{ GeV}^2 < -t < 0.8 \text{ GeV}^2$
- $1.1 \text{ GeV} < M < 1.7 \text{ GeV}$

and the R' ratio was extracted and compared to the CLAS results. The CLAS12 dataset allows to have a thinner binning and to keep statistical error bars well below the CLAS ones. The ratio R' is calculated for eight $-t$ bins and the size of each bin is indicated by the horizontal error bars in the plot bellow. Note that the CLAS12 points are obtained without performing a complete acceptance calculation; the lowest-mass and lowest-energy bin of the acceptance presented in Section 3.4 is used.

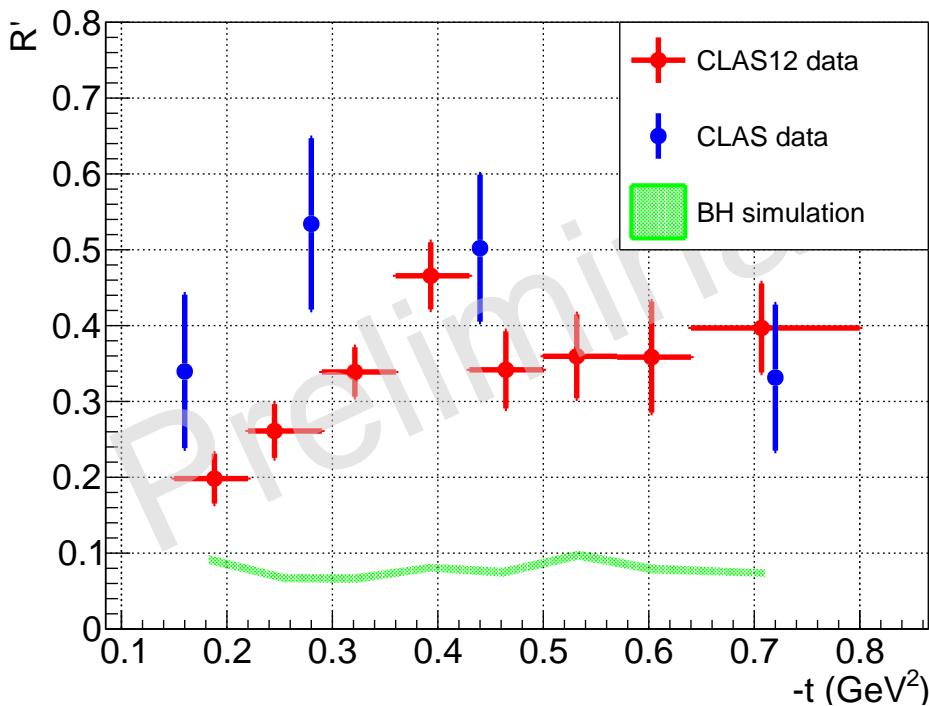


Figure 4.16: CLAS12 results for the R' ratio as a function of $-t$ in the kinematic region accessible by CLAS, together with the CLAS results obtained in [3] and BH-only simulations for comparison. The error bars correspond to the statistical error only.

Figure 4.16 shows the data point from CLAS (blue), CLAS12 (red) as well as the R' ratio calculated from BH-only simulation events with kinematics inside the CLAS12 acceptance (green). Both CLAS and CLAS12 datasets give roughly compatible results, given the error bars, indicating an asymmetry well above the one obtained from BH simulations. The difference between the CLAS and CLAS12 results can be explained by the fact that the ratio is calculated within the respective detector acceptance. Although in this mass region the energy scale might be too low to ensure factorization and vector-meson resonances might be too important to extract any information on TCS, both analyses have coherent results, indicating that the extraction method is under control.

4.5 Comparison Data/Models and physical interpretations

The TCS reaction is of great interest as it allows to measure the D-term via the the R' ratio and the A_{FB} asymmetry, both sensitive to the real part of the \mathcal{H} CFF, as well to verify the universality of GPDs by extracting the photon polarization asymmetry (or BSA), which is sensitive to the imaginary part of \mathcal{H} . In this section, the data points obtained for the A_{FB} and the BSA are compared to model predictions provided by M.Vanderhaeghen using the VGG model [21, 22, 13] and by P.Sznajder using the GK model [23] in the *PARTONS* software [16].

BSA interpretation

Most of the data used to constrain GPDs have been measured in DVCS and Deeply Virtual Meson Production (DVMP). However, DVMP and DVCS data are difficult to compare directly as the former reaction involves Meson Distribution Amplitudes that must be measured using other reactions before being able to interpret DVMP in terms of GPDs. Contrary to DVMP, TCS does not involve distribution amplitudes and is only parameterized by GPDs, making it directly comparable to DVCS. The comparison between these two processes is an important test of the universality of the GPDs. This can be done using the TCS BSA, as it is directly sensitive to the imaginary part of the CFF \mathcal{H} which is itself well constrained by DVCS data.

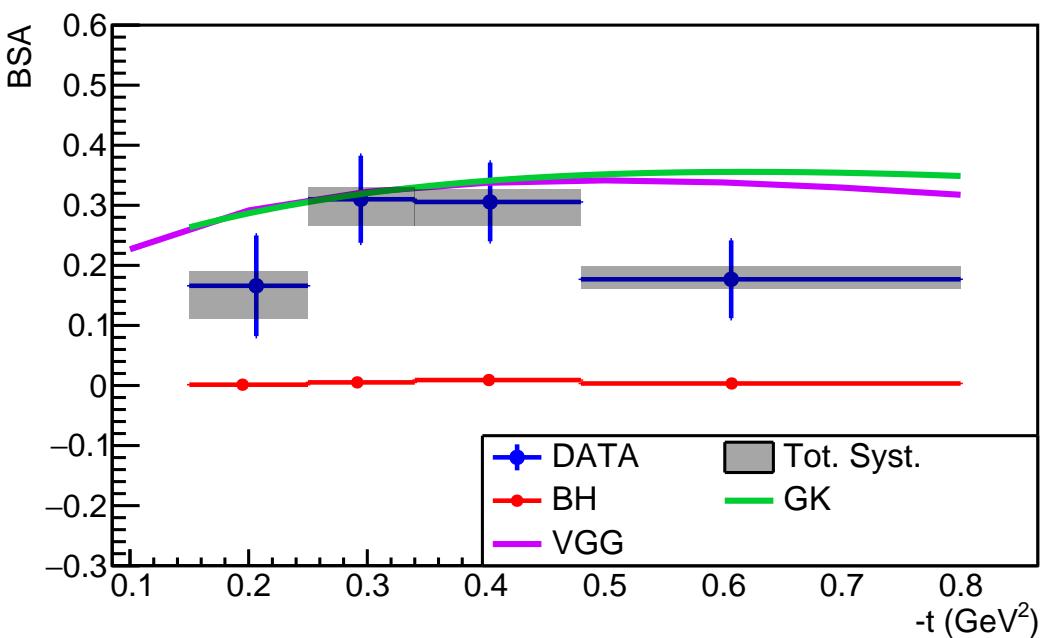


Figure 4.17: CLAS12 data points for the TCS BSA as a function of $-t$, evaluated at $\phi = 90^\circ$, integrated over CLAS12 acceptance and over all the other variables. The vertical blue error bars are statistical uncertainties while the grey bands correspond to systematic uncertainties. Three model predictions, obtained using the VGG and GK models, are also displayed. The model predictions are calculated at the mean kinematic point given above the plot. The red points are the expected values of the BSA for BH-only events, obtained using BH-weighted simulations.

Figure 4.17 shows the TCS BSA extracted from the CLAS12 data as a function of $-t$, compared to three theoretical predictions. The two VGG curves (cyan and magenta) display the $-t$ dependence of the BSA evaluated at $\phi = 90^\circ$ and for θ integrated from $\pi/4$ to $3\pi/4$. The other variables, E_γ and M , are set to 7 GeV and 1.8 GeV respectively. The BSA is calculated for two different values of the sea skewness parameter. The hypothesis $b_{sea} = 1$ (cyan) is the default value of the VGG model, while the $b_{sea} = 5$ hypothesis (magenta) seems to be favored by the analysis of DVCS data in [24]. The GK prediction (orange) is evaluated at the mean kinematic point of the plot, and the angular kinematics

and integration are identical to those of VGG.

The values of the BSA extracted from the CLAS12 data are in agreement, within error bars, with the three theoretical predictions. This observation tends to validate the use of the GPD formalism to describe TCS data and is a hint for the universality of the GPDs, as the VGG model also describes well the DVCS data [24]. However, our data points do not strongly favor any of the VGG hypothesis. Further studies on the dependence of the TCS BSA with the b_{sea} parameter should be made in order to identify the kinematic regions where one could discern between both values.

Figure 4.18 shows the measured BSA as a function of the invariant mass of the lepton pair, M . The prediction obtained with the GK model, for the mean kinematic point specified above the plot and for angular dependencies identical to the ones in Figure 4.17, is superimposed (orange line). The GK curve is only displayed in the mass range between 1.5 GeV to 2.3 GeV, as values of the mass higher than 2.3 GeV are not kinematically allowed at the mean kinematic point of this plot. Nevertheless, the data points are in agreement, within error bars, with the theoretical prediction in the [1.5 GeV – 2.3 GeV] mass range. This is an indication that the GPD factorization seems to apply to TCS in a large part of the mass range accessible by CLAS12.

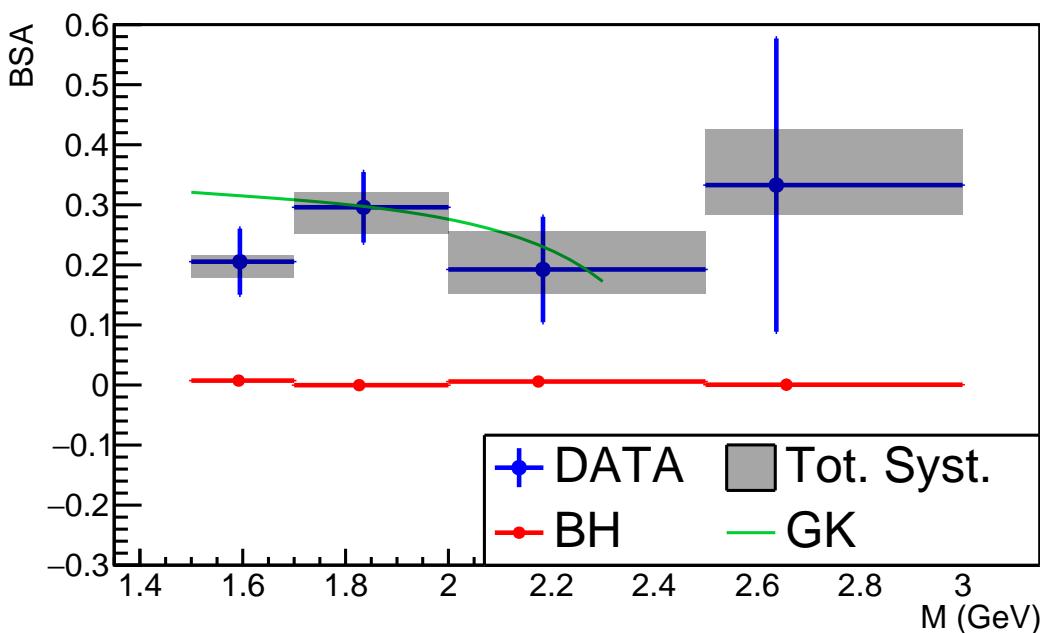


Figure 4.18: CLAS12 data points for the BSA as a function of M . The blue error bars are statistical uncertainties while the grey bands represent the systematic errors. The orange curve is the GK model prediction at the mean kinematic point of the data. The red points are the expected values for BH only (from simulations).

A_{FB} interpretation

The A_{FB} asymmetry has the advantage that it can be easily compared to theory, as it does not involve acceptance limits. Also, as shown in Section 3.7, it has a large sensitivity to the D-term, making it a valuable observable to extract this quantity.

The t -dependence of the A_{FB} extracted from data is compared to theoretical predictions in three cases. In all three cases the GK predictions are calculated for the average E_γ and the invariant mass of the lepton pair given above each plot. For the VGG predictions the mean kinematics are given in the corresponding figure captions.

In Figure 4.19 the A_{FB} asymmetry data points are plotted against $-t$, and all the other variables are integrated over the phase space detailed in Section 3.11. They are compared to predictions obtained

1489 using the VGG model for $M = 1.8$ GeV (red lines) and for different values of θ_0 (at the lower edge
 1490 of the angular bin $\theta_0 = 50^\circ$ (dashed), at its center $\theta_0 = 65^\circ$ (solid), and at its upper edge $\theta_0 = 80^\circ$
 1491 (dash-double-dotted)). The effect of changing the average mass is also illustrated with the green curve
 1492 calculated for $M = 1.5$ GeV. The cyan curve is calculated without the contribution of the D-term to
 1493 the GPD H (the D-term contribution used for this calculation is the one described in [25]). Finally the
 1494 GK prediction (orange) shown here is obtained by integrating the BH-TCS cross section in the forward
 1495 and backward angular bin defined in Section 3.8. Also it has to be noticed that the GK prediction
 1496 does not include the contribution of the D-term.

1497 As already mentionned in Section 3.7, the VGG curves produced with and without the D-term
 1498 clearly indicate that the D-term contribution to the GPD H has a large effect on the value of FB
 1499 asymmetry. The data points are better described by the VGG model when the D-term is included,
 1500 although error bars are still too large to completely rule out the case without the D-term. The GK
 1501 model prediction seems to largely underestimate the asymmetry. This could be explained by the
 1502 absence of D-term in this prediction. In order to ensure that this interpretation is valid for the whole
 1503 mass range studied, the same comparison is done in the low-mass region [1.5 GeV – 2 GeV] and in
 1504 the high-mass region [2 GeV – 3 GeV]. Indeed low mass events dominate when the asymmetry is
 1505 integrated over the full mass range [1.5 GeV – 3 GeV], therefore low-mass vector-meson resonances
 1506 (e.g. $\rho(1450)$) could be the origin of the observed FB asymmetry.

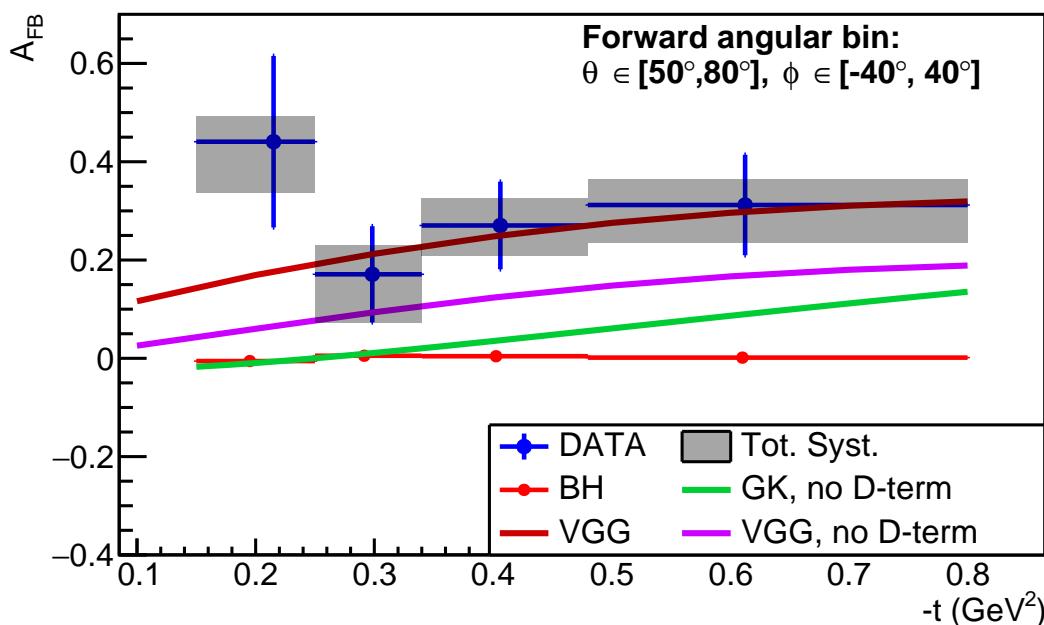


Figure 4.19: CLAS12 A_{FB} as a function of $-t$, integrated over all other variables. The data points are compared with theory predictions realized using the VGG and GK models. The cyan line is calculated with VGG and without the D-term contribution to H . The red lines correspond to three different θ_0 values for the forward direction, using the VGG model. The green line is the prediction of the VGG model for a lower invariant mass value (1.5 GeV). All the VGG curves are calculated at $E_\gamma = 7.0$ GeV, and at the invariant masses indicated in the legend. The orange line is the GK model prediction obtained with *PARTONS*, at the mean kinematic point of the plot, and integrated in the same angular range as the experimental A_{FB} . The red points are the expected values for BH only (from simulations).

1507 Figure 4.20 shows the data points extracted in the [1.5 GeV – 2 GeV] low-mass region. They are
 1508 compared with the same VGG predictions as in the full mass range case, as the average mass and the
 1509 average photon energy do not change dramatically. The GK prediction is however recalculated at the
 1510 mean kinematic point of this plot. The data points do not change substantially when the mass range
 1511 is restricted to the [1.5 GeV – 2 GeV] range as the events included in this analysis mostly have a low

invariant mass, as seen in the mass distribution in Figure 3.7c. Therefore the conclusions drawn for Figure 4.19 also apply to Figure 4.20. Again the data points are better described when the D-term contribution is taken into account.

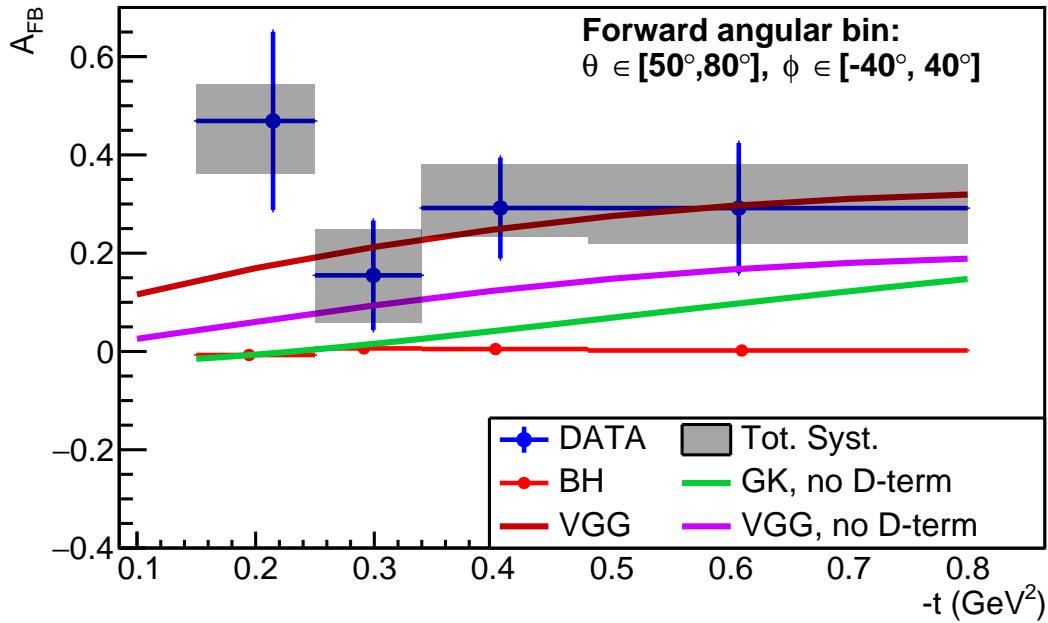


Figure 4.20: CLAS12 A_{FB} as a function of $-t$, integrated over all the other variables, with the invariant mass of the lepton pair integrated in the range [1.5 GeV – 2 GeV]. The VGG curves are the same as in Figure 4.19. The GK prediction is calculated at the mean kinematic point. The red points are the expected values for BH only (from simulations).

The A_{FB} measured in the high-mass region [2 GeV – 3 GeV] is shown in Figure 4.21. The data points are compared with predictions for the FB asymmetry calculated with the VGG model at the center of the angular bin ($\theta_0 = 65^\circ$) and for two mass and photon-energy hypotheses. The prediction obtained when neglecting the D-term is displayed (cyan). The GK model prediction (orange) is also shown. While the effect of changing the kinematic point has little effect on the predicted asymmetry, the D-term plays again a very important role in the value of the asymmetry. As in the case of the full invariant mass integration, the data points tend to indicate that the D-term contribution to the asymmetry is necessary to explain its value. Indeed, both the VGG curve without the D-term and the GK curve underestimate the value of the asymmetry. The error bars shown in this plot do not allow for further conclusions. However, by increasing the available statistics, one could fit directly the D-term contribution and extract the pressure distribution inside the proton from this observable.

The mass and the photon-energy dependences of the extracted A_{FB} data points are also compared to GK model predictions. In both cases the GK model is evaluated at the mean kinematic point of the plots and integrated over the experimental forward and backward bins. Figure 4.22 shows the CLAS12 A_{FB} as a function of M . The GK prediction is only plotted in the [1.5 GeV – 2.3 GeV], as higher mass are kinematically forbidden at the mean kinematic point used to calculate the curve. A prediction for the GK model with a slightly different mean $-t$ ($-t = 0.33 \text{ GeV}^2$) is also shown. The data points are not well reproduced by the GK model in the [1.5 GeV – 2.3 GeV] mass range. This discrepancy could originate from a possible vector-meson contamination, although the conclusions drawn from Figure 4.19 seem to indicate that the absence of the D-term in the model is the reason why the data and the GK model do not agree.

Figure 4.23 displays the A_{FB} data points as a function of the photon energy, E_γ . The GK prediction for this observable is also shown. As already observed in Figures 4.19 and 4.22, the GK predictions largely underestimate the measured FB asymmetry.

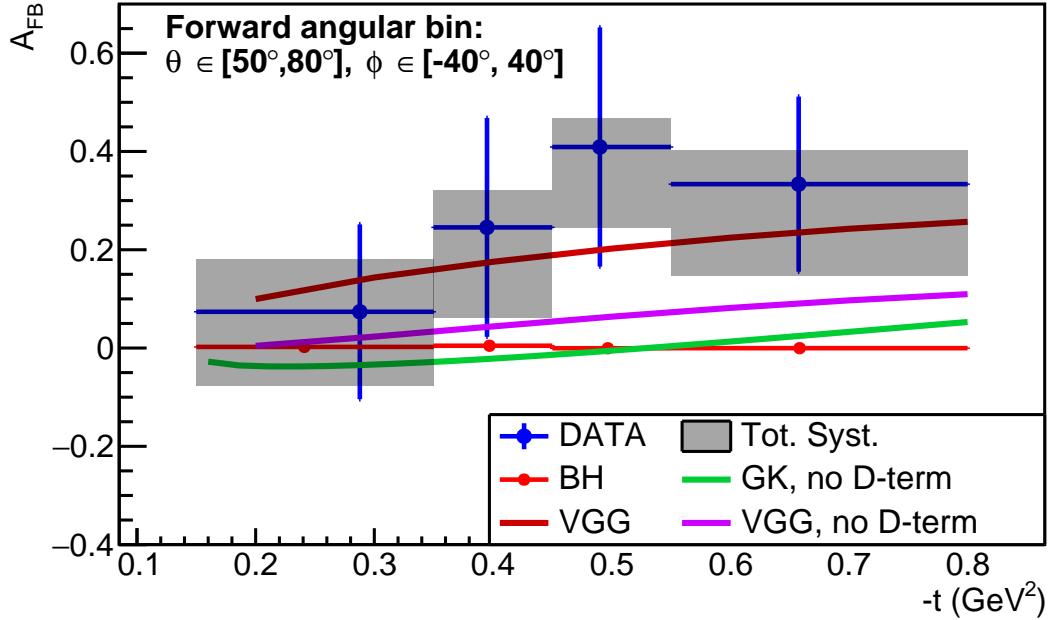


Figure 4.21: CLAS12 A_{FB} as a function of $-t$, integrated over all the other variables and with the invariant mass integrated in the range [2 GeV – 3 GeV]. The VGG model cyan curve does not take into account the D-term contribution. The plain red curve is calculated using VGG with the invariant mass set to 2.2 GeV and $E_\gamma = 7.88$ GeV while for the dot-dashed curve M is set to 2 GeV and $E_\gamma = 7.0$ GeV. The orange line is the GK model prediction at the mean kinematic point. The red points are the expected value for BH only (from simulations).

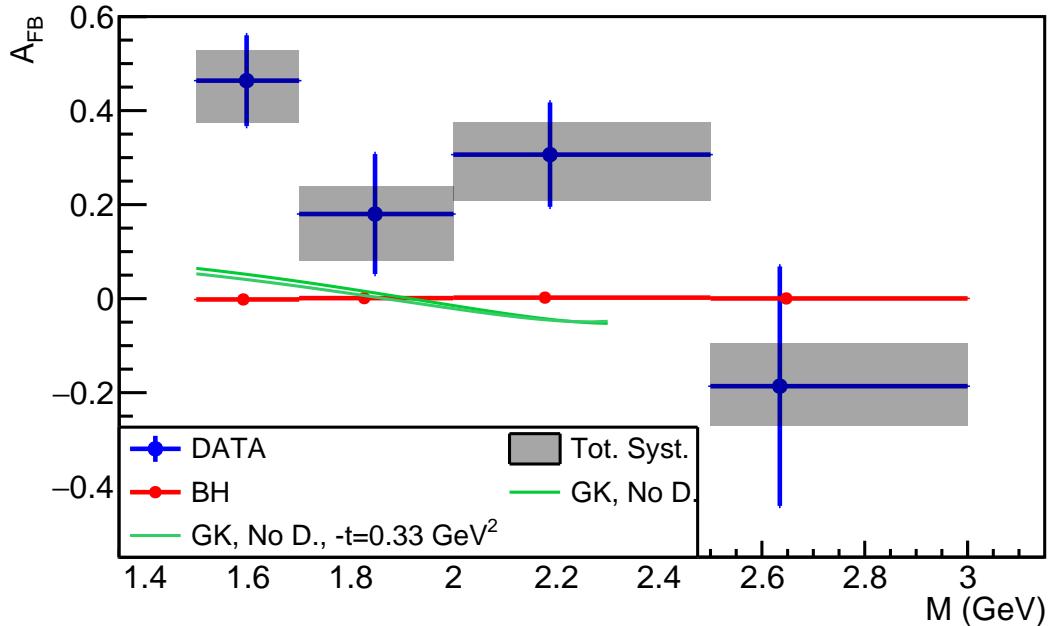


Figure 4.22: CLAS12 A_{FB} as a function of M , integrated over all the other variables. The vertical blue error bars are statistical errors and the grey bands are systematic uncertainties. The orange and brown curves are two predictions obtained with the GK model, for $-t = 0.37 \text{ GeV}^2$ and $-t = 0.33 \text{ GeV}^2$, respectively. The red points are the expected values for BH only (from simulations).

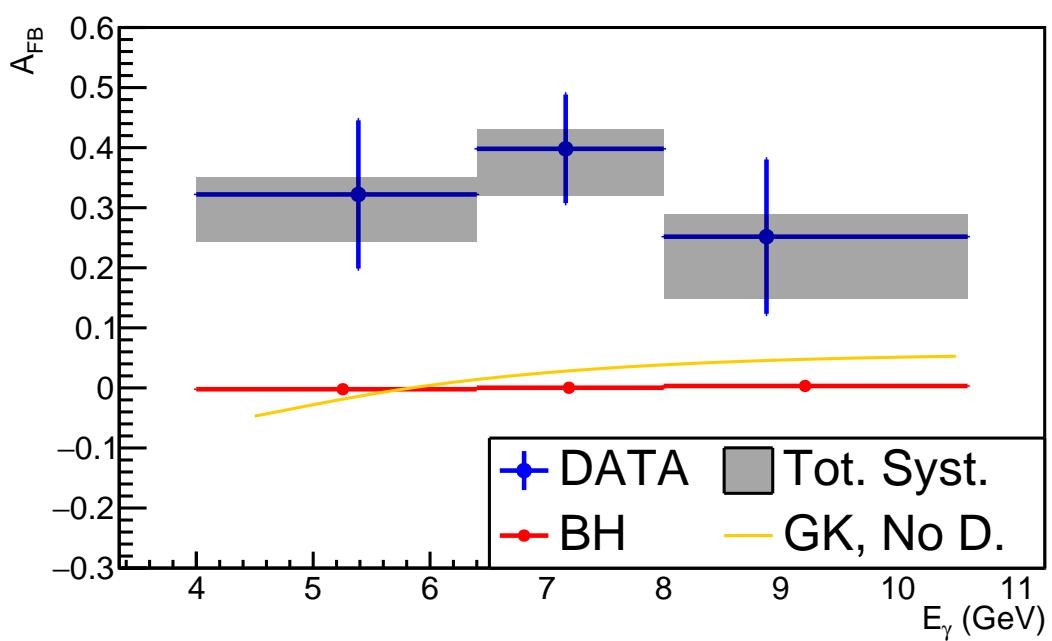


Figure 4.23: CLAS12 A_{FB} as a function of E_γ , integrated over all the other variables. The vertical blue error bars are statistical errors and the grey bands are systematic uncertainties. The orange curve is obtained with the GK model. The red points are the expected value for BH only (from simulations).

Appendices

1539

¹⁵⁴⁰ **Appendix A**

¹⁵⁴¹ **Derivation of the background/signal ratio**

¹⁵⁴² Let $B(x)$ and $S(x)$ respectively be the number of background (mis-id. pions) and signal (true positron)
¹⁵⁴³ events in the TCS sample for a background strength x ($x \in [0, 1]$).

¹⁵⁴⁴ The number of background events in the TCS sample is linear with the background strength x :

$$B(x) = \beta x, \quad (\text{A.1})$$

¹⁵⁴⁵ where β is the number of background events in the TCS sample when no cut is applied.

¹⁵⁴⁶ The function $S(x)$ is unknown but we assume it is increasing with x (when background is removed,
¹⁵⁴⁷ signal events might also be removed by mistake), and does not vary much with x (signal events should
¹⁵⁴⁸ not be removed by the classifier).

¹⁵⁴⁹ The normalized number of TCS events can then be described in the linear region with the following
¹⁵⁵⁰ expression:

$$y(x) = \frac{S(x) + \beta x}{S(1) + \beta}. \quad (\text{A.2})$$

¹⁵⁵¹ Although this formula is only applicable in the linear region, we can extrapolate it to $x = 0$:

$$y(0) = \frac{S(0)}{S(1) + \beta}. \quad (\text{A.3})$$

¹⁵⁵² Assume the chosen cut yields to a normalized background strength x_0 in the linear region. The
¹⁵⁵³ corresponding normalized number of TCS events is:

$$y(x_0) = \frac{S(x_0) + \beta x_0}{S(1) + \beta}. \quad (\text{A.4})$$

¹⁵⁵⁴ We want to estimate the background/signal ratio for a given normalized background strength x_0 ,
¹⁵⁵⁵ $\frac{B(x_0)}{S(x_0)}$. Solving $B(x_0)$ from Equations (A.1) and (A.4) one can write for the B/S ratio at x_0 :

$$\frac{B(x_0)}{S(x_0)} = y(x_0) \frac{S(1) + \beta}{S(x_0)} - 1 = \frac{y(x_0)}{y(0)} \frac{S(0)}{S(x_0)} - 1. \quad (\text{A.5})$$

¹⁵⁵⁶ We can estimate the ratio $\frac{S(x_0)}{S(0)}$ using simulations. It is most of the time very slightly bigger than 1 as
¹⁵⁵⁷ seen in Figure 2.20 of Chapter 2. Therefore one finds:

$$\frac{y(x_0)}{y(0)} - 1 \simeq \frac{B(x_0)}{S(x_0)}, \quad (\text{A.6})$$

¹⁵⁵⁸ and the quantity $\frac{y(x_0)}{y(0)} - 1$ gives a good estimate of the ratio $\frac{B(x_0)}{S(x_0)}$.

₁₅₅₉ Appendix B

₁₅₆₀ Generator checks: comparison between
₁₅₆₁ *GRAPE* and *TCSGen*

₁₅₆₂ Bethe-Heitler events have been generated using *TCSGen* and *GRAPE* within the following phase
₁₅₆₃ space:

- ₁₅₆₄ • Lepton momenta bigger than 1 GeV
- ₁₅₆₅ • Lepton polar angle in the lab in CLAS12 acceptance $5^\circ < \theta_{\text{Lab e}^{+/-}} < 45^\circ$
- ₁₅₆₆ • $0.01 \text{ GeV}^2 < -t < 1 \text{ GeV}^2$
- ₁₅₆₇ • $1.7 \text{ GeV} < M < 3.08 \text{ GeV}$
- ₁₅₆₈ • $2 \text{ GeV} < E_\gamma < 10.6 \text{ GeV}$
- ₁₅₆₉ • $40^\circ < \theta_{\text{COM}} < 140^\circ$,

₁₅₇₀ and the results of both generators have been compared.

₁₅₇₁ The following plots in Figures B.1 to B.5 show the distributions of all five relevant TCS variables,
₁₅₇₂ obtained for both generators as well as their ratio. Each distribution is normalized by the integral of
₁₅₇₃ the E_γ distribution obtained for each generator.

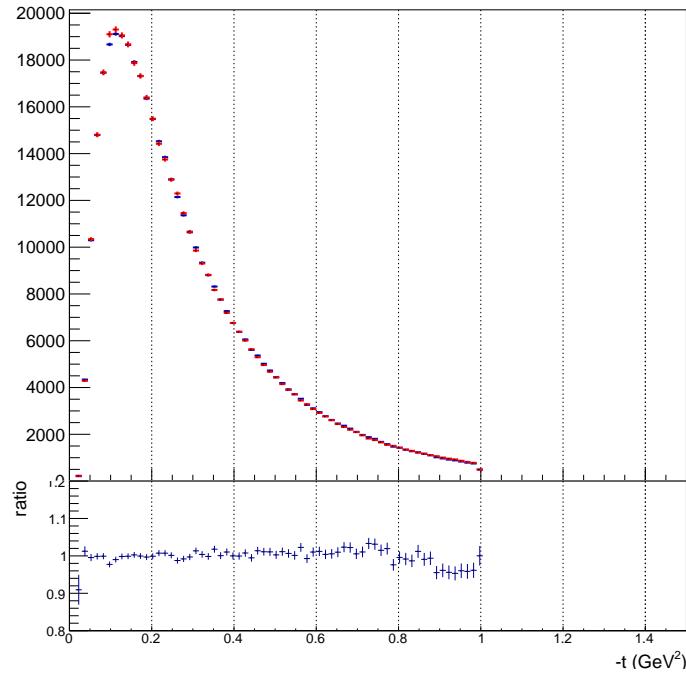


Figure B.1: Comparison of the generated proton momentum transfer squared $-t$ distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

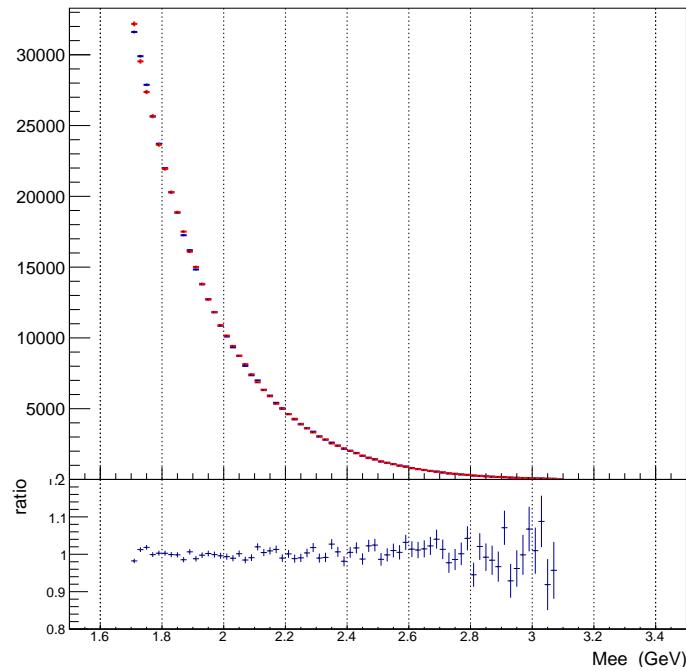


Figure B.2: Comparison of the generated invariant mass of the lepton pair M distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

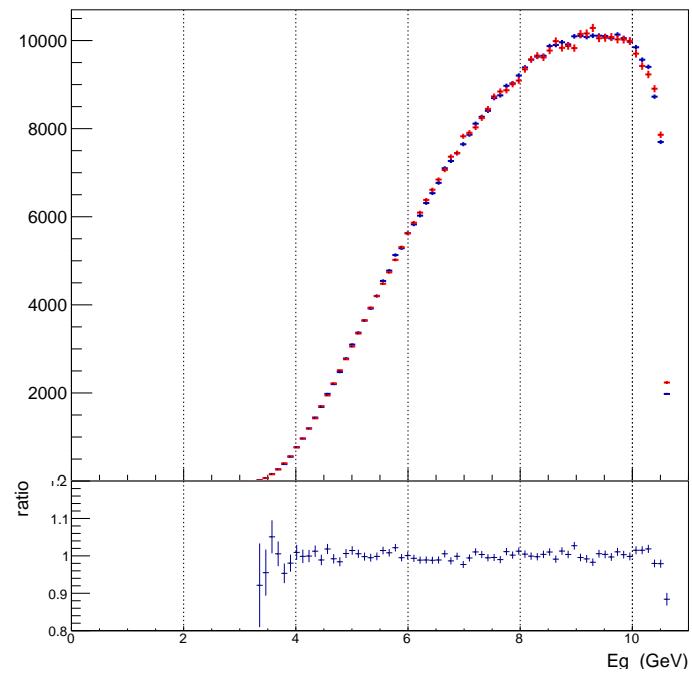


Figure B.3: Comparison of the generated photon energy E_γ distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

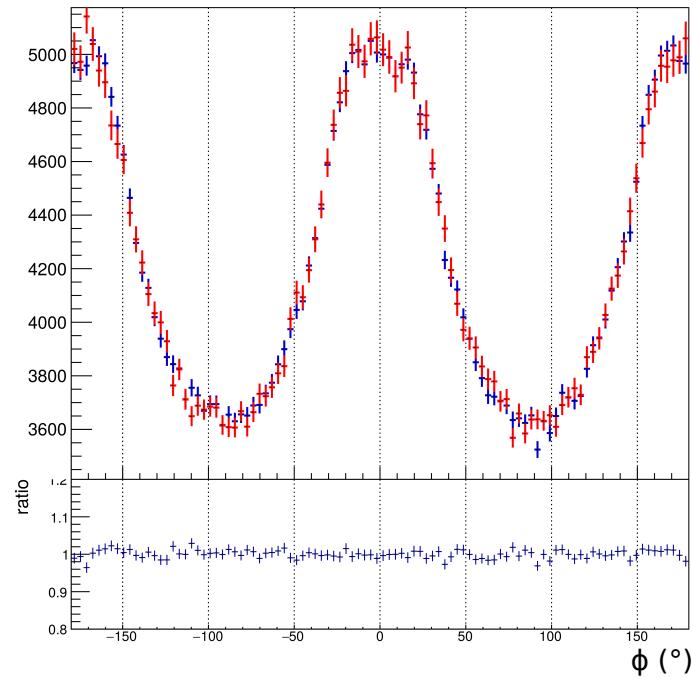


Figure B.4: Comparison of the generated COM azimuthal angle ϕ distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

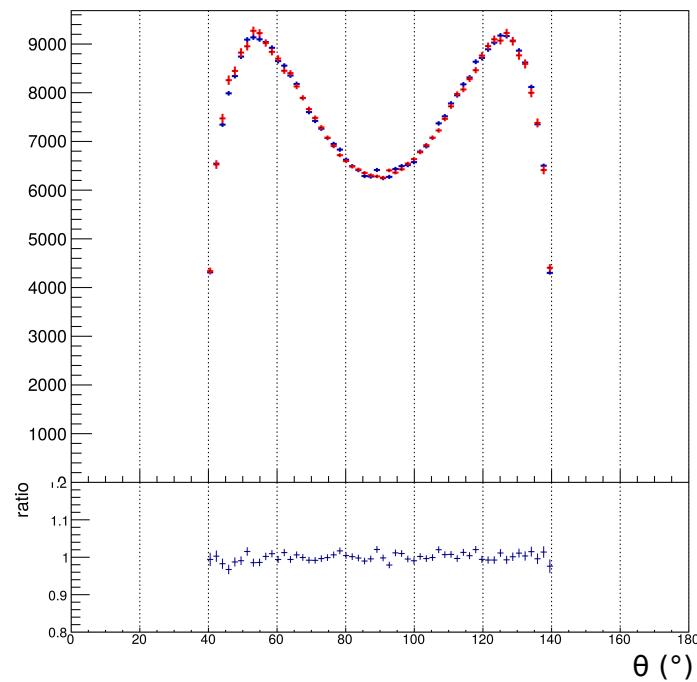


Figure B.5: Comparison of the generated COM polar angle θ distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

¹⁵⁷⁴ **Appendix C**

¹⁵⁷⁵ **Final state particle kinematics**

¹⁵⁷⁶ Figures C.1 and C.2 in this appendix show the kinematics of the three final state particles of the TCS
¹⁵⁷⁷ reaction in the CLAS12 data. The top three plots of Figure C.1 show these kinematics for all events
¹⁵⁷⁸ in CLAS12 data in which a proton, an electron and a positron are detected. The three bottom plots
¹⁵⁷⁹ in the same figure, show the kinematic distributions once the exclusivity cuts presented in Section 3.2
¹⁵⁸⁰ are applied. The plots displayed in Figure C.2 show the kinematics of the final state particles for the
¹⁵⁸¹ events selected in the analysis of this manuscript.

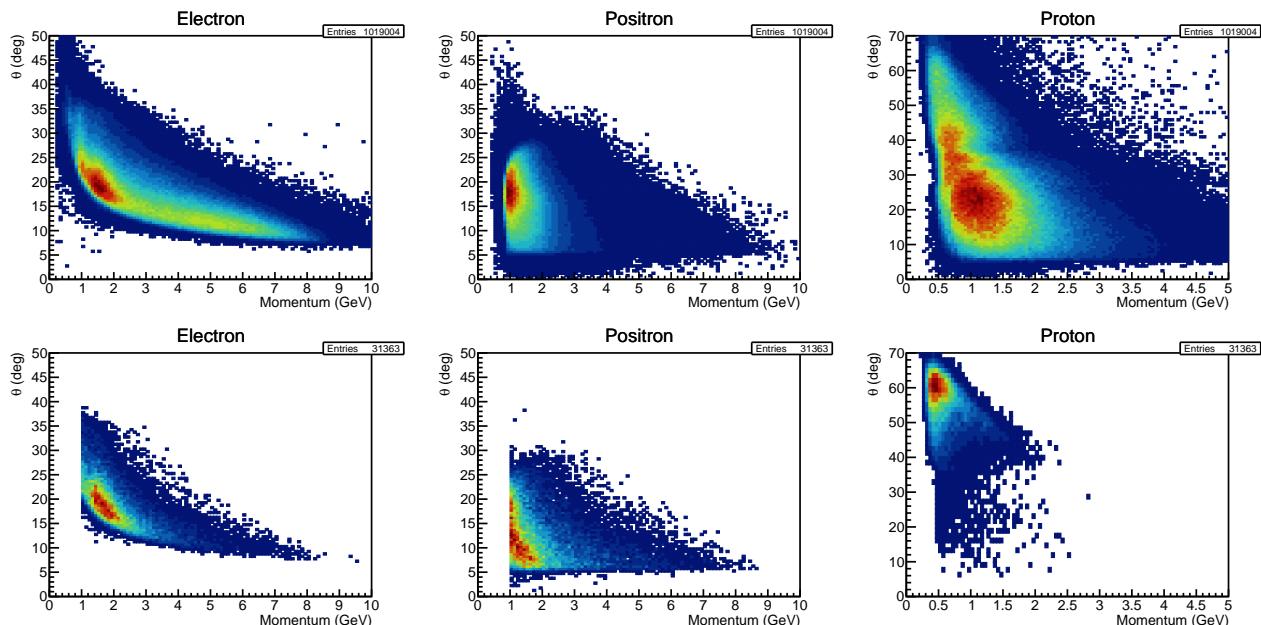


Figure C.1: Top: Kinematics of the TCS final state particles (polar angle in the laboratory frame vs momentum) for events with one proton and two opposite-sign leptons. Bottom: Same plots after the exclusivity cuts to select the photoproduction events are applied.

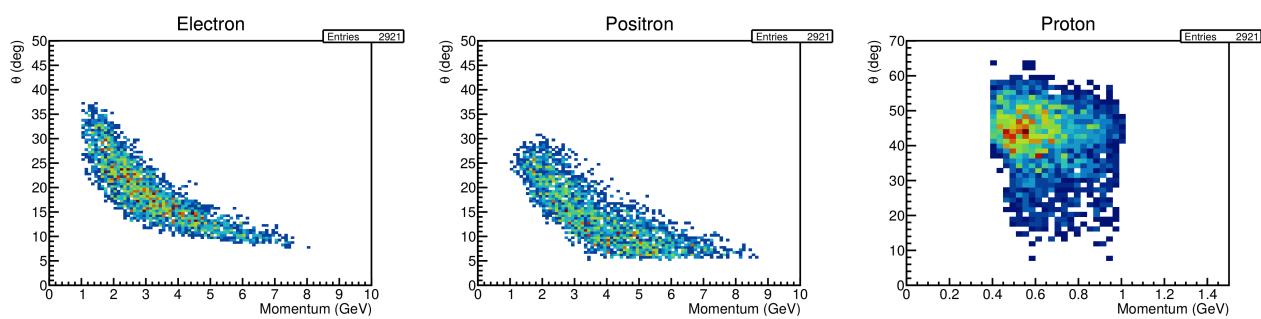


Figure C.2: Kinematics of the final state particles for TCS events with invariant mass between 1.5 and 3 GeV.

¹⁵⁸² Bibliography

- ¹⁵⁸³ [1] E.R. Berger, M. Diehl, and B. Pire. Timelike compton scattering: exclusive photoproduction of
¹⁵⁸⁴ lepton pairs. *The European Physical Journal C - Particles and Fields*, 23(4):675–689, Apr 2002.
- ¹⁵⁸⁵ [2] Boér, M., Guidal, M., and Vanderhaeghen, M. Timelike compton scattering off the proton and
¹⁵⁸⁶ generalized parton distributions. *Eur. Phys. J. A*, 51(8):103, 2015.
- ¹⁵⁸⁷ [3] R. Paremuzyan. *Timelike Compton Scattering*. PhD thesis, Yerevan Physics Institute, 2010.
- ¹⁵⁸⁸ [4] Markus Diehl, Thierry Gousset, Bernard Pire, and John P. Ralston. Testing the handbag contribu-
¹⁵⁸⁹ tion to exclusive virtual compton scattering. *Physics Letters B*, 411(1):193 – 202, 1997.
- ¹⁵⁹⁰ [5] Oleksii Gryniuk and Marc Vanderhaeghen. Accessing the real part of the forward $j/\psi-p$ scattering
¹⁵⁹¹ amplitude from j/ψ photoproduction on protons around threshold. *Phys. Rev. D*, 94:074001, Oct
¹⁵⁹² 2016.
- ¹⁵⁹³ [6] Y.G. Sharabian et al. The clas12 high threshold cherenkov counter. *Nuclear Instruments and*
¹⁵⁹⁴ *Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated*
¹⁵⁹⁵ *Equipment*, page 163824, 2020.
- ¹⁵⁹⁶ [7] Andreas Hoecker et al. Tmva: Toolkit for multivariate data analysis. *PoS*, ACAT:040, 2007.
- ¹⁵⁹⁷ [8] CLAS collaboration. Fiducial studies. https://clasweb.jlab.org/wiki/index.php/Run_Group_A#tab=Fiducial_Studies, 2020.
- ¹⁵⁹⁹ [9] Tetsuo Abe. Grape-dilepton (version 1.1): A generator for dilepton production in ep collisions.
¹⁶⁰⁰ *Computer Physics Communications*, 136(1):126 – 147, 2001.
- ¹⁶⁰¹ [10] Paul Kessler. Photon fluxes and the EPA. In *Two-Photon Physics at LEP and HERA - Status of*
¹⁶⁰² *Data and Models*, pages 183–190, 5 1994.
- ¹⁶⁰³ [11] Eero Byckling and K. Kajantie. *Particle Kinematics*. University of Jyvaskyla, Jyvaskyla, Finland,
¹⁶⁰⁴ 1971.
- ¹⁶⁰⁵ [12] CLAS collaboration. Genev webpage. https://clasweb.jlab.org/wiki/index.php/Event_generator, 2014.
- ¹⁶⁰⁷ [13] M. Guidal, M. V. Polyakov, A. V. Radyushkin, and M. Vanderhaeghen. Nucleon form factors
¹⁶⁰⁸ from generalized parton distributions. *Phys. Rev. D*, 72:054013, Sep 2005.
- ¹⁶⁰⁹ [14] Michel Guidal, Hervé Moutarde, and Marc Vanderhaeghen. Generalized parton distributions
¹⁶¹⁰ in the valence region from deeply virtual compton scattering. *Reports on Progress in Physics*,
¹⁶¹¹ 76(6):066202, 2013.
- ¹⁶¹² [15] Paweł Sznajder. private communication.
- ¹⁶¹³ [16] B. Berthou, D. Binosi, N. Chouika, L. Colaneri, M. Guidal, C. Mezrag, H. Moutarde, J. Rodríguez-
¹⁶¹⁴ Quintero, F. Sabatié, P. Sznajder, and J. Wagner. Partons: Partonic tomography of nucleon
¹⁶¹⁵ software. *The European Physical Journal C*, 78(6):478, Jun 2018.

- 1616 [17] K. Goeke, M.V. Polyakov, and M. Vanderhaeghen. Hard exclusive reactions and the structure of
1617 hadrons. *Progress in Particle and Nuclear Physics*, 47(2):401 – 515, 2001.
- 1618 [18] Raphael Dupre, Michel Guidal, and Marc Vanderhaeghen. Tomographic image of the proton.
1619 *Physical Review D*, 95, 06 2016.
- 1620 [19] Haakon Olsen and L. C. Maximon. Photon and electron polarization in high-energy
1621 bremsstrahlung and pair production with screening. *Phys. Rev.*, 114:887–904, May 1959.
- 1622 [20] Jonathan Dumas. *Feasibility studies of a polarized positron source based on the bremsstrahlung of*
1623 *polarized electrons*. PhD thesis, Université Grenoble Alpes, 2011.
- 1624 [21] M. Vanderhaeghen, P. A. M. Guichon, and M. Guidal. Hard electroproduction of photons and
1625 mesons on the nucleon. *Phys. Rev. Lett.*, 80:5064–5067, Jun 1998.
- 1626 [22] M. Vanderhaeghen, P. A. M. Guichon, and M. Guidal. Deeply virtual electroproduction of photons
1627 and mesons on the nucleon: Leading order amplitudes and power corrections. *Phys. Rev. D*,
1628 60:094017, Oct 1999.
- 1629 [23] S. V. Goloskokov and P. Kroll. An attempt to understand exclusive π^+ electroproduction. *The*
1630 *European Physical Journal C*, 65(1):137, Nov 2009.
- 1631 [24] R. Dupré, M. Guidal, S. Niccolai, and M. Vanderhaeghen. Analysis of deeply virtual compton scat-
1632 tering data at jefferson lab and proton tomography. *The European Physical Journal A*, 53(8):171,
1633 Aug 2017.
- 1634 [25] B. Pasquini, M.V. Polyakov, and M. Vanderhaeghen. Dispersive evaluation of the d-term form
1635 factor in deeply virtual compton scattering. *Physics Letters B*, 739:133 – 138, 2014.