

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

<b>1 Physics motivations</b>	<b>7</b>
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
<b>2 Analysis strategy, particle identification and momentum corrections</b>	<b>13</b>
2.1 Analysis strategy for TCS . . . . .	13
2.2 Particle identification . . . . .	14
2.3 Positron identification . . . . .	15
2.3.1 Evidence of $\pi^+$ 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
2.4 Proton momentum corrections . . . . .	31
2.4.1 Monte-Carlo corrections . . . . .	32
2.4.2 Data-driven momentum corrections . . . . .	35
2.5 Lepton momentum corrections . . . . .	37
2.5.1 Monte-Carlo corrections . . . . .	38
2.5.2 Detected radiated photon correction . . . . .	38
2.6 Fiducial cuts . . . . .	39
2.7 Background merging . . . . .	41
2.8 Proton efficiency correction . . . . .	42
2.8.1 Efficiency correction in the central detector . . . . .	43
2.8.2 Efficiency correction in the forward detector . . . . .	43
<b>3 Simulations and extraction of the TCS observables</b>	<b>49</b>
3.1 TCS event generator and simulations . . . . .	49
3.1.1 GRAPE . . . . .	49
3.1.2 TCSGen . . . . .	49
3.2 Event selection . . . . .	50
3.2.1 Final state selection . . . . .	50
3.2.2 Exclusivity cuts . . . . .	50
3.3 Phase space of interest and Simulations/Data comparison . . . . .	52
3.4 Acceptance estimation . . . . .	54

63	3.5	Background estimations . . . . .	60
64	3.5.1	Electro-production of a lepton pair $ep \rightarrow pe^+(e^-)e_{scattered}$ . . . . .	60
65	3.5.2	Pion contamination . . . . .	65
66	3.5.3	$J/\psi$ contamination . . . . .	65
67	3.6	Experimental cross-section ratio . . . . .	65
68	3.7	Phenomenological study of the TCS Forward-Backward asymmetry . . . . .	66
69	3.7.1	Early considerations and comparison with other models . . . . .	67
70	3.7.2	TCS $A_{FB}$ kinematic dependencies . . . . .	69
71	3.7.3	TCS $A_{FB}$ model dependencies . . . . .	70
72	3.8	Experimental Forward-Backward asymmetry . . . . .	71
73	3.9	Experimental beam-spin asymmetry . . . . .	73
74	3.10	Photon polarization transfer . . . . .	74
75	3.11	Binning of the data . . . . .	76
76	3.12	Systematic errors estimation . . . . .	77
77	<b>4</b>	<b>Results and comparison with model predictions</b> . . . . .	<b>83</b>
78	4.1	Complete CLAS12 results for the TCS observables . . . . .	83
79	4.1.1	$R'$ ratio . . . . .	83
80	4.1.2	$A_{FB}$ . . . . .	83
81	4.1.3	BSA . . . . .	84
82	4.2	Tabulated results . . . . .	97
83	4.3	Comparison with CLAS results . . . . .	100
84	4.4	Comparison Data/Models and physical interpretations . . . . .	101
85	<b>Appendices</b>		<b>107</b>
86	<b>A</b>	<b>Derivation of the background/signal ratio</b>	<b>109</b>
87	<b>B</b>	<b>Generator checks: comparison between <i>GRAPE</i> and <i>TCSGen</i></b>	<b>111</b>
88	<b>C</b>	<b>Final state particle kinematics</b>	<b>115</b>
89	<b>Bibliography</b>		<b>118</b>

# <sup>90</sup> Introduction

<sup>91</sup> This note describes the Timelike Compton Scattering analysis using the first CLAS12 data. It is  
<sup>92</sup> divided in four chapters. In the first chapter, the motivations for the measurement of the TCS process  
<sup>93</sup> are briefly presented. In the following chapters the analysis procedures are detailed. The final-state  
<sup>94</sup> particles are identified using the CLAS12 EB described in the RGA analysis note. An enhanced positron  
<sup>95</sup> identification algorithm, especially developed for the TCS measurement, is presented in the second  
<sup>96</sup> chapter of this note. The second step of the analysis consists in correcting the data. These corrections  
<sup>97</sup> aim at matching the measured momenta of the detected particles with their actual momenta at the  
<sup>98</sup> vertex. This step is also presented in Chapter 2 of this note. Once the final state particles are well  
<sup>99</sup> identified and their momenta are corrected, exclusivity cuts are applied in order to make sure that each  
<sup>100</sup> event is a TCS event (i.e. from the  $\gamma p \rightarrow p' e^+ e^-$  reaction). The observables are then computed with  
<sup>101</sup> the kinematic variables of these good events. This step is performed in Chapter 3. Finally in Chapter  
<sup>102</sup> 4 the results of the full analysis of the CLAS12 data are displayed and discussed against theoretical  
<sup>103</sup> predictions.

<sup>104</sup> This note includes sections and subsections which can also be found in the RG-A common analysis  
<sup>105</sup> note. These sections have been developed during this analysis but are of common interest to all  
<sup>106</sup> analysers. In order to keep consistency and clarity, these sections are clearly identified with the  
<sup>107</sup> following markers:

---

<sup>108</sup>  
<sup>109</sup> *Start of common analysis section*  
<sup>110</sup> ...  
<sup>111</sup> *End of common analysis section*  
<sup>112</sup> 

---

<sup>113</sup> **Data set**

<sup>114</sup> **Analysis :** Fall 2018 - Inbending - Skim 1 (TCS/Jpsi)  
<sup>115</sup> Run list : 5032 5036 5038 5039 5040 5041 5043 5045 5046 5047 5051 5052 5053 5116 5117 5119 5120  
<sup>116</sup> 5124 5125 5126 5127 5128 5129 5130 5137 5138 5139 5153 5158 5159 5160 5162 5163 5164 5165 5166  
<sup>117</sup> 5167 5168 5169 5169 5180 5181 5182 5183 5189 5190 5191 5193 5194 5195 5196 5197 5198 5199 5200 5201  
<sup>118</sup> 5202 5203 5204 5205 5206 5208 5211 5212 5215 5216 5219 5220 5221 5222 5223 5225 5229 5230 5231  
<sup>119</sup> 5232 5233 5234 5235 5237 5238 5239 5247 5248 5249 5250 5252 5253 5257 5258 5259 5261 5262 5300  
<sup>120</sup> 5301 5302 5303 5304 5305 5306 5307 5310 5311 5315 5316 5317 5318 5319 5320 5323 5324 5325 5333  
<sup>121</sup> 5334 5335 5336 5339 5340 5341 5342 5343 5344 5345 5346 5347 5349 5351 5354 5355 5356 5357 5358  
<sup>122</sup> 5359 5360 5361 5362 5366 5367 5368 5369 5370 5371 5372 5373 5374 5375 5376 5377 5378 5379 5380  
<sup>123</sup> 5381 5382 5383 5386 5390 5391 5392 5393 5394 5398 5399 5400 5401 5402 5403 5404 5406 5407 5414  
<sup>124</sup> 5415 5416 5417 5418 5419

<sup>125</sup> **Positron PID :** Fall 2018 - Inbending - Skim 1 and Skim 4 (inclusive)

<sup>126</sup> **Simulation settings**

<sup>127</sup> GEMC 4.3.2 (no bg merging) / 4.4.0 (with bg merging)  
<sup>128</sup> CoatJava 6.5.8 (no bg merging) / 6.5.9 (with bg merging)



<sub>129</sub> **Chapter 1**

<sub>130</sub> **Physics motivations**

<sub>131</sub> Generalized Partons Distributions (GPD) have been experimentally studied mainly through Deeply  
<sub>132</sub> Virtual Compton Scattering (DVCS) polarization observables. Such observables are mainly sensitive to  
<sub>133</sub> the imaginary part of the Compton Form Factors (CFF), which are quantities directly related to GPDs.  
<sub>134</sub> To access the real part of CFFs, DVCS doubly polarized beam-target asymmetries or unpolarized cross  
<sub>135</sub> sections are needed. It is also possible to access the real part of CFFs through the time-reversal  
<sub>136</sub> symmetric process of the DVCS: the Timelike Compton Scattering (TCS)[1, 2, 3].

<sub>137</sub> Measuring TCS observables also provides a test for the universality of GPDs. Indeed photon-  
<sub>138</sub> polarization dependent cross section of TCS is sensitive to the imaginary part of CFFs. Comparing  
<sub>139</sub> the results obtained in TCS and DVCS will help proving that GPDs are universal functions and are  
<sub>140</sub> not only related to DVCS. In this chapter, the motivations to measure TCS are presented.

<sub>141</sub> **1.1 Phenomenology of TCS**

<sub>142</sub> TCS is the time-reversal symmetric process of DVCS. The reaction of interest is  $\gamma p \rightarrow \gamma^* p'$ , where  
<sub>143</sub> the incoming photon is real ( $Q^2 = 0$ ) and the outgoing photon is virtual. The virtual photon decays  
<sub>144</sub> in a lepton pair which can be detected. The full reaction is therefore  $\gamma p \rightarrow p' l^+ l^-$ . Note that in the  
<sub>145</sub> following we refer to the  $\gamma p \rightarrow p' l^+ l^-$  reaction directly as the *TCS reaction*. Contrary to DVCS, where  
<sub>146</sub> the large spacelike virtuality of the incoming photon gives a hard scale which ensures factorization, the  
<sub>147</sub> TCS hard scale is given by the timelike virtuality  $Q'^2$  of the outgoing photon. For large  $Q'^2$  such that  
<sub>148</sub>  $\frac{t}{Q'^2} \ll 1$ , factorization can be applied. The real incoming photon scatters off a single quark, which  
<sub>149</sub> emits a virtual photon. The leading order, leading twist diagram for TCS is given in Figure 1.1.

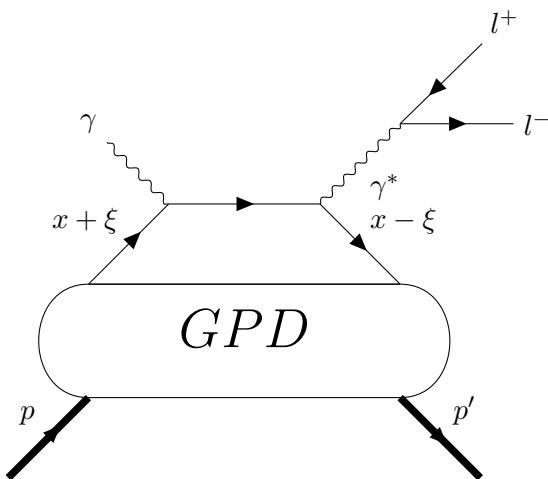


Figure 1.1: The Feynman diagram for TCS in the reaction  $\gamma p \rightarrow p' l^+ l^-$

<sub>150</sub> The relevant variables to study the TCS reaction are shown in Figure 1.2. They are the virtuality  
<sub>151</sub> of the outgoing photon  $Q'^2 = (k + k')^2$ , the transferred momentum to the nucleon  $t = (p' - p)^2$ , the

square of the  $\gamma p$  Center-of-Mass (COM) energy  $s = (p + q)^2$  (or equivalently the real photon energy  $E_\gamma$ ), the azimuthal angle  $\phi$  between the leptonic plane and the hadronic plane, and the angle of the outgoing electron in the lepton COM frame,  $\theta$ . 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.

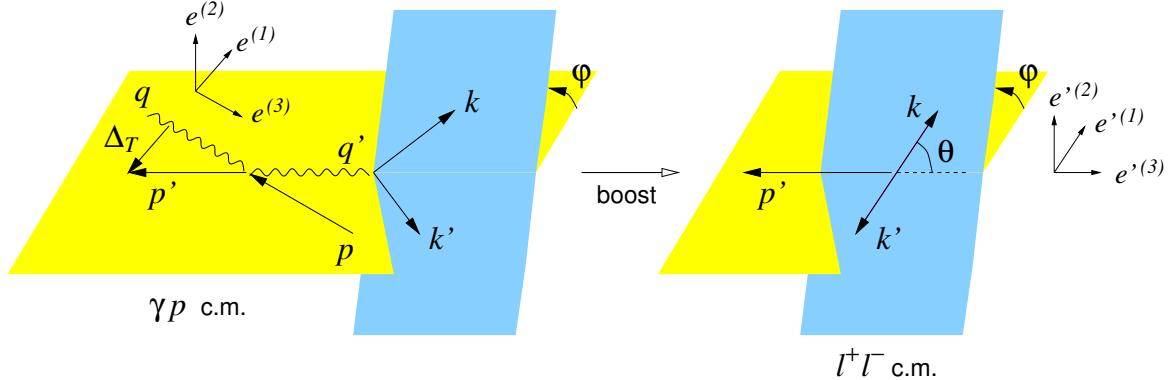


Figure 1.2: Frame definition and relevant variables for TCS, taken from [1]. The yellow plane containing the momenta of the target and recoil protons is called the hadronic plane. The blue plane containing the momenta of the two leptons is the leptonic plane. The angle between these planes is called  $\phi$ . The angle  $\theta$  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.

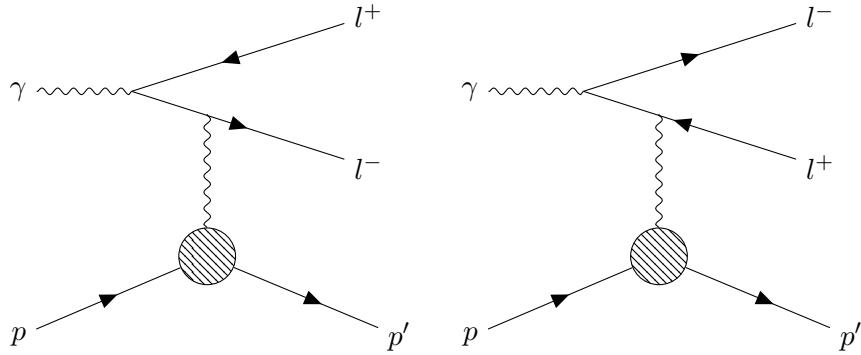


Figure 1.3: Feynman diagrams for the Bethe-Heitler processes that contribute to the  $\gamma p \rightarrow p' l^+ l^-$  reaction.

### 1.1.1 The CFFs of TCS

Similarly to DVCS, the quark loop 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:

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

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

### 162 1.1.2 TCS cross section

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

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

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

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

166 where

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

167 with  $m_l$  the lepton mass and

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

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

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

168 The BH cross section is plotted in Figure 1.4 for different  $\theta$  and  $\phi$ . One can see that the cross  
169 section is largely enhanced around  $\phi = 0^\circ$  for high values of  $\theta$ , and around  $\phi = 180^\circ$  for low values of  
170  $\theta$ .

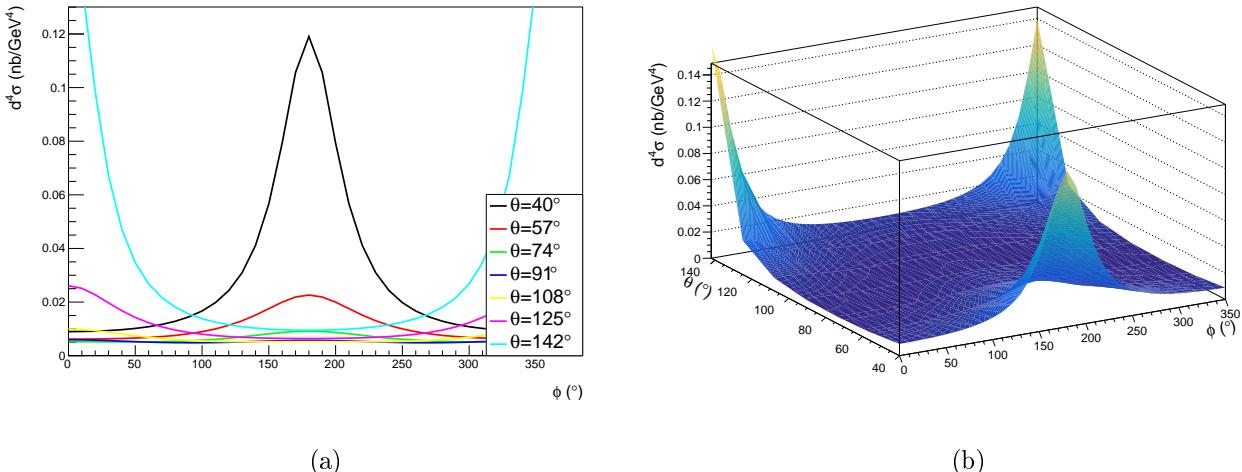


Figure 1.4: BH cross section at the kinematic point  $Q'^2 = 4 \text{ GeV}^2$ ,  $\xi = 0.1$  and  $-t = 0.2 \text{ GeV}^2$ , (a) for different values of  $\theta$  as a function of  $\phi$ , (b) as a function of both  $\theta$  and  $\phi$ .

171 The TCS contribution reads:

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

172 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:

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

173 where

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

174 and the  $\tilde{M}$  terms are CFFs combinations defined in [4]. At leading order and leading twist, the only 175 term contributing to the cross section is the one proportional to  $\tilde{M}^{--}$ . The CFF dependence of  $\tilde{M}^{--}$  176 is given by:

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

177 where

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

178 Furthermore, the dominant term in Equation (1.13) at JLab kinematics is the one containing the CFF 179  $\mathcal{H}$  ( $\tilde{\mathcal{H}}$  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 180 of the interference cross section thus gives direct access to the real part of the CFF  $\mathcal{H}$ .

181 **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 183 to TCS. In the case of transversely polarized photons, the interference cross section is expressed as:

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

184 where  $\nu$  is the circular polarization of the incoming real photon. The additional polarization term 185 exhibits the same CFF content as the unpolarized cross section, except that it now depends on the 186 imaginary parts via  $\sin(n\phi)$  factors. Extracting the  $\sin(\phi)$  component of the polarized cross section 187 enables to access the imaginary part of  $\mathcal{H}$ . This is an important test of the universality of GPDs once 188 compared with DVCS data.

## 189 1.2 TCS observables

190 In this section, the TCS observables extracted in this analysis are presented. These observables are 191 the  $R$  ratio and the Forward/Backward asymmetry sensitive to the real parts of the CFFs, and the 192 photon polarization asymmetry sensitive to the imaginary parts of the CFFs.

193 **1.2.1 R ratio**

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

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

195 where

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

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

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

200 **1.2.2 Forward-Backward asymmetry**

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

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

203 where only the  $\theta$ - $\phi$  dependence of the cross section is explicitly written.

204 The transformation  $(\theta \rightarrow 180^\circ - \theta, \phi \rightarrow \phi + 180^\circ)$  corresponds to inverting the vectors  $k$  and  $k'$  in the  
205 COM frame of the lepton pair. This transformation leaves both  $L$  (Equation (1.8)) and  $L_0$  (Equation  
206 (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  
207 unchanged under this transformation. Assuming leading order and leading twist, only the  $\cos(\phi)\tilde{M}^{--}$   
208 term contributes in Equation (1.11) and the interference cross section is transformed as:

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

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

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

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

214 **1.2.3 Photon polarization asymmetry**

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

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

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

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

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



220 Chapter 2

221 **Analysis strategy, particle identification  
222 and momentum corrections**

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

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

234 **2.1 Analysis strategy for TCS**

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

- 239 • exactly one proton,
- 240 • exactly one electron,
- 241 • exactly one positron,
- 242 • any other particles.

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

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

250 Once the final state particles are well identified and their momenta are corrected, exclusivity cuts  
251 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).  
252 The observables are then computed with the kinematic variables of these good events. This step is  
253 presented in Chapter3.

**254 2.2 Particle identification**

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

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

$$260 \quad t_{tof} = t_{TOF} - t_S, \quad (2.1)$$

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

$$263 \quad t_{track} = \frac{P_L \sqrt{p^2 + m^2}}{pc}, \quad (2.2)$$

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

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

$$260 \quad SF = \frac{E_{dep}}{P}, \quad (2.3)$$

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

- 273 • A track in the DCs and an associated EC shower (the curvature of the track in the torus magnetic  
274 field gives the charge of the particle)
- 275 • A minimum deposited energy in the PCAL:  $E_{PCAL} > 60$  MeV
- 276 • A total measured SF,  $SF_M(E_{dep})$ , within  $5\sigma$  of the parametrized SF (see common analysis note  
277 for more details).
- 278 • a minimum number of HTCC photo-electrons:  $N_{PHE}(HTCC) > 2$ .

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

290

---

291 *Start of common analysis section*

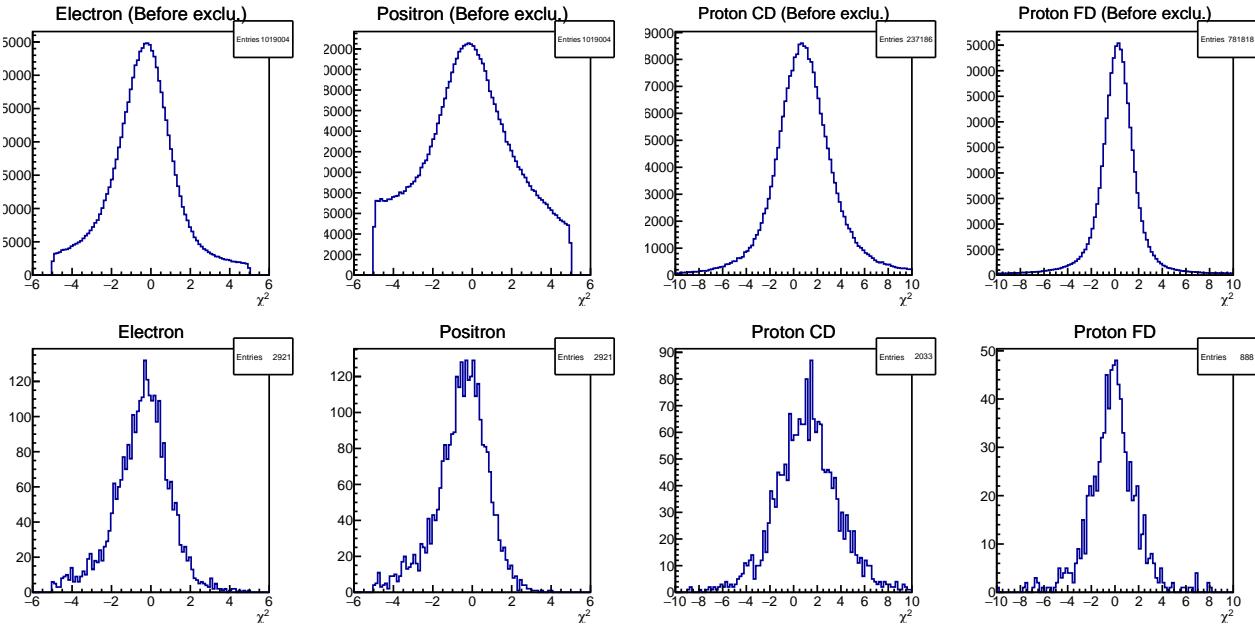


Figure 2.1: Pid  $\chi^2$  for the three final state particles of the TCS reaction. In the proton case, the protons detected in the CD or the FD are shown separately. One can see that the selected events have low  $\chi^2$  ensuring a good pid quality.

## 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  $\pi^+$  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 $\pi^+$ 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  $\pi^+$  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.

#### $\pi^+$ 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.

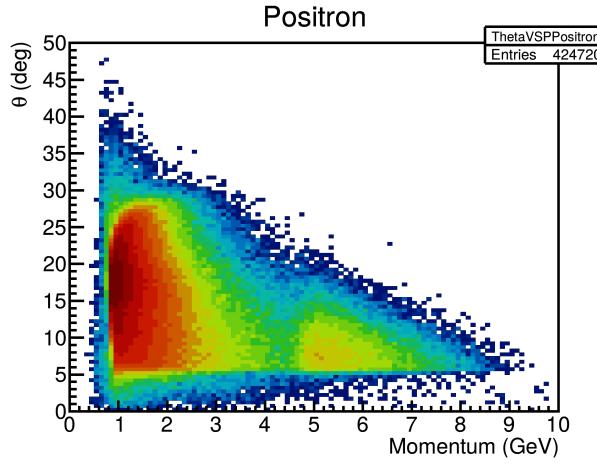


Figure 2.2: Polar angle  $\theta$  versus momentum of positrons for events with one electron, one proton and one positron in CLAS12.

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

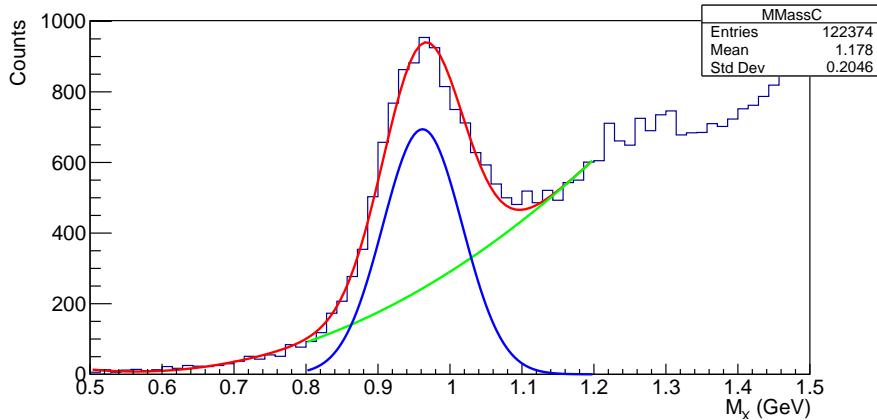


Figure 2.3: Missing mass spectrum of the reaction  $ep \rightarrow ee_{m_\pi}^+ X$  in the neutron mass region.

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

### 329 $\pi^+$ contamination from simulations

330 The  $\pi^+$  contamination is also visible in simulations. Two test samples were generated, one sample  
 331 of positrons and one of positively charged pions. They were passed to the GEMC CLAS12 simulation  
 332 and reconstructed using the standard CLAS12 software. Particles were simulated within the CLAS12

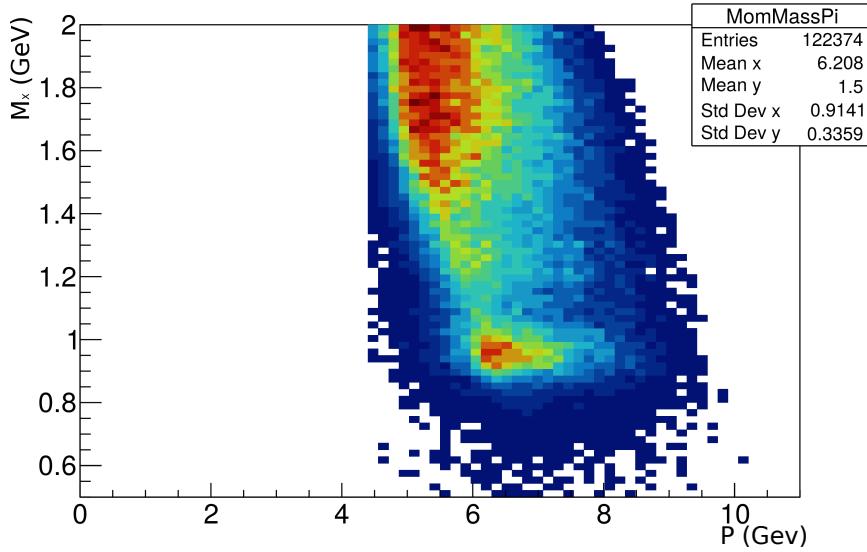


Figure 2.4: Mass of the missing particle versus momentum of the particle with ID -11 .

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.

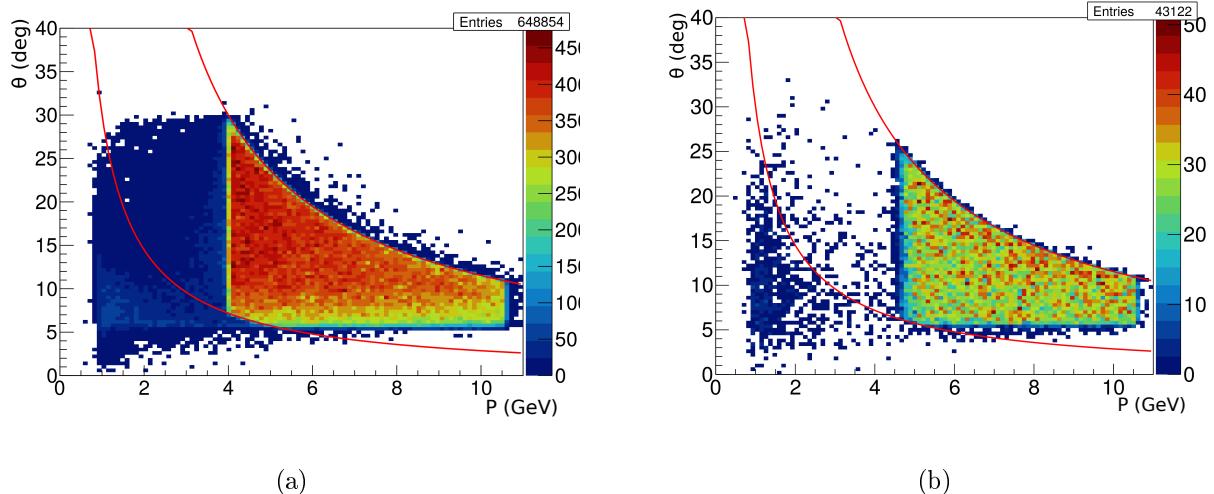


Figure 2.5: (a) Polar angle  $\theta$  versus momentum of the simulated true-positrons (b) Polar angle  $\theta$  versus momentum of the simulated mis-id. pions. The red lines represent the transverse momentum limits ( $0.5 \text{ GeV} < P_t < 2 \text{ GeV}$ ) applied on the generated particles. The 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  $\pi^+$ ) 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.

342  **$\chi^2$  cut**

343 The CLAS12 EB gives the deviation of the total sampling fraction from a parametrized model. This  
 344 quantity is referred as  $\chi^2$  in the CLAS12 software (although the name commonly used in literature is  
 345 *pull* value) and it is defined as:

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

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

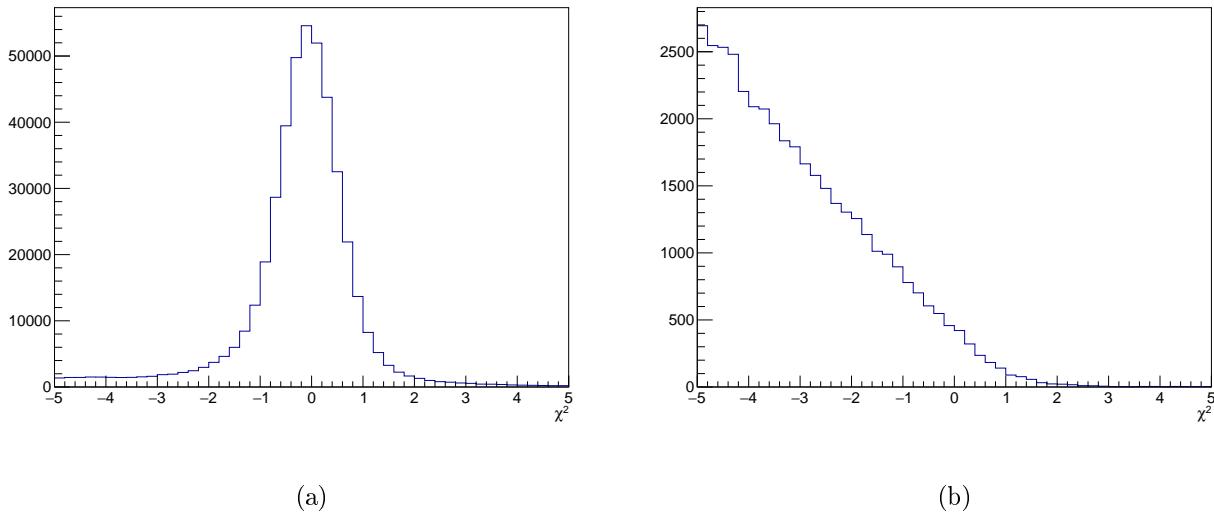


Figure 2.6: Sampling fraction  $\chi^2$  of (a) true-positrons and (b) mis-id. pions.

352 **2D Sampling Fraction cuts**

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

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

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

362 **2.3.3 Multivariate analysis approaches**

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

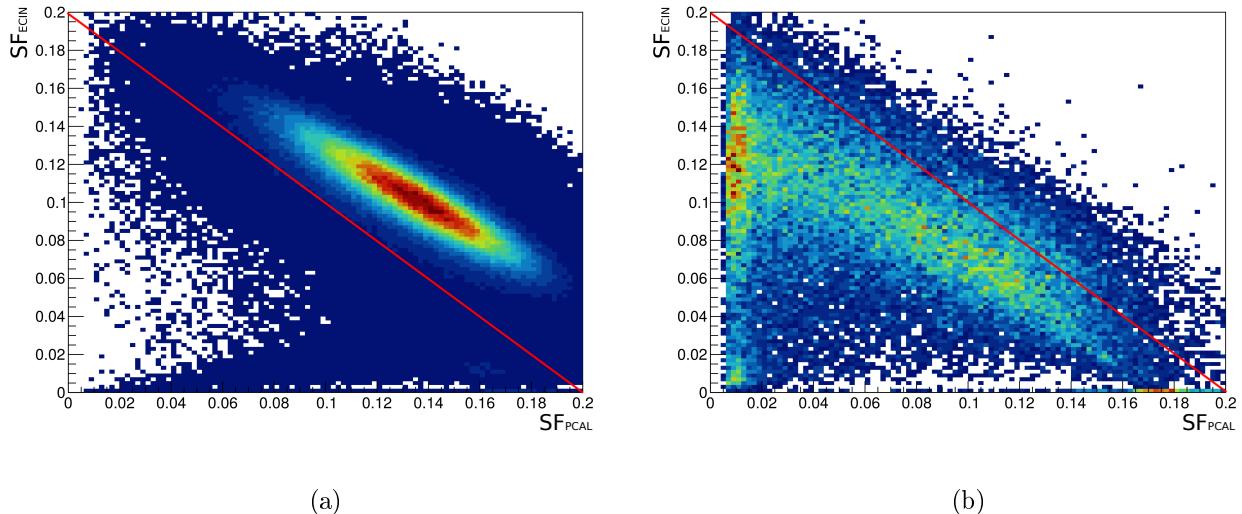


Figure 2.7: SF in the ECIN versus SF in the PCAL for simulated (a) true-positrons and (b) mis-id. pions. A possible cut to remove mis-id. pions is to keep only particles above the anti-diagonal represented by the red lines. Most mis-id. pions are removed while most positrons are kept.

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

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

367 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  $\pi^+$  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  $\pi^+ / e^+$  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).

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

382 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.

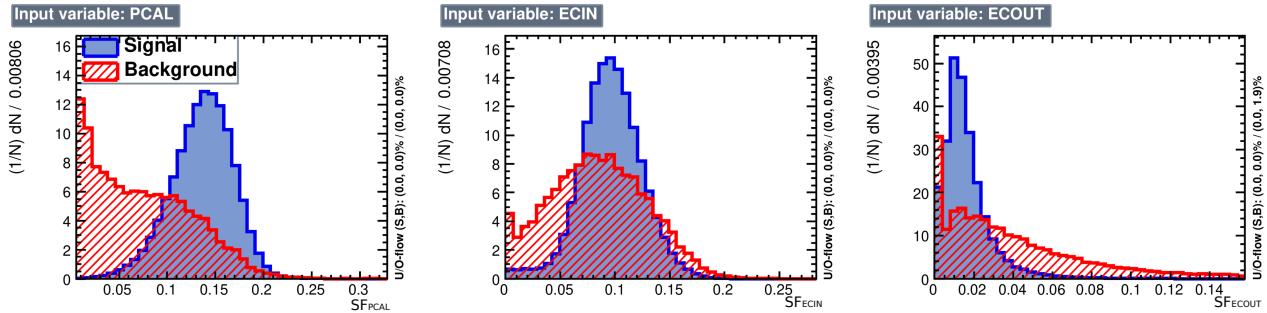


Figure 2.8: Input variables provided for the training of the multivariate classifiers. The leftmost plot shows the  $SF_{PCAL}$  distributions for the signal and the background. The middle plot  $SF_{ECIN}$ , the rightmost  $SF_{ECOUT}$ .

## 387 6D analysis

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

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

391 where  $M_{2/U/PCAL}$  is the square of the width of the shower along the  $U$  direction as defined in Equation  
 392 (2.6). Similar equations apply for the  $V$  and  $W$  directions. The distribution of the input squared shower  
 393 widths are shown in Figure 2.9. Figure 2.10 shows the architecture of the neural network used for the  
 6D analysis.

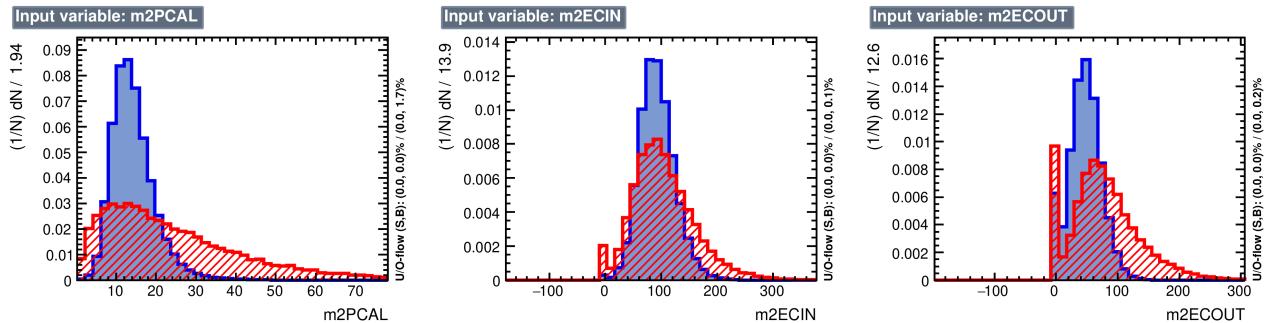


Figure 2.9: Average squared shower width for the PCAL (left), ECIN (middle), and ECOUT (right).

## 394 2.3.4 Training, testing and comparison of MVA classifiers on simulations

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

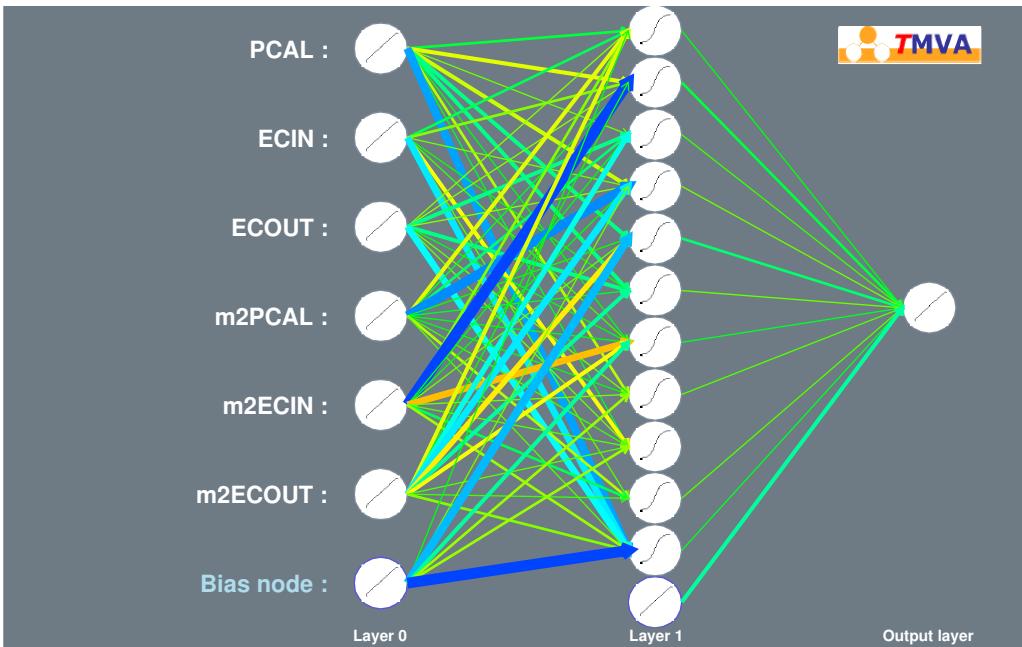


Figure 2.10: MLP using six input variables developed for the positron identification algorithm.

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

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

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

### 422 2.3.5 Test and comparison of MVA classifiers on data

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

### 426 A benchmark to assess the $\pi^+/e^+$ separation on data

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

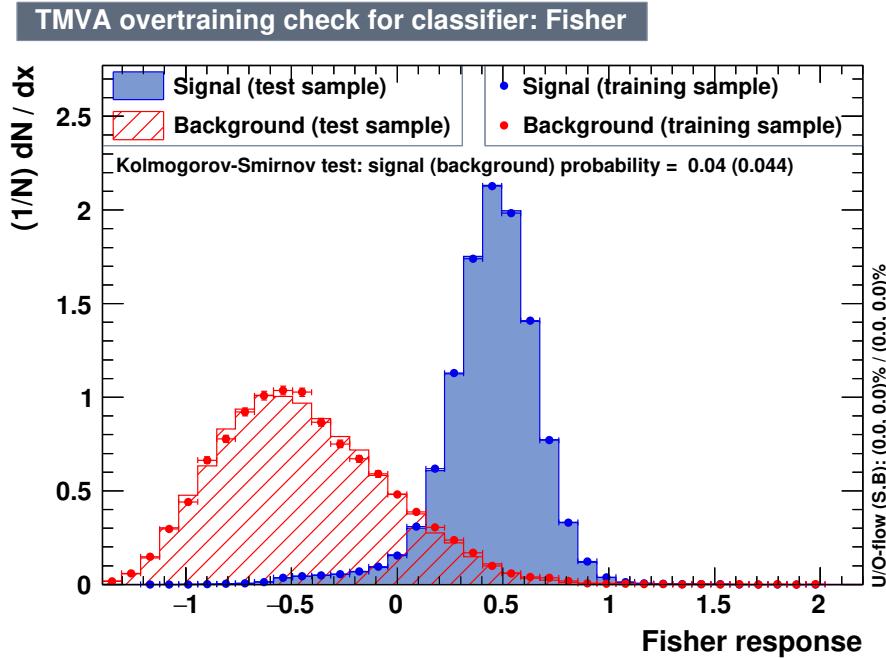


Figure 2.11: Training and testing outputs of the 3D Fisher classifier. Both distributions match very well for both the background and the signal.

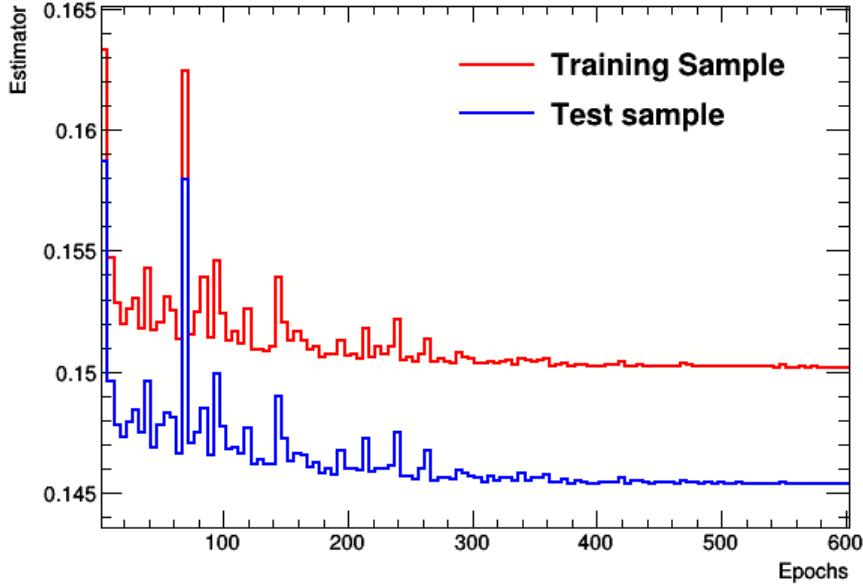


Figure 2.12: MLP 3D convergence test for the training and testing samples. The MLP error function summed over all the events in the training (red) and testing (blue) sets, respectively, is shown as a function of the training iterations. One should make sure that the convergence is reached after a certain 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 GeV<sup>2</sup>, 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

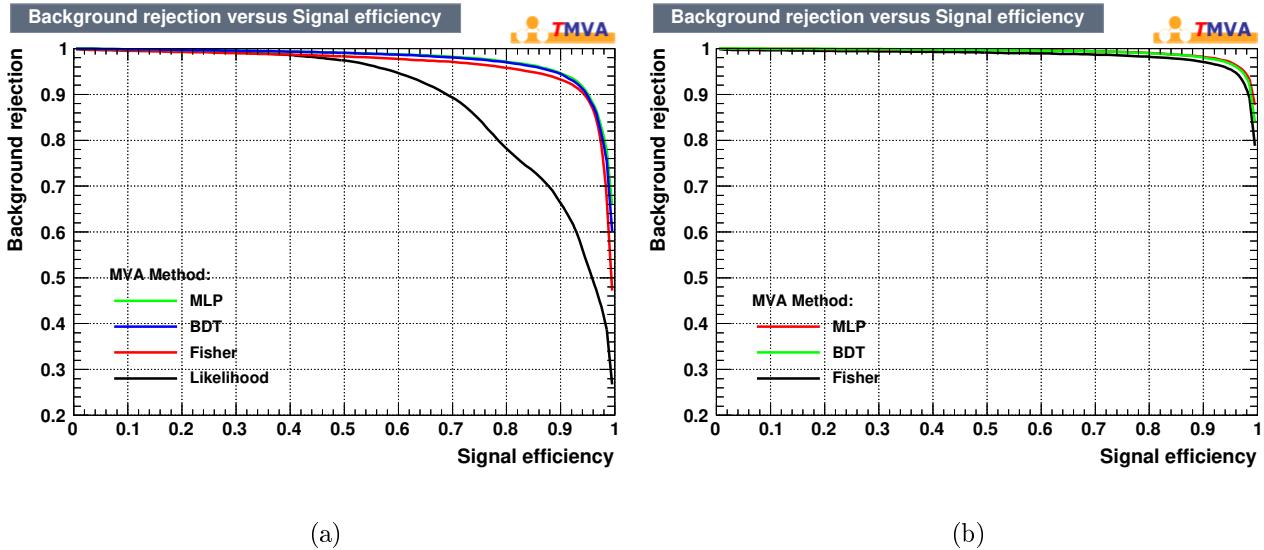


Figure 2.13: ROC curves for four classifiers (a) in the 3D case and (b) in the 6D case. In the 3D case, a Likelihood classifier was also tested, but not kept for later analysis due to its poor performance. The ROC curves of the 6D classifiers are consistently above 3D ones.

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  $\chi^2$  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  $\pi^+/e^+$  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

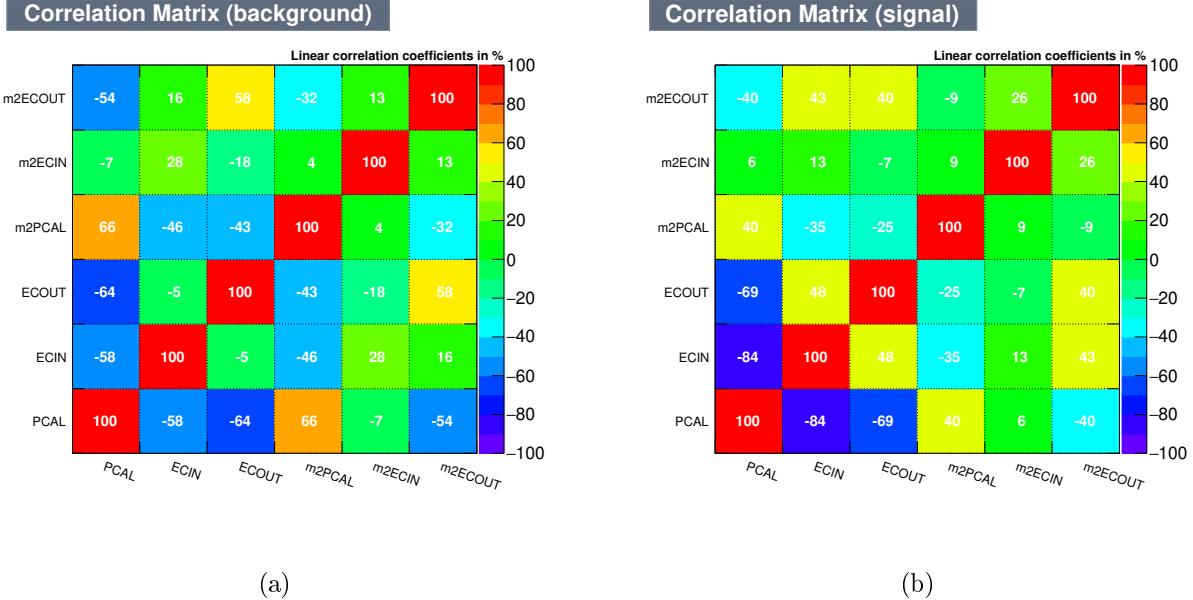


Figure 2.14: Linear correlation  $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - E[X])(Y - E[Y])]}{\sqrt{E[(X - E[X])^2]}\sqrt{E[(Y - E[Y])^2]}}$  of the 6 input variables for (a) Background (mis-id. pions) and (b) Signal (true-positrons)

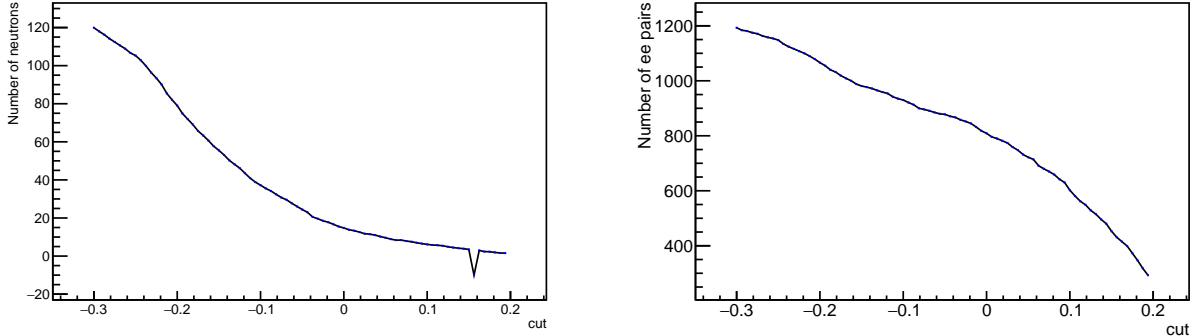


Figure 2.15: Number of neutrons as a function of the cut applied on the output of the 3D BDT (left). Number of TCS events as a function of the cut applied on the output of the 3D BDT (right).

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

#### 471 Adding skewness to the classifier

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

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

476 The pseudo-ROC curves for these two classifiers are compared to the 6D MLP and BDT in Figure  
477 2.17. One can see that all curves superimpose. Adding the skewness to the inputs of the classifiers  
478 does not improve their separation power.

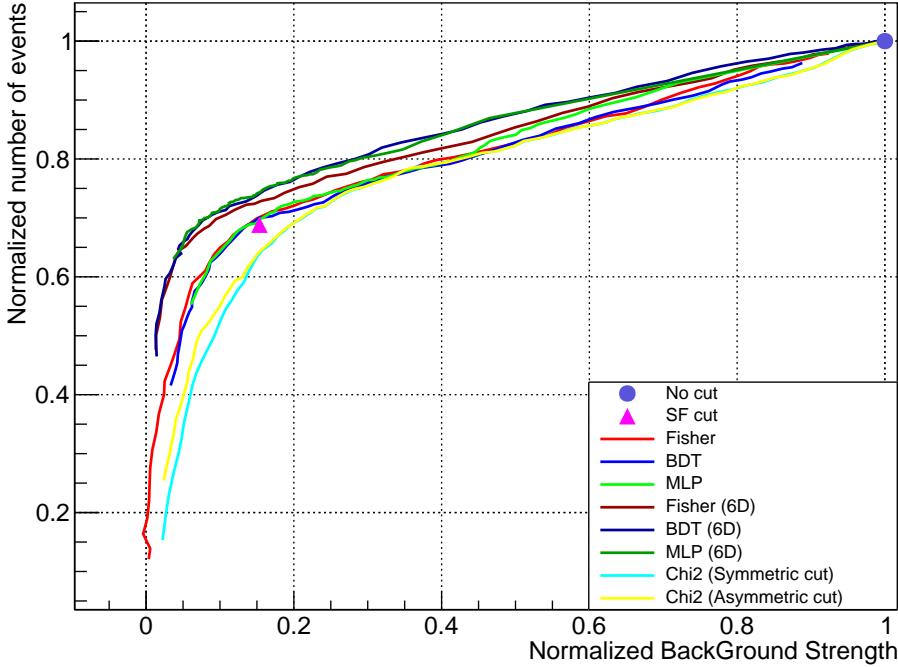


Figure 2.16: Pseudo-ROC curves obtained from data for different classifiers (3D classifiers are referred by their name only) and for the symmetric and asymmetric  $\chi^2$  cuts. The number of events in the TCS sample is plotted against the integral of the neutron peak in the neutron sample. Both axes 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.

#### 479 2.3.6 Remaining background estimation

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

486 Let  $y$  be the variable describing the *normalized number of TCS events* and  $x$  describing the *normalized*  
 487 *background strength* in Figure 2.16. In the region  $0.1 < x < 1$  the pseudo-ROC curves in Figure  
 488 2.16 are linear. Let  $B(x)$  and  $S(x)$  respectively be the number of background (mis-id. pions) and  
 489 signal events (true-positrons) in the TCS sample for the corresponding  $x$  background strength.

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

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

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

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

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

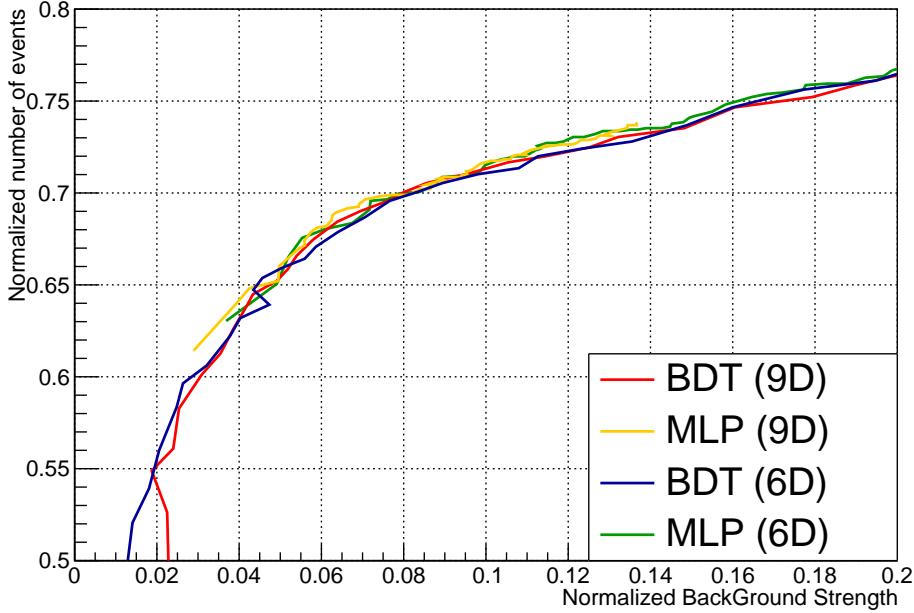


Figure 2.17: Pseudo-ROC curve obtained from data for 6D (SFs, shower widths) and 9D (SFs, shower widths and shower skewness) classifiers. The plot is zoomed in the region where the background is minimal while the signal is maximal. Adding the skewness information in the input of the classifiers does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

<sup>498</sup> Although this formula is only applicable in the linear region, we can extrapolate it to  $x = 0$ :

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

<sup>499</sup> At a given value  $x_0$  of the normalized background strength achieved with the chosen cut, the corresponding normalized number of TCS events is:

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

<sup>501</sup> Solving Equation (2.13) for  $B(x_0)$  and assuming that the signal is almost constant  $S(x_0) = S(0)$ , one <sup>502</sup> can then obtain an estimate of the background/signal ratio at  $x_0$ :

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

<sup>503</sup> The full derivation is given in Appendix A. In order to get a simple reading of the background/signal <sup>504</sup> ratio as a function of  $x$ , the pseudo-ROC curves of the 6D BDT and MLP are fitted with a linear <sup>505</sup> function in the range  $0.1 < x < 1$ . The pseudo-ROC curves  $y(x)$  are then transformed as:

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

<sup>506</sup> where  $y(0)$  is the intercept of the fit. The obtained curves are shown in Figure 2.18, from which one can <sup>507</sup> get an estimation of the background/signal ratio as a function of the normalized background strength <sup>508</sup> in the region where the function is linear.

<sup>509</sup> The B/S ratio is estimated to be close to 0.05 for a cut at 0.5 on the MLP output corresponding <sup>510</sup> to a Background strength of 0.15 (see Figure 2.19 for values of the background as a function of the cut <sup>511</sup> applied on the output of the 6D MLP classifier). Finally the pion contamination  $C_{\pi^+}$  above 4 GeV is <sup>512</sup> estimated as:

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

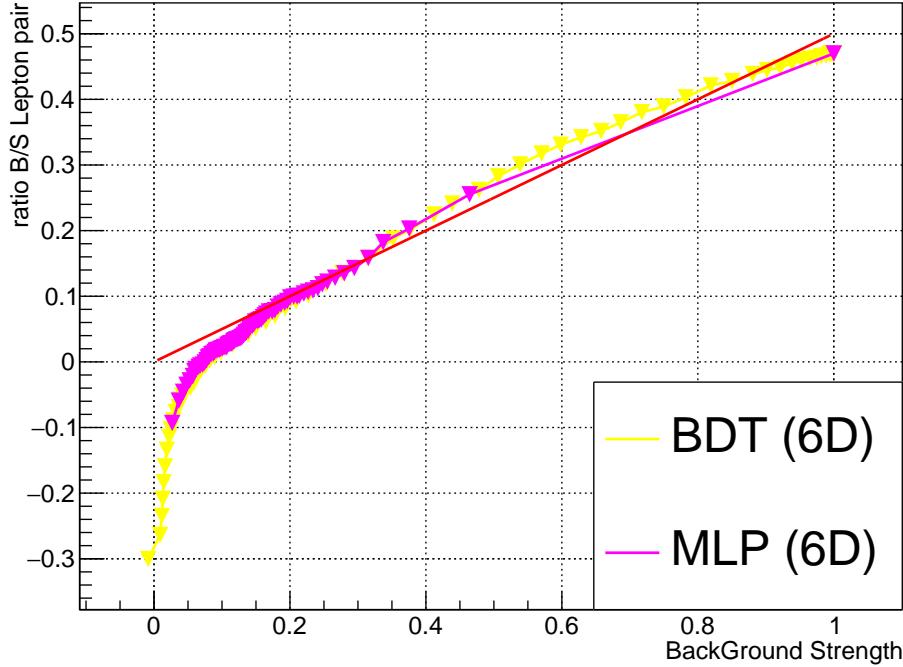


Figure 2.18: Background/Signal ratio in the TCS sample as a function of the background strength evaluated with the neutron sample. The red line corresponds to the linear fit of the BDT curve between 0.1 and 1.

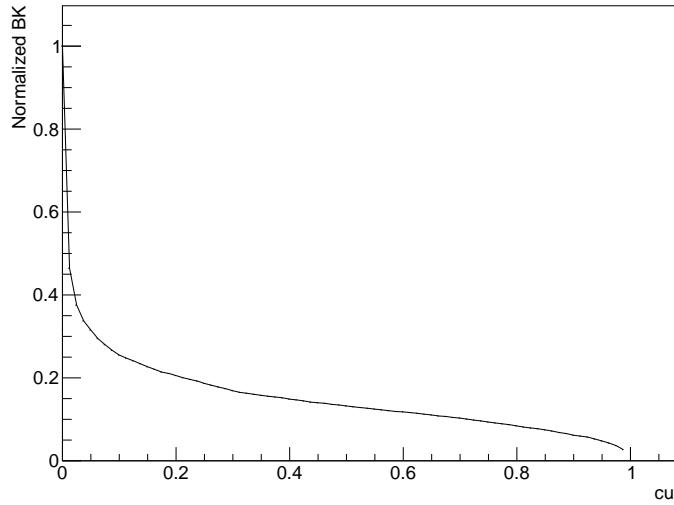


Figure 2.19: Normalized remaining background as a function of the cut applied to the output of the 6D MLP classifier.

### <sup>513</sup> 2.3.7 Systematic checks on simulations

<sup>514</sup> A second systematic check of the efficiency of the 6D MLP classifier was performed on simulations.  
<sup>515</sup> A sample of BH-weighted events (see Chapter 3 for details on the TCS simulations) was used to test  
<sup>516</sup> the classifier. This sample is completely uncorrelated from the training sample. It also assumes no  
<sup>517</sup> hypothesis on momentum, polar angle and transverse momentum, apart from the ones arising from  
<sup>518</sup> 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 ( $\phi$  between  $100^\circ$  and  $150^\circ$ ) and 5 (between  $-150^\circ$  and  $-100^\circ$ ) 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  $\pi^+$  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  $\theta$  distributions of  $\pi^+$  before and after the cut. No large systematic efficiency shift is seen.

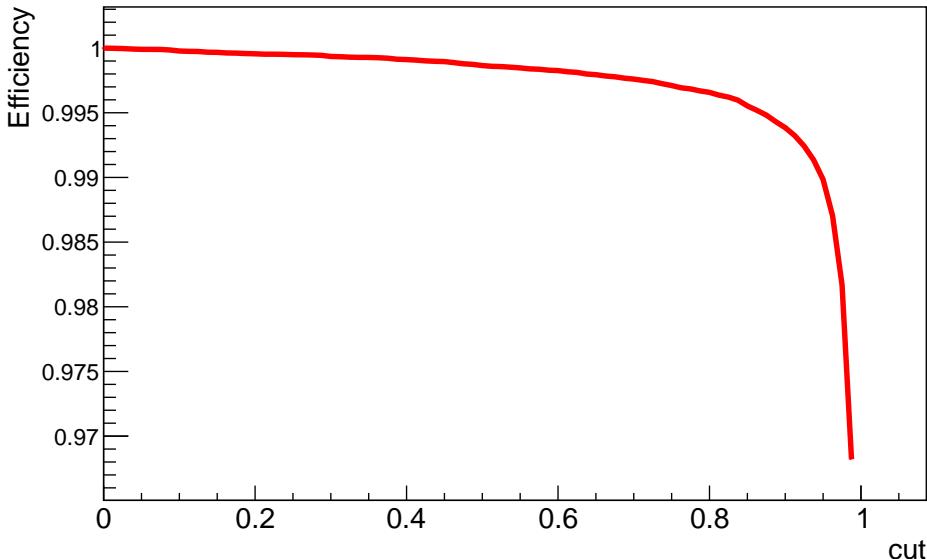


Figure 2.20: Signal efficiency for simulated TCS events (inbending electron) using the 6D MLP classifier as a function of the cut applied on the classifier output. At cut=0.5, the fraction of signal which is lost is less than a 1%.

### 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.

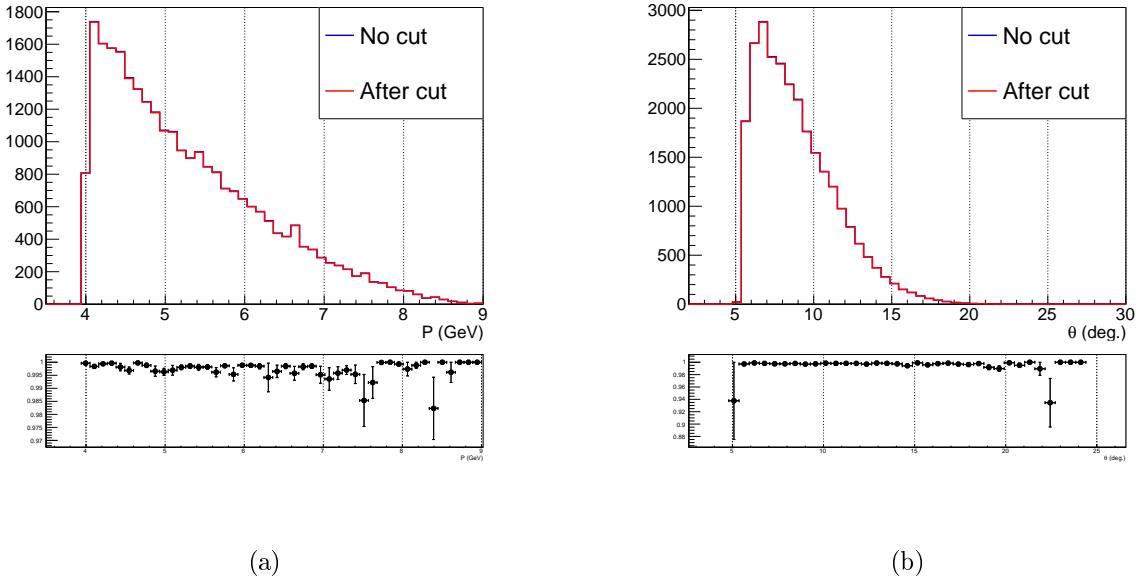


Figure 2.21: (a) Momentum spectrum of positrons for simulated TCS events. (b)  $\theta$  spectrum of positrons for simulated TCS events. In both cases the histogram before and after MVA pid cut, of value 0.5, superimpose. Ratios of both histograms are also shown. The number of cut events per bin is always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

### 547 2.3.9 What about electron PID?

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

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

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

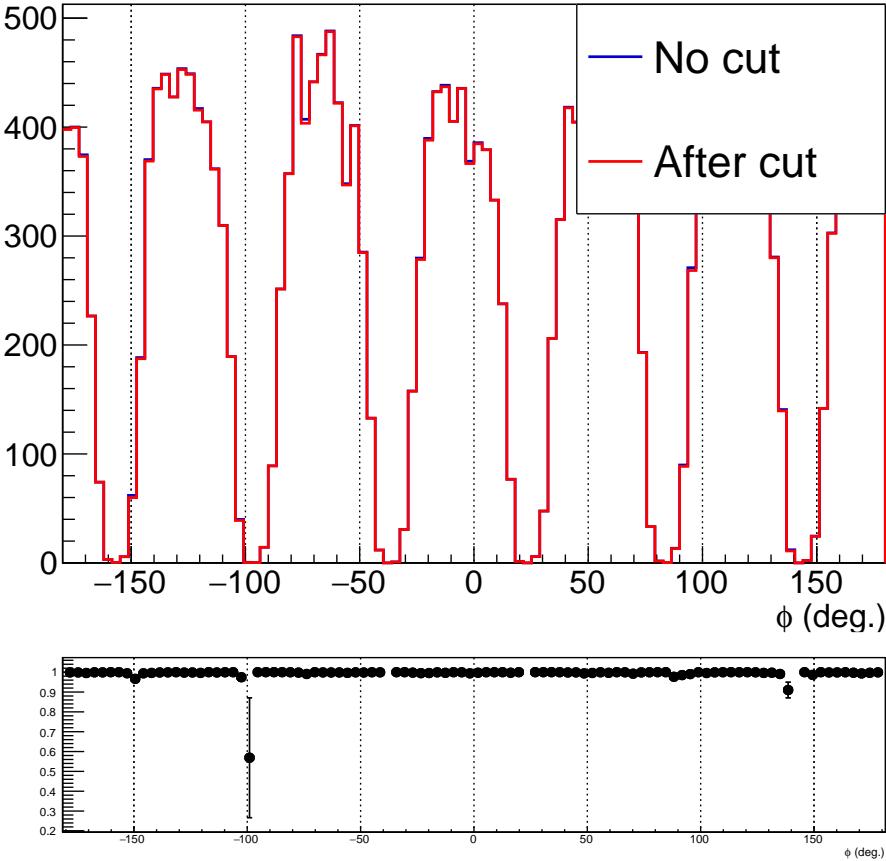


Figure 2.22:  $\phi$  spectrum of positrons for simulated TCS events. The loss of positrons happens at very specific  $\phi$  angles corresponding to the edges of sectors which accommodate an LTCC module.

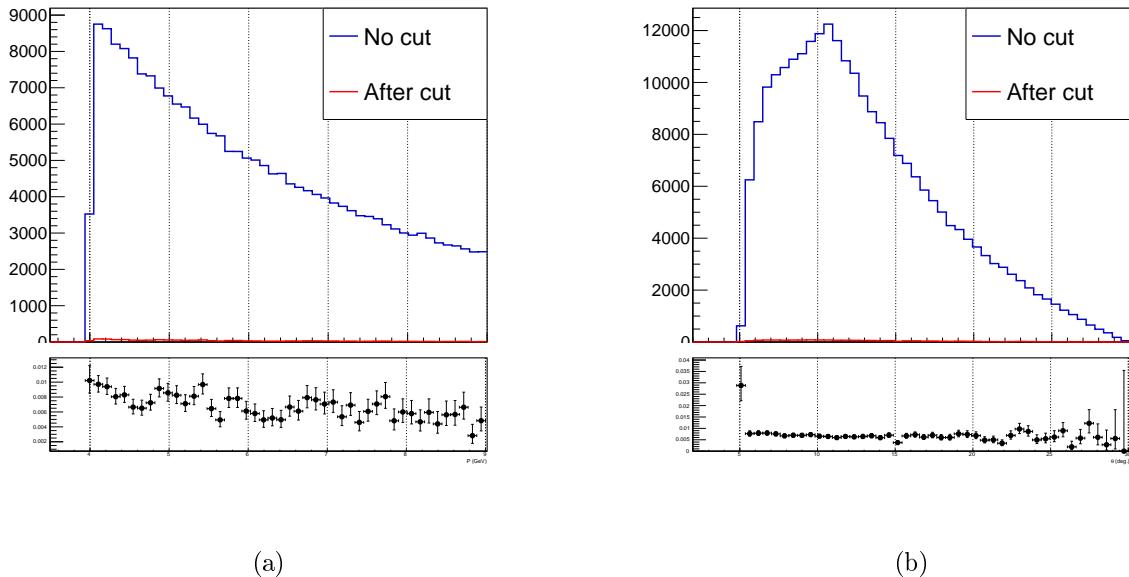


Figure 2.23: (a) Momentum spectrum of mis-id. pions. (b)  $\theta$  spectrum of mis-id. pions. The histograms before and after the MLP cut at 0.5 are shown.

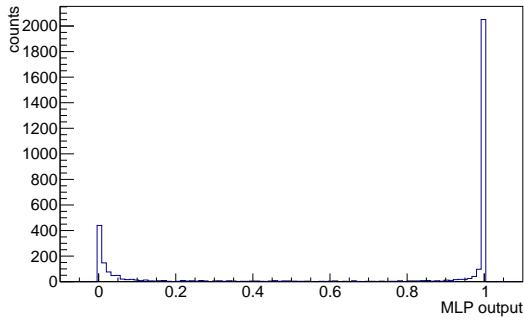


Figure 2.24: Output of the 6D MLP neural network on the TCS events used in the analysis.

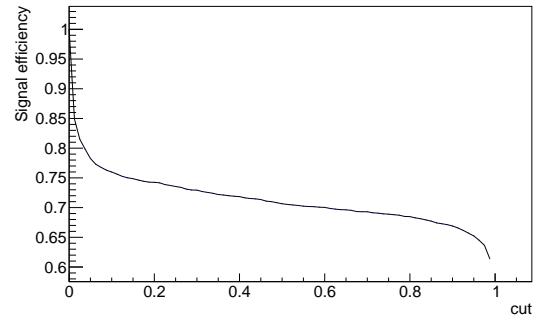


Figure 2.25: Signal efficiency of the cut applied on the MLP output. This corresponds to the proportion of remaining events for a given cut.

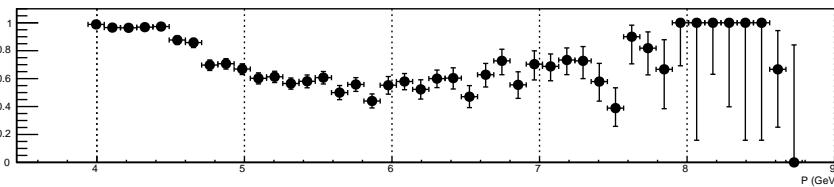
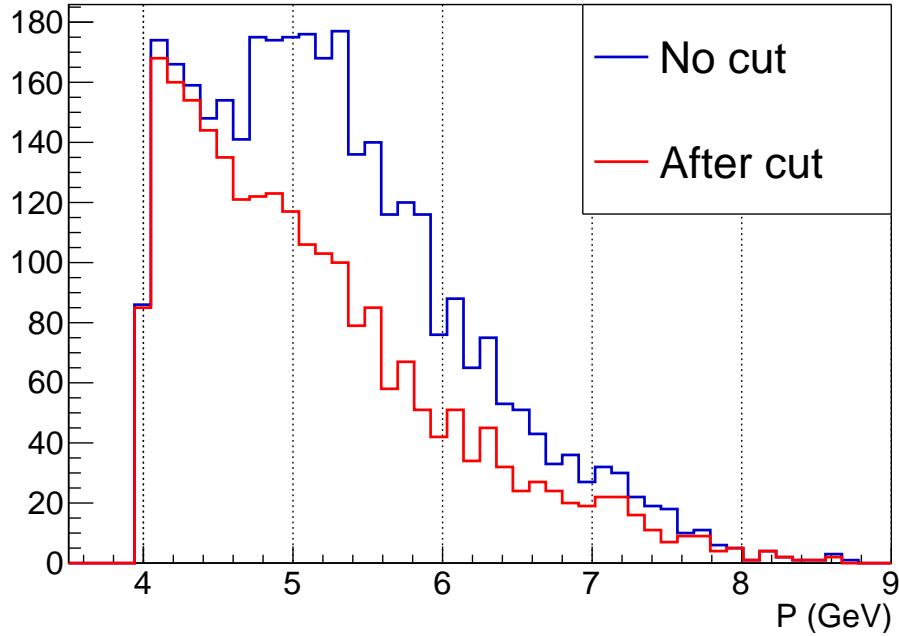


Figure 2.26: Momentum of positrons for TCS events. The blue histogram corresponds to all events. The red histogram corresponds to events that passed a 6D MLP output cut of 0.5.

## 570 2.4 Proton momentum corrections

571 The proton momentum corrections are split in two main contributions. The first corrections, pre-  
 572 sented in Subsection 2.4.1, are determined by comparing Monte-Carlo generated and reconstructed  
 573 proton kinematics. The shifts observed in this case is attributed to the energy lost by the proton while  
 574 propagating in the various detector materials. The determination of the parameters of this correction

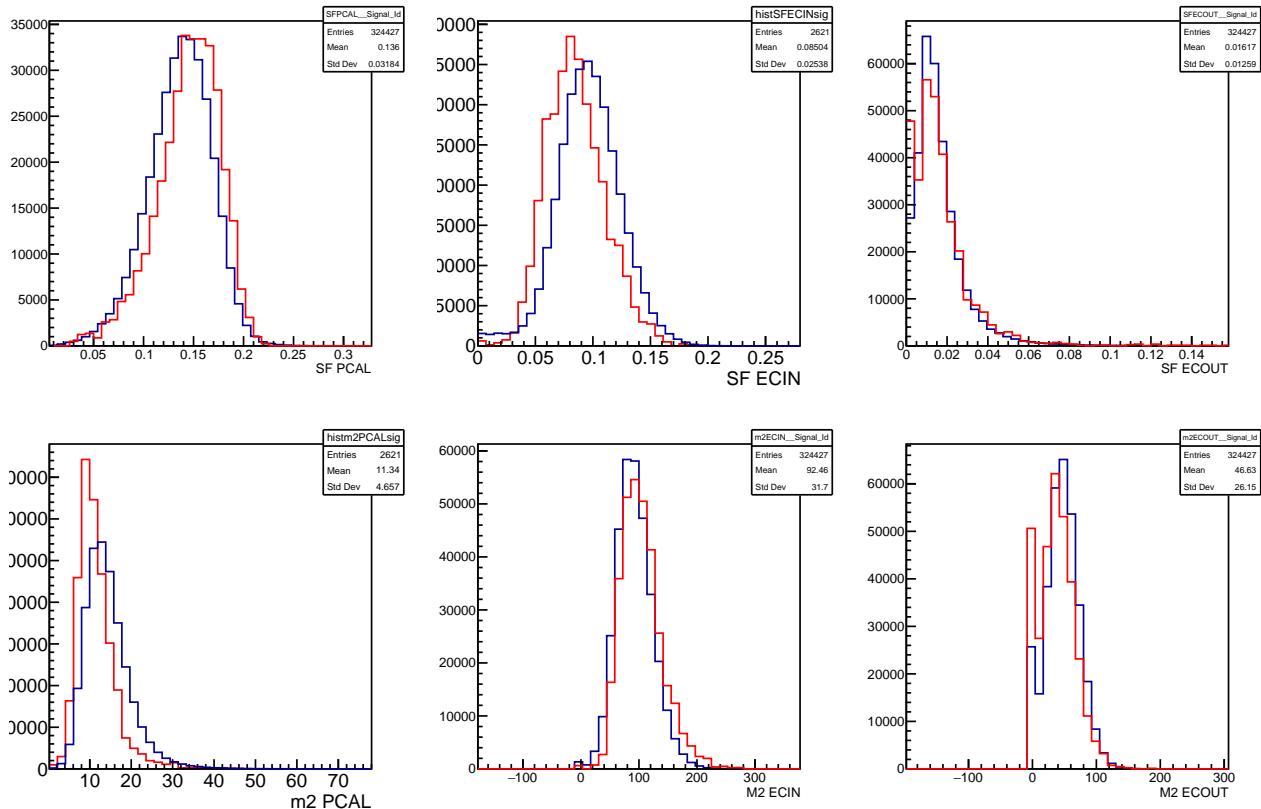


Figure 2.27: Comparison of the data and simulation spectra of the 6 variables used in the neural network for signal events. The red curves corresponds to data, the blue curves to simulation. The curves are normalized to have equal integrals.

575 solely relies on simulations. The second contribution is a data-only based correction, which aims at  
 576 correcting mis-alignments and inefficiencies of the actual detectors, not accounted for by the simula-  
 577 tion. In particular, data-driven corrections are developed for the momenta of the protons in the CD  
 578 in Subsection 2.4.2.

#### 579 2.4.1 Monte-Carlo corrections

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

584 The difference between the generated and reconstructed momenta,

$$\Delta P = P_{Gen.} - P_{Rec.}, \quad (2.17)$$

585 is studied as a function of the polar angle of the proton,  $\theta$ . The plot in Figure 2.31 shows the difference  
 586 between the generated and reconstructed proton momenta as a function of  $\theta$  for protons detected in the  
 587 FD. The momenta difference shows different behaviors in two distinct regions. Below  $27^\circ$  protons cross  
 588 little material before being detected and the momentum difference is small (below 20 MeV). Above  $27^\circ$   
 589 the material budget between the target and the DCs is larger, especially due to the HTCC and the  
 590 forward CTOF light-guides. In this region the momentum resolution is degraded and the momentum  
 591 difference can reach up to 80 MeV.

592 The MC corrections are derived in three different CLAS12 regions: the two regions in the FD  
 593 described above and one region in the CD. In each region the momenta difference is parametrized as  
 594 a function of the reconstructed momentum as shown in Figure 2.32. These corrections are at most of

## 2.4. Proton momentum corrections

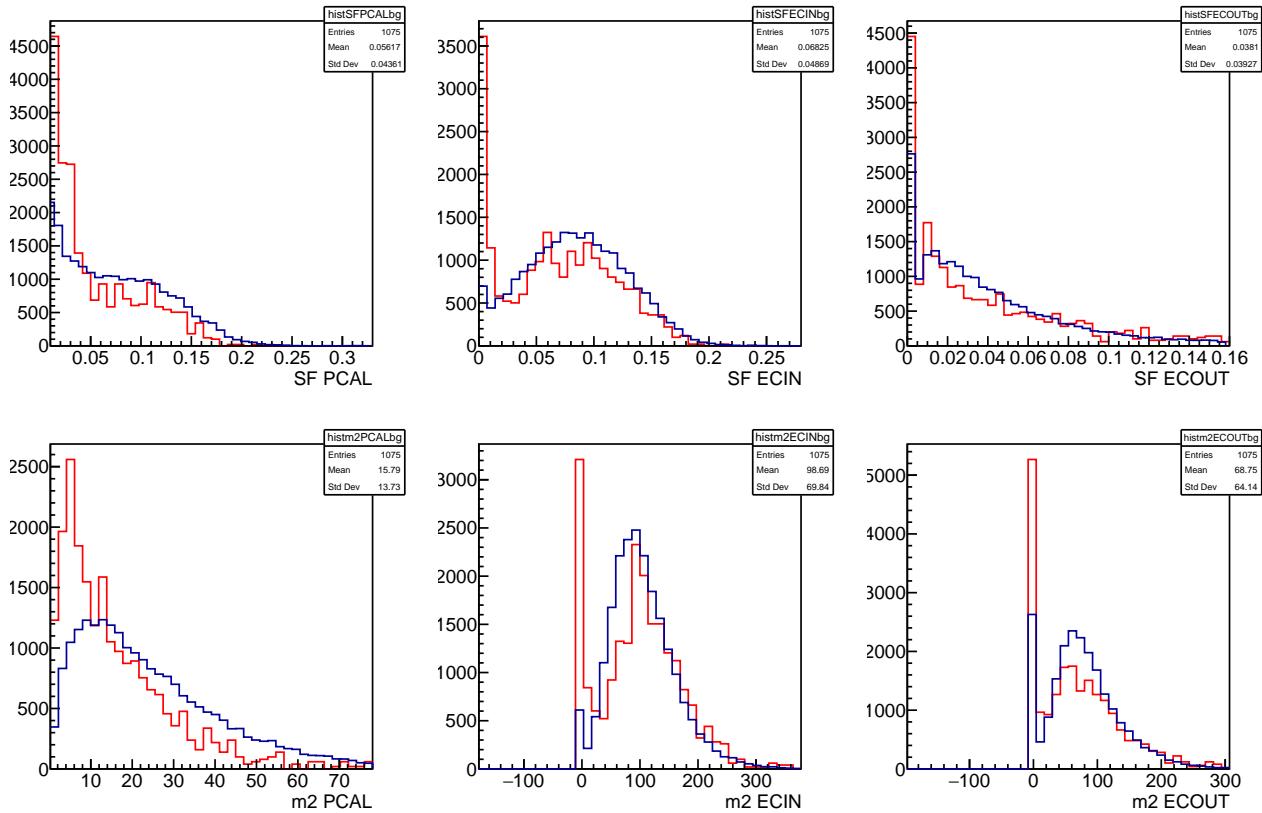


Figure 2.28: Comparison of the data and simulation spectra of the 6 variables used in the neural network for background events. The red curves corresponds to data, the blue curves to simulation. The curves are normalized to have equal integrals.

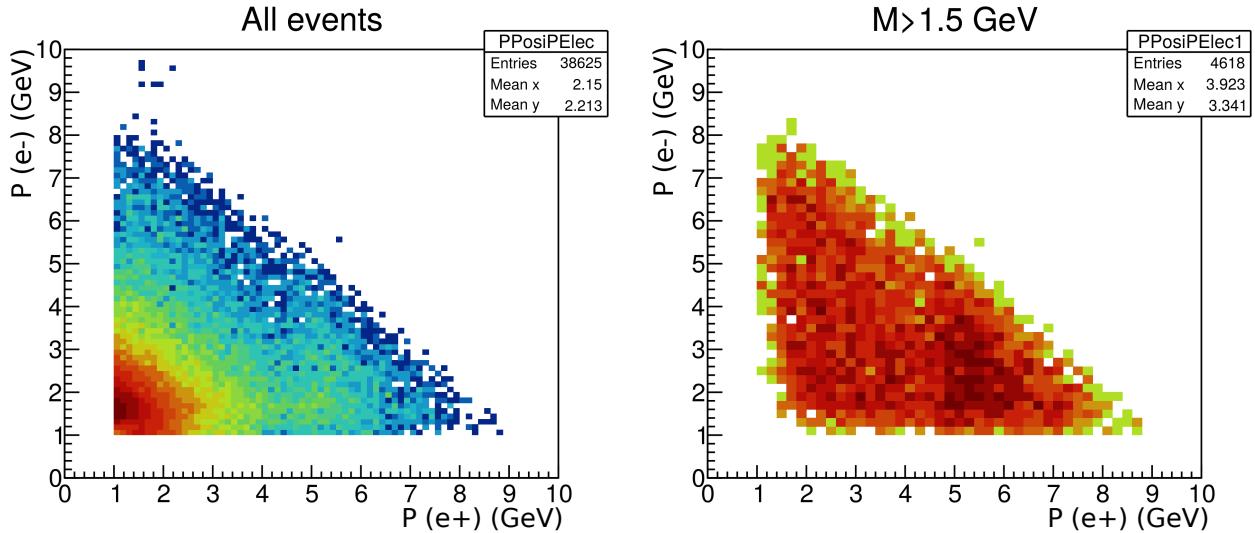


Figure 2.29: Momentum of the positron versus momentum of the electron in  $e^-e^+p$  events for all invariant masses (left) and only invariant masses higher than 1.5 GeV (right). No additional PID cuts than the EB ones are used in these two plots. The  $\pi^+$  contamination is clearly visible in the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

595 the order of 4% for low-momenta proton ( $\sim 0.45$  GeV) in the high-polar angle region of the FD. The

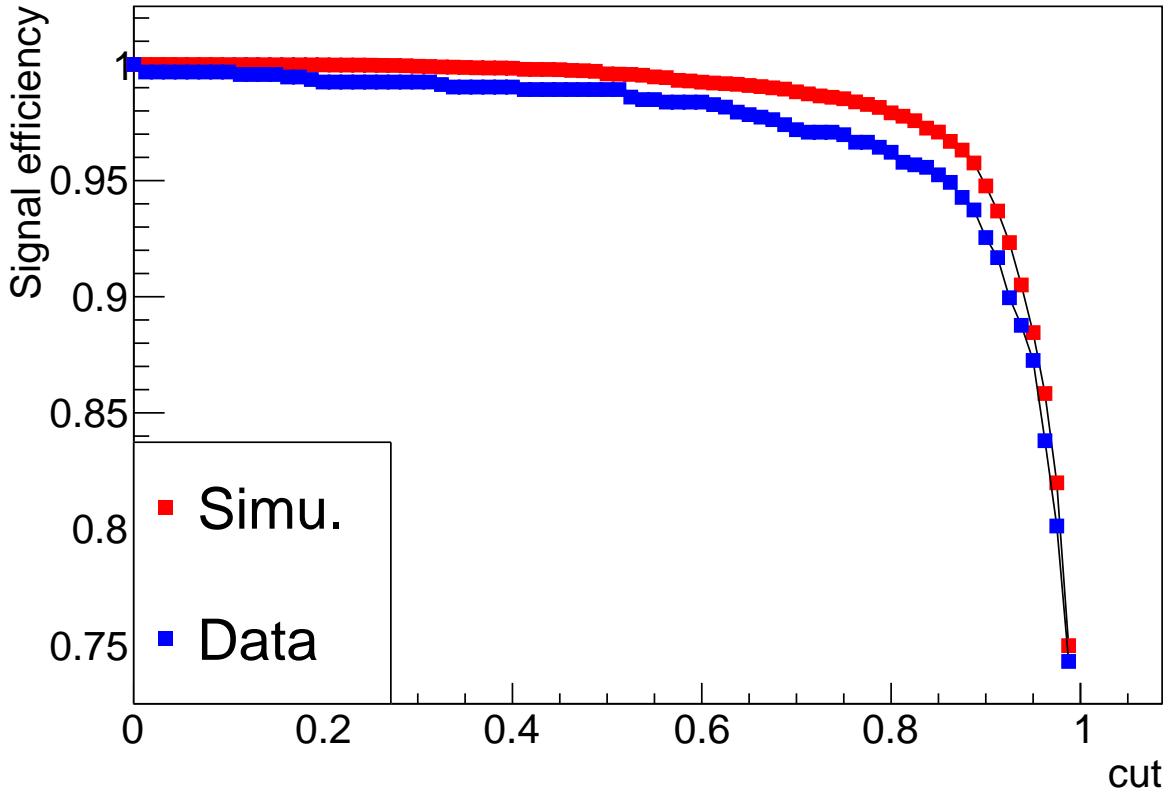


Figure 2.30: Signal efficiency of the MLP classifier for electrons with momentum bigger than 4 GeV, for simulated TCS events (red) and data (blue).

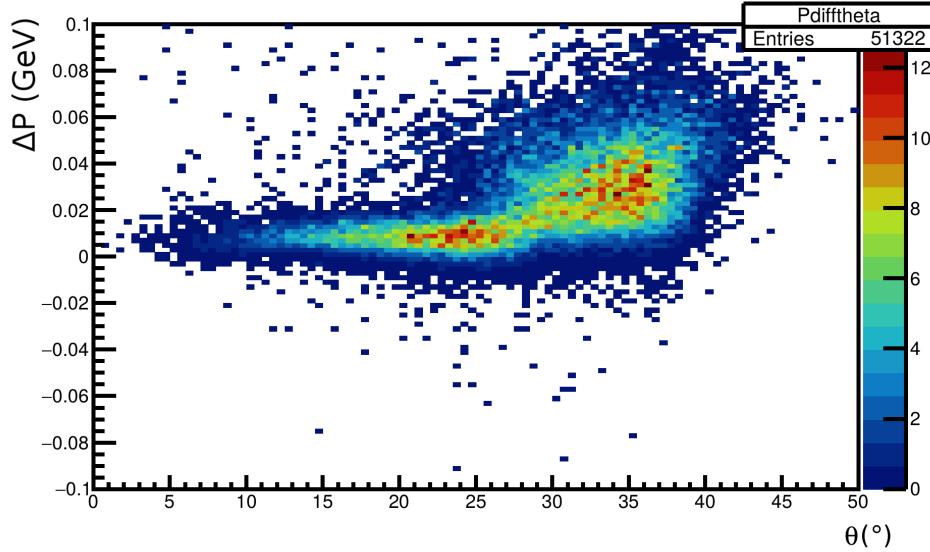


Figure 2.31: Difference between the generated and reconstructed momenta for protons detected in the FD of CLAS12. One can see two distinct regions, below and above  $\theta = 27^\circ$ .

596 corrections are applied to protons in both the simulations and the data.

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

$$f_{MC}(P) = a_0 + a_1 \cdot P + a_2 \cdot P^2, \quad (2.18)$$

599 where the coefficient  $a_i$  are given in Table 2.1. The corrected momentum then reads:

$$P_{Corr. MC} = P_{Uncorr.} + f_{MC}(P). \quad (2.19)$$

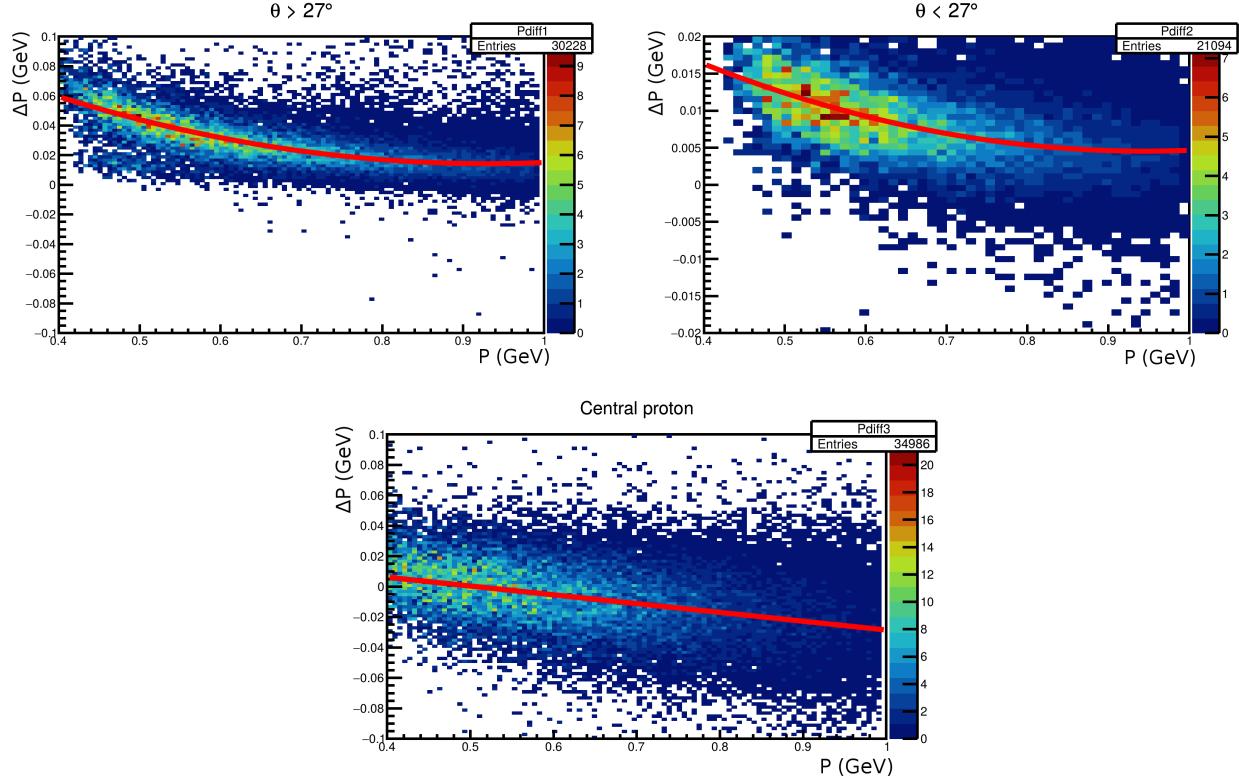


Figure 2.32: Top left: momenta difference (defined in Equation (2.17)) as a function of the reconstructed momentum for polar angles above  $27^\circ$  in the FD. The distribution of the mean of each momentum slice (obtained by a gaussian fit) is fitted with a 2nd order polynomial. The result of this fit is superimposed to the original distribution (red line). Top right: Corresponding figure for protons with polar angles below  $27^\circ$  in the FD. Bottom: Corresponding figure for protons in the CD.

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

Table 2.1: Values of the coefficients used in the MC-based momentum for proton.

#### 600 2.4.2 Data-driven momentum corrections

601 Data-driven momentum corrections for the proton are motivated by the fact that the simulations  
 602 depict an "ideal" detector, and therefore do not perfectly reproduce the data. Because of issues mainly  
 603 due to the reconstruction software of the CVT, the reconstructed momentum in the CD can be shifted  
 604 from its actual value. To investigate this issue a method using exclusive two-pion production events  
 605 was developed. This method relies on the exclusive measurement of the  $ep \rightarrow e'p'\pi^+\pi^-$  reaction, where  
 606 the scattered electron and the pions are detected in the FD. The kinematics of the scattered proton can  
 607 then be studied in two different ways. In one case the proton can be detected by CLAS12, in the other  
 608 case its kinematics can be inferred by calculating the missing 4-momentum of  $X$  in the  $ep \rightarrow e'\pi^+\pi^-X$   
 609 reaction.

610 The missing mass spectrum obtained from the latter analysis is shown in Figure 2.33. One can see  
 611 that the missing mass shows a clear peak at the proton mass. Furthermore, looking at the dependence

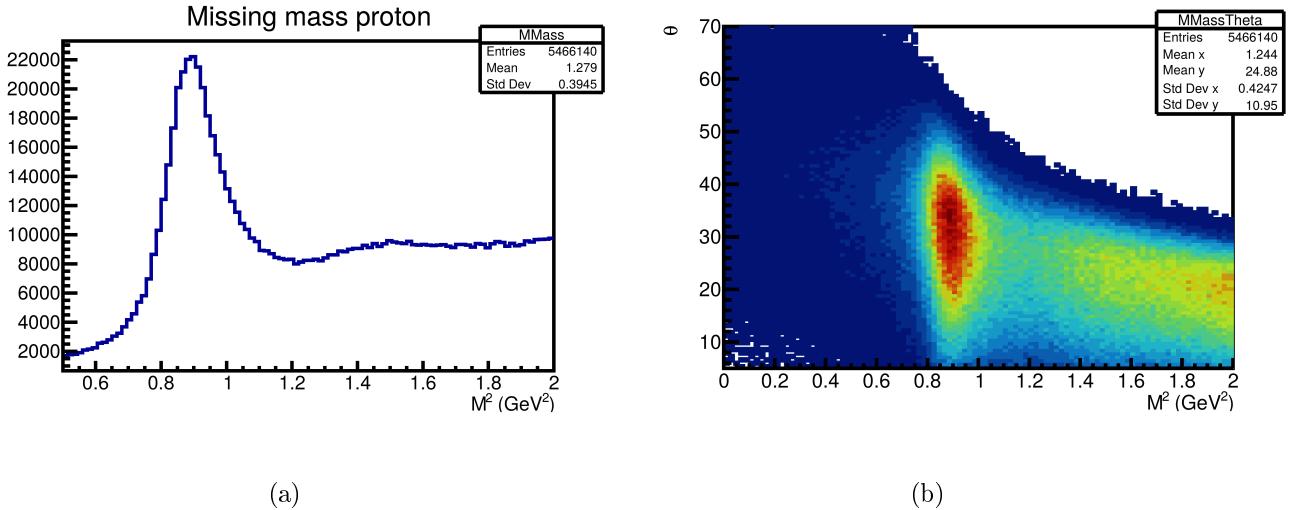


Figure 2.33: (a) Squared missing mass of the  $ep \rightarrow e'\pi^+\pi^-X$  reaction. One can see a clear peak at the proton mass and a higher-mass continuum. (b) Missing polar angle as a function of the squared missing mass for the same reaction. Once can see that the high-polar-angle region, corresponding to topologies where the missing proton goes in the CD, is free of high-mass background.

of the missing mass as a function of the missing polar angle in Figure 2.33b, one can see that at high polar angles (above  $35^\circ$ ) 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.

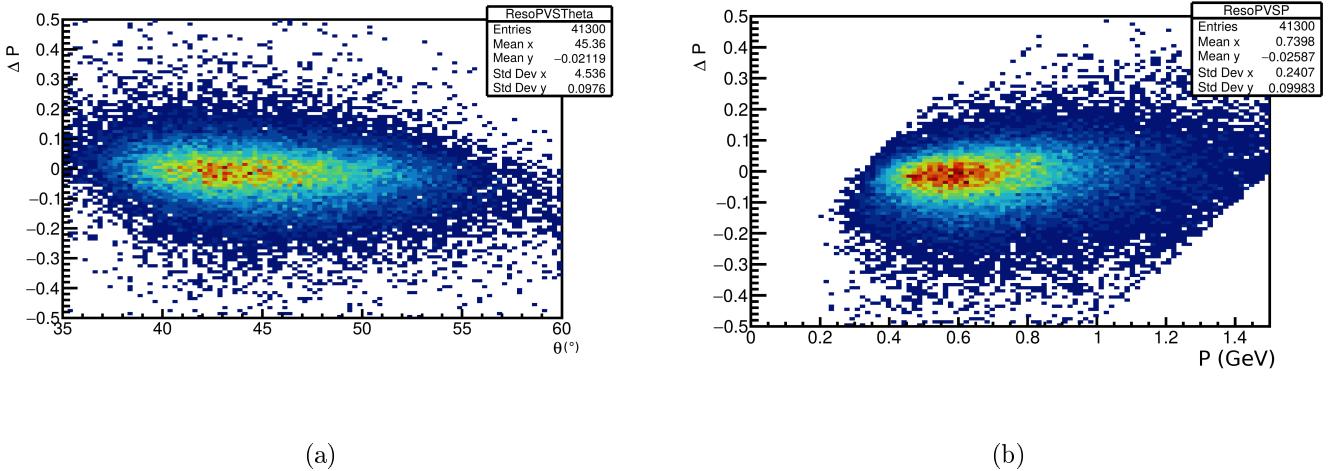


Figure 2.34: (a) Data-driven momentum difference as a function of the polar angle for protons in the CD. (b) Momentum difference as a function of the momentum.

Figure 2.34 shows the momentum difference:

$$\Delta P = P_{Rec.} - P_{Missing.} \quad (2.20)$$

as a function of the detected polar angles (Figure 2.34a) and the detected momenta (Figure 2.34b). No large dependencies are seen. The momentum resolution defined as:

$$\frac{\Delta P}{P} = \frac{P_{Rec.} - P_{Missing.}}{P_{Rec.}} \quad (2.21)$$

is also plotted as a function of the local azimuthal angle in the last layer of the CVT,  $\phi_{CVT}$ , as shown in Figure 2.35.

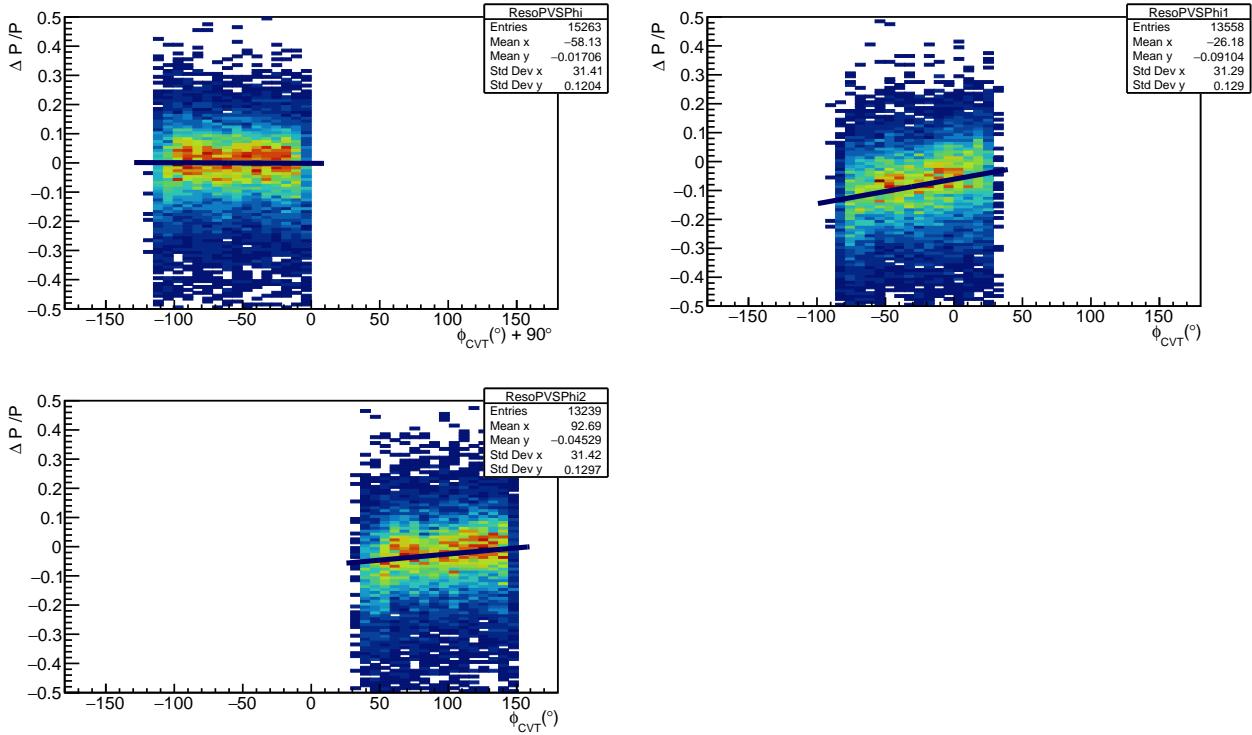


Figure 2.35: Momentum resolution as a function of the local azimuthal angle for protons in the CD, for the three regions of the CVT. The superimposed black line is the correction function for each region.

621     Each subplot corresponds to one of the three CVT regions. The azimuthal coordinate of the last  
622     layer of the CVT is used for this correction (layer ID 12 in the CLAS12 reconstruction nomenclature).  
623     The distribution of the gaussian means of each  $\phi_{CVT}$  slices is fitted with a linear function, for each  
624     subplot. The obtained coefficients are detailed in Table 2.2, where the functions can be written:

$$f_{DATA}(\phi_{CVT}) = c_0 + c_1 \cdot \phi_{CVT}. \quad (2.22)$$

625     It has to be noted that for the range  $-210 < \phi_{CVT} < -90^\circ$  a  $90^\circ$  shift is used for the angle to be in the  
626      $[-180^\circ, 180^\circ]$  range. The coefficient given in Table 2.2 are given after this translation. The resulting  
627     corrections are applied to the proton momenta in the data. The corrected momentum reads:

$$P_{Corr. DATA} = P_{Uncorr.} \cdot (1 - f_{DATA}(\phi_{CVT})). \quad (2.23)$$

628     The corrections range from almost zero for protons in the region  $-210^\circ < \phi_{CVT} < -90^\circ$  to up to  
629     14% at the lower edge of the  $-90^\circ < \phi_{CVT} < 30^\circ$  region. These corrections are performed after the  
630     MC corrections presented in the previous subsection.

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

Table 2.2: Values of the coefficients used in the data based momentum for proton in the CD.

631     *End of common analysis section*

## 633     2.5 Lepton momentum corrections

634     In this section we present the corrections applied to leptons. First, in Subsection 2.5.1, the shift be-  
635     tween the Monte-Carlo generated and the reconstructed kinematics is analyzed. Second, in Subsection

636 2.5.2, a specific detected-photon correction is presented.

637 **2.5.1 Monte-Carlo corrections**

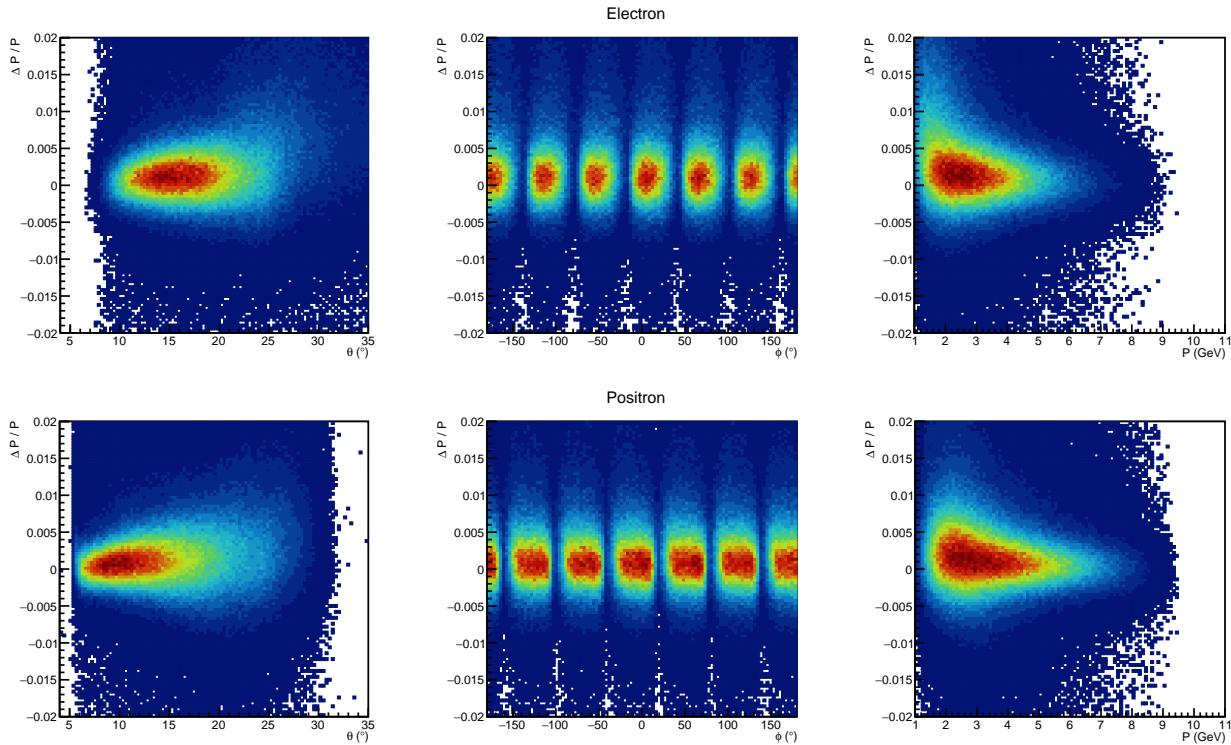


Figure 2.36: Reconstructed momentum resolution for electrons (top) and positrons (bottom), as a function of the lab angles  $\theta$ ,  $\phi$  and the momentum.

638 As for proton, BH-simulation events are used to investigate the impact of the energy lost by the lepton  
 639 in the material of CLAS12 before being detected. This energy loss ultimately affects the reconstruction  
 640 of the kinematic variables, especially the momentum. Figure 2.36 shows, for the electrons and the  
 641 positrons, the momentum resolution:

$$\frac{\Delta P}{P} = \frac{P_{\text{Gen.}} - P_{\text{Rec.}}}{P_{\text{Rec.}}}, \quad (2.24)$$

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

646 **2.5.2 Detected radiated photon correction**

647 Leptons propagating through the target material can lose energy by radiating photons. These  
 648 photons are emitted at low angles around the direction of the lepton. The process is represented by  
 649 the diagram of Figure 2.37.

650 One way to recover the initial momentum of the lepton, before any radiative energy loss, is to use  
 651 the detected photon in the CLAS12 EC. Figure 2.38a shows, in the CLAS12 data, the uncorrected  
 652 momenta of electrons versus the difference in the polar angle at the vertex between electrons and  
 653 detected photons,  $\Delta\theta_\gamma$ , while Figure 2.38b shows the cone angle between electrons and detected photons  
 654 (the angle between the 3-momentum of the electron and the 3-momentum of each photon). One can  
 655 see that in both plots there is a significant number of photons detected in the close vicinity of the

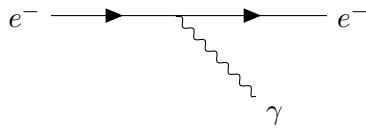


Figure 2.37: Diagram representing the radiation of a photon from an electron.

656 electron. The same behaviour is seen for positrons. Similar results are also obtained for simulations,  
657 as seen in Figure 2.39.

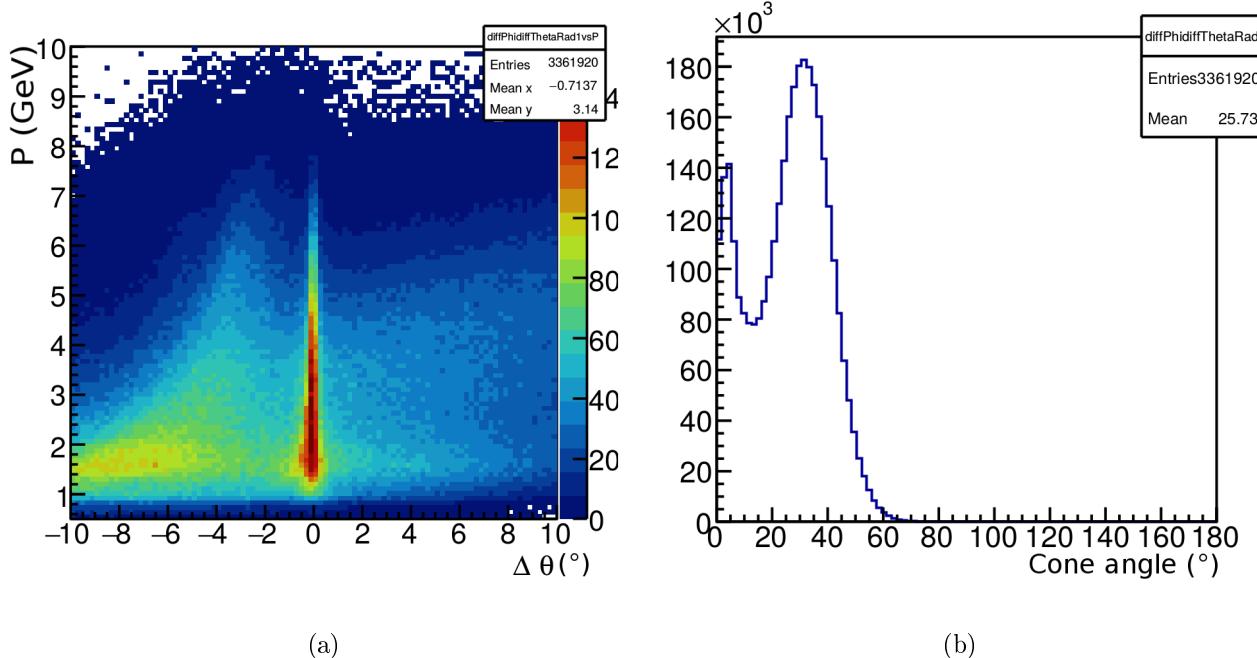


Figure 2.38: (a) Electron momentum as a function of the difference of polar angle between the electron and the photons detected in each event, (b) Cone angle between the electron and same event photons. These plots are produced with real CLAS12 data.

658 The 4-momenta of photons within  $-1.5^{\circ} < \Delta\theta_{\gamma} < 1.5^{\circ}$  and with a cone angle below  $10^{\circ}$  are added  
659 to the reconstructed 4-momentum of the corresponding lepton. The effect of this correction can be  
660 seen on simulation in Figure 2.40. The simulations used for this study are described in Section 3.1.  
661 The improvements in the momentum and angular resolutions are shown for the electrons for which  
662 this correction is applied.

## 663 2.6 Fiducial cuts

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

671 The PCAL fiducial cut developed in this section is based on the width of the measured electromagnetic  
672 shower. The definition of the square of the shower width in the PCAL is given in Formula (2.6).  
673 The mean shower size  $\sqrt{M_2}$  is calculated for each of the three calorimeter coordinates (U, V and W).  
674 Figures 2.41 and 2.42 show the distributions, for sectors 1 to 3, of the mean shower size as a function

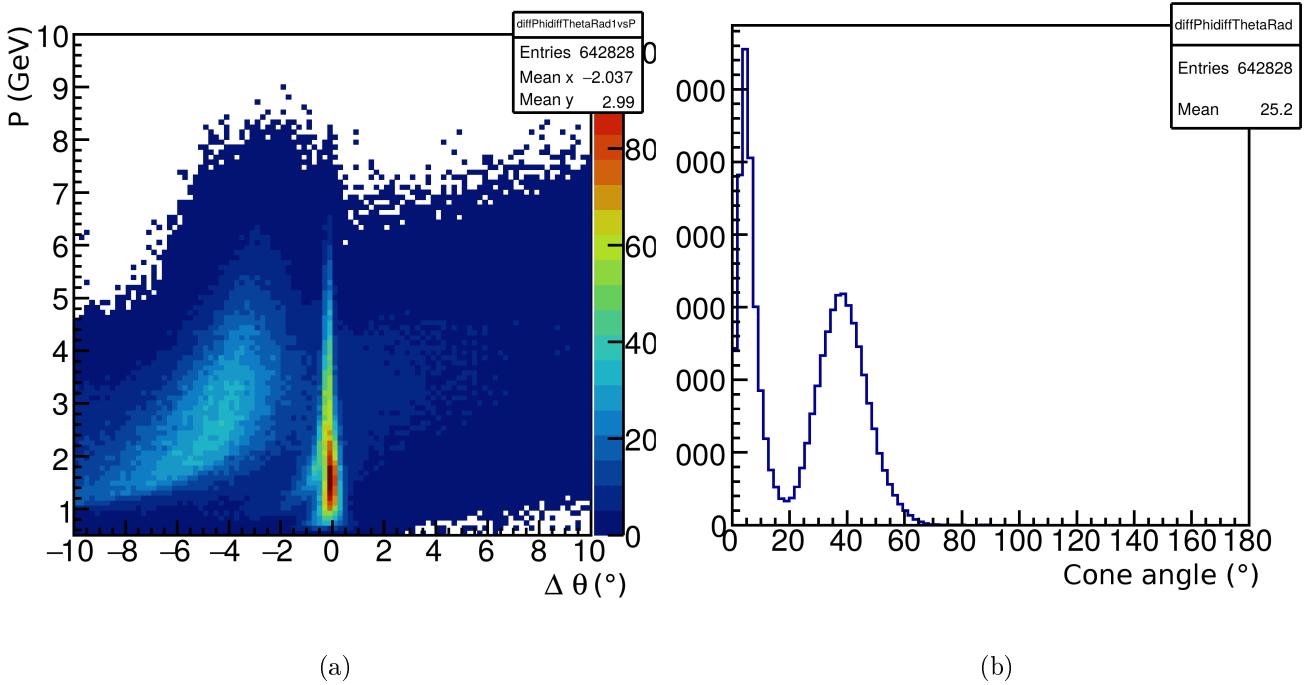


Figure 2.39: (a) Electron momentum as a function of the difference of polar angle between the electron and the photons detected in each event, (b) Cone angle between the electron and same event photons. These plots show simulated electrons.

675 of respectively low and high values of  $U$ ,  $V$  and  $W$ . The shower mean size is estimated away from the  
 676 edge of the calorimeter (red lines). The value of the cut is then set at a distance equal to the mean  
 677 shower size from the edge of the distribution (black lines).

678 The cut values used in the TCS analysis are summarized in Table 2.3, where the maximum and  
 679 minimum values of  $V$  and  $W$  are given for each sector of CLAS12. Only the cuts along  $V$  and  $W$  are  
 680 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 2.3: 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.

681 The effect of the PCAL fiducial cuts on the electron distribution is shown in Figure 2.43. The  
 682 left plot shows the removed electrons in black. The plot on the right shows the regions kept for the  
 683 analysis. The fraction of electrons lost by these fiducial cuts was estimated on inclusive events (at least  
 684 one electron detected in CLAS12) from the inbending data set. It ranges from 5% at 1 GeV to up to  
 685 30% at 10 GeV.

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

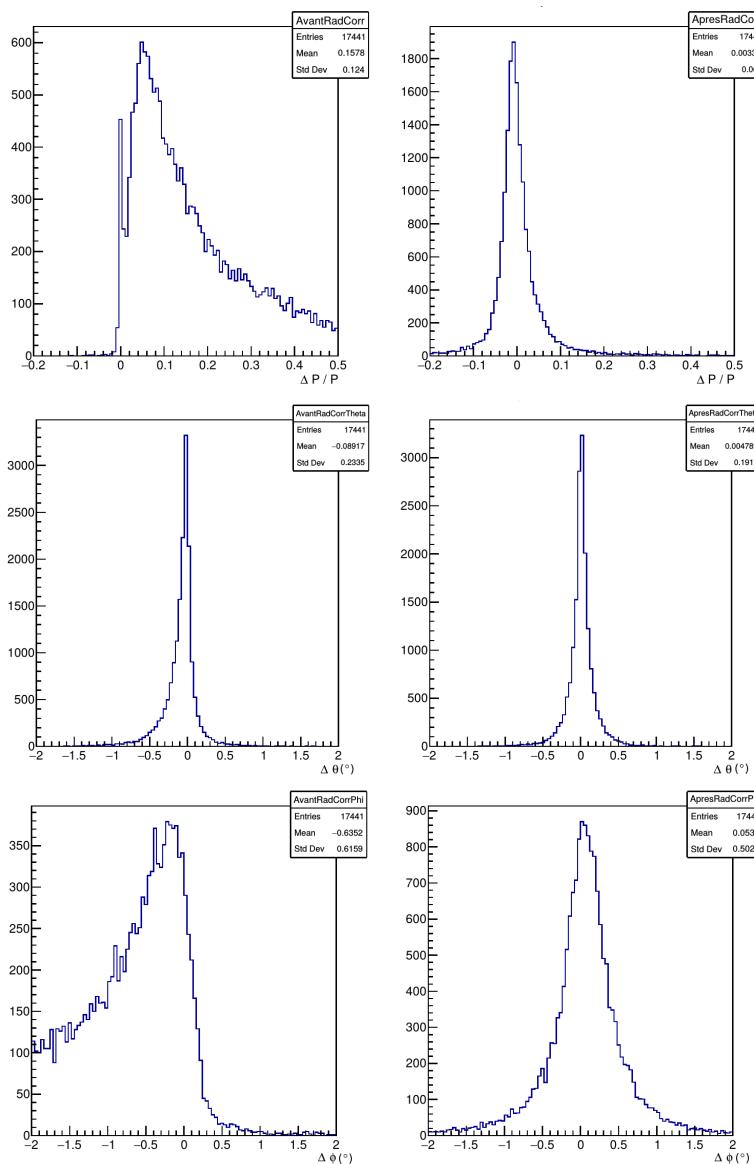


Figure 2.40: Distributions of the momentum resolution (top), the  $\theta$  difference (middle) and the  $\phi$  difference (bottom) before and after applying the detected radiated photon correction to the simulated electrons (in the left and right plots, respectively).

## 689 2.7 Background merging

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

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

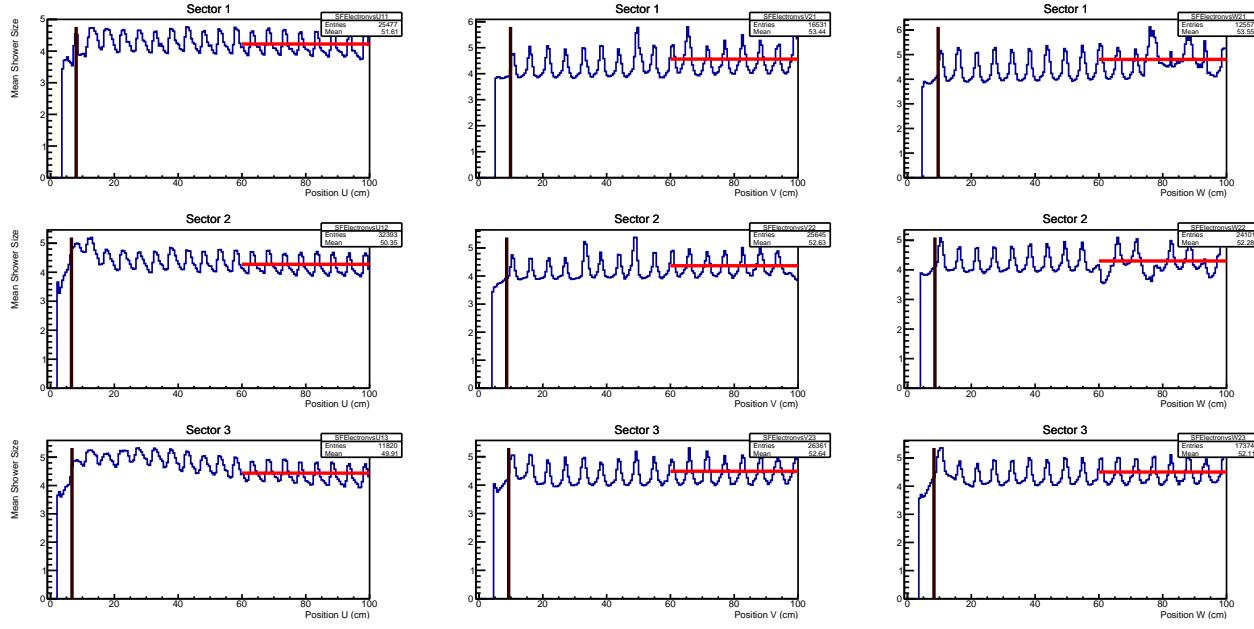


Figure 2.41: Mean shower width as a function of U, V and W (low values) for three sectors of CLAS12. The red line corresponds to the mean shower size calculated away from the edge of the distribution. The black line corresponds to the fiducial cut.

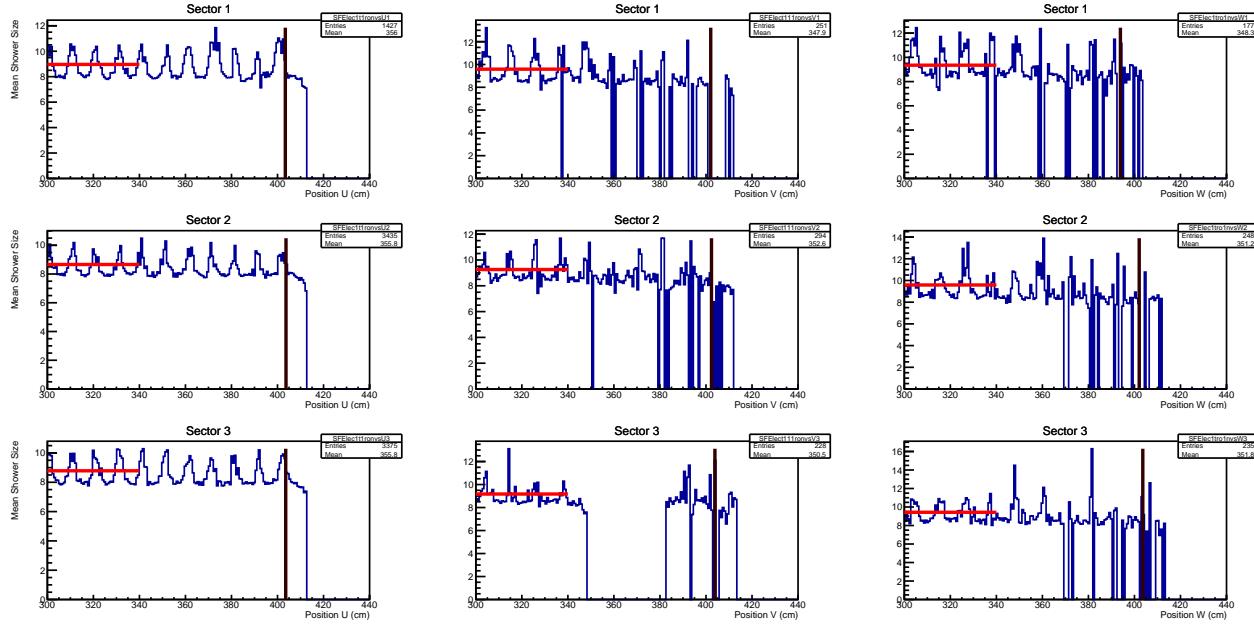


Figure 2.42: Same plots as in Figure 2.41 for the case of high values of U, V and W.

## 2.8 Proton efficiency correction

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

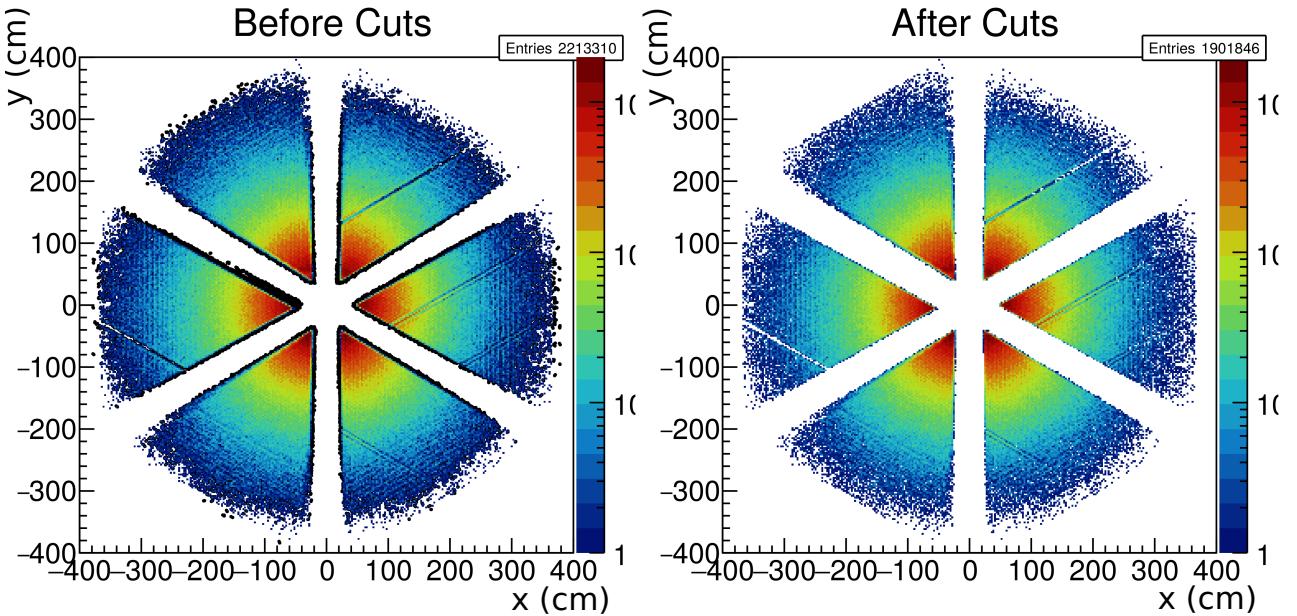


Figure 2.43: Effect of the fiducial cuts on the electron distribution in the natural coordinate plane of the PCAL. The left plot shows the electrons which are removed in black. The plot on the right shows all the electrons that passed the PCAL fiducial cuts.

710 The proton efficiency is measured for data and simulations as:

$$A \cdot Eff^{\text{Data/Simu.}}(\Omega_{Mis.}) = \frac{N_{Rec.}^{\text{Data/Simu.}}(\Omega_{Mis.})}{N_{Mis.}^{\text{Data/Simu.}}(\Omega_{Mis.})}, \quad (2.25)$$

711 where  $N_{Mis.}^{\text{Data/Simu.}}(\Omega_{Mis.})$  is the number of events with a missing proton in the kinematic bin  $\Omega_{Mis.} =$   
 712  $P_{Mis.}; \theta_{Mis.}; \phi_{Mis.}$ ,  $N_{Rec.}^{\text{Data/Simu.}}(\Omega_{Mis.})$  is the corresponding number of events with a detected proton,  
 713 and  $A$  is the geometrical acceptance of CLAS12. The proton efficiency correction is then encoded in  
 714 the ratio:

$$Eff_{Corr} = \frac{A \cdot Eff^{\text{Data}}}{A \cdot Eff^{\text{Simu.}}} = \frac{Eff^{\text{Data}}}{Eff^{\text{Simu.}}}. \quad (2.26)$$

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

### 717 2.8.1 Efficiency correction in the central detector

718 As shown in Figure 2.33b, there is very little background under the missing-proton mass peak in  
 719 the high-polar-angle region. The number of events with a missing proton or a detected proton is then  
 720 given by the number of events in each bin. The integrated efficiencies as a function of the momentum,  
 721 the polar and the azimuthal angles of the missing proton are shown in Figure 2.44. The efficiency  
 722 calculated in the simulations case is higher than for the data. The efficiency correction is calculated  
 723 as a function of the three variables, with 2 bins in  $\theta$  (from  $37^\circ$  to  $45^\circ$  and from  $45^\circ$  to  $65^\circ$ ), 4 bins  
 724 in momentum (spanning the 0.4 to 1.5 GeV range evenly) and 18 bins in  $\phi$  (from  $-180^\circ$  to  $180^\circ$ , 20°  
 725 bins). The limits of the binning are driven by the variation of the correction as a function of each  
 726 variable. Figure 2.45 shows the value of the corrections.

### 727 2.8.2 Efficiency correction in the forward detector

728 The proton efficiency is also derived for FD protons. Unlike in the case of the CD, where there  
 729 is no background under the proton peak, there is a large high-mass background in the missing mass  
 730 spectrum for protons at polar angles below  $37^\circ$ . In this region the number of events with a missing

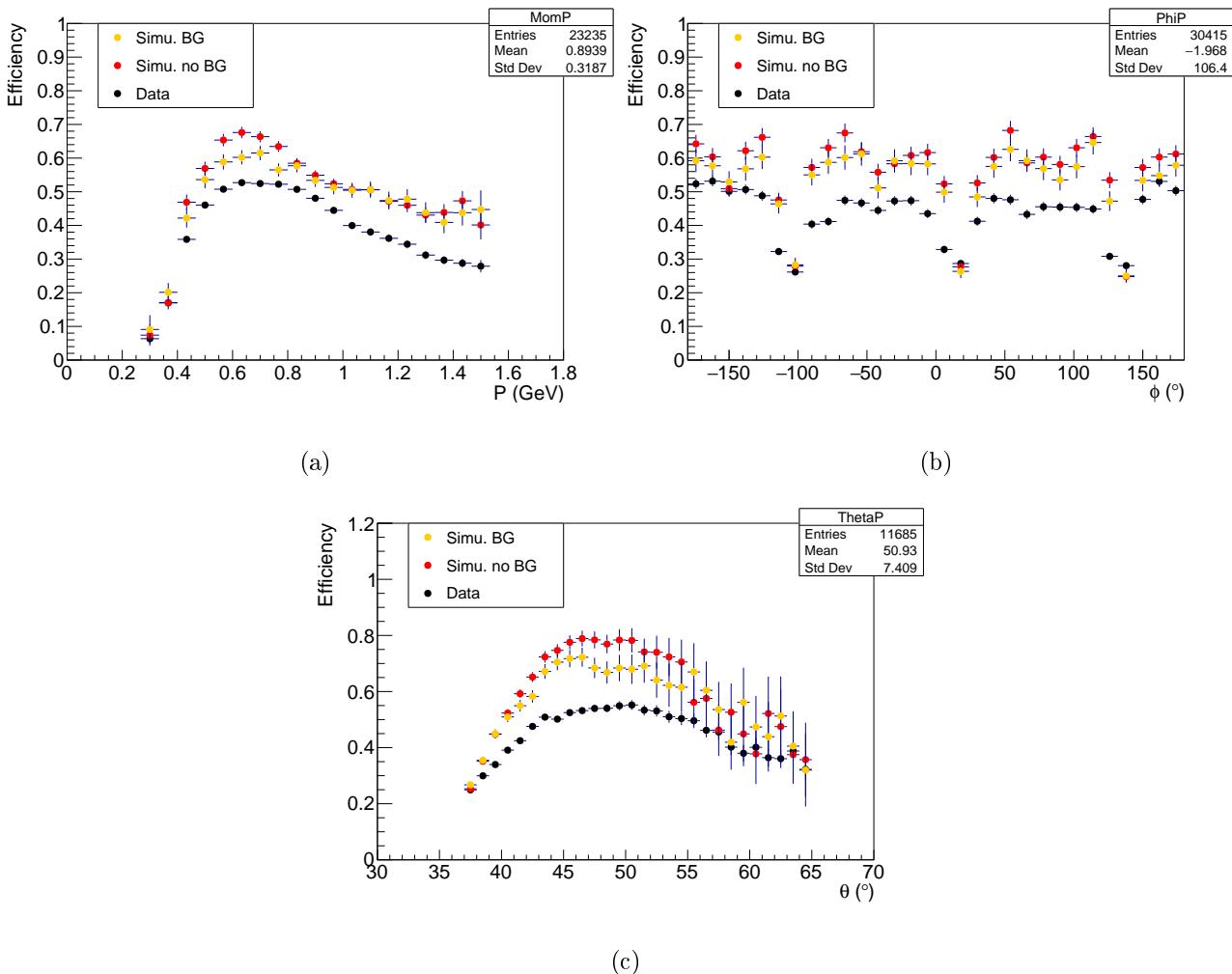


Figure 2.44: Proton efficiency in the CD, as a function of (a) the momentum, (b) the azimuthal angle and (c) the polar angle of the missing proton; for simulations without BG merging in red, with BG merging in yellow and data in black.

731 or reconstructed proton is calculated by fitting the missing proton peak with a gaussian plus a linear  
 732 background as shown in Figure 2.46. The integral of the gaussian defines the number of events. The fit  
 733 on the missing mass peak is done in both the missing-proton case and the reconstructed-proton case.  
 734 In the case of the reconstructed proton, only the events with a reconstructed proton are kept in the  
 735 missing mass spectrum fit.

736 Figure 2.47 shows the efficiency as a function of the momentum, the azimuthal and the polar angles  
 737 of the missing proton. As the ratios between simulations and data efficiencies are fairly constant in  $\theta$   
 738 and  $\phi$ , a single differential correction in momentum is applied according to the efficiencies shown in  
 739 Figure 2.47a. The correction values are given in Table 2.4. Note that the momentum bin [0.3, 0.42] is  
 740 assigned the same correction factor as the next bin as the number of data points is too low to allow  
 741 correct extraction of the efficiency. This change has no effect on the TCS analysis, as most protons  
 742 have momenta above 0.4 GeV in the CD.

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

743 *End of common analysis section*

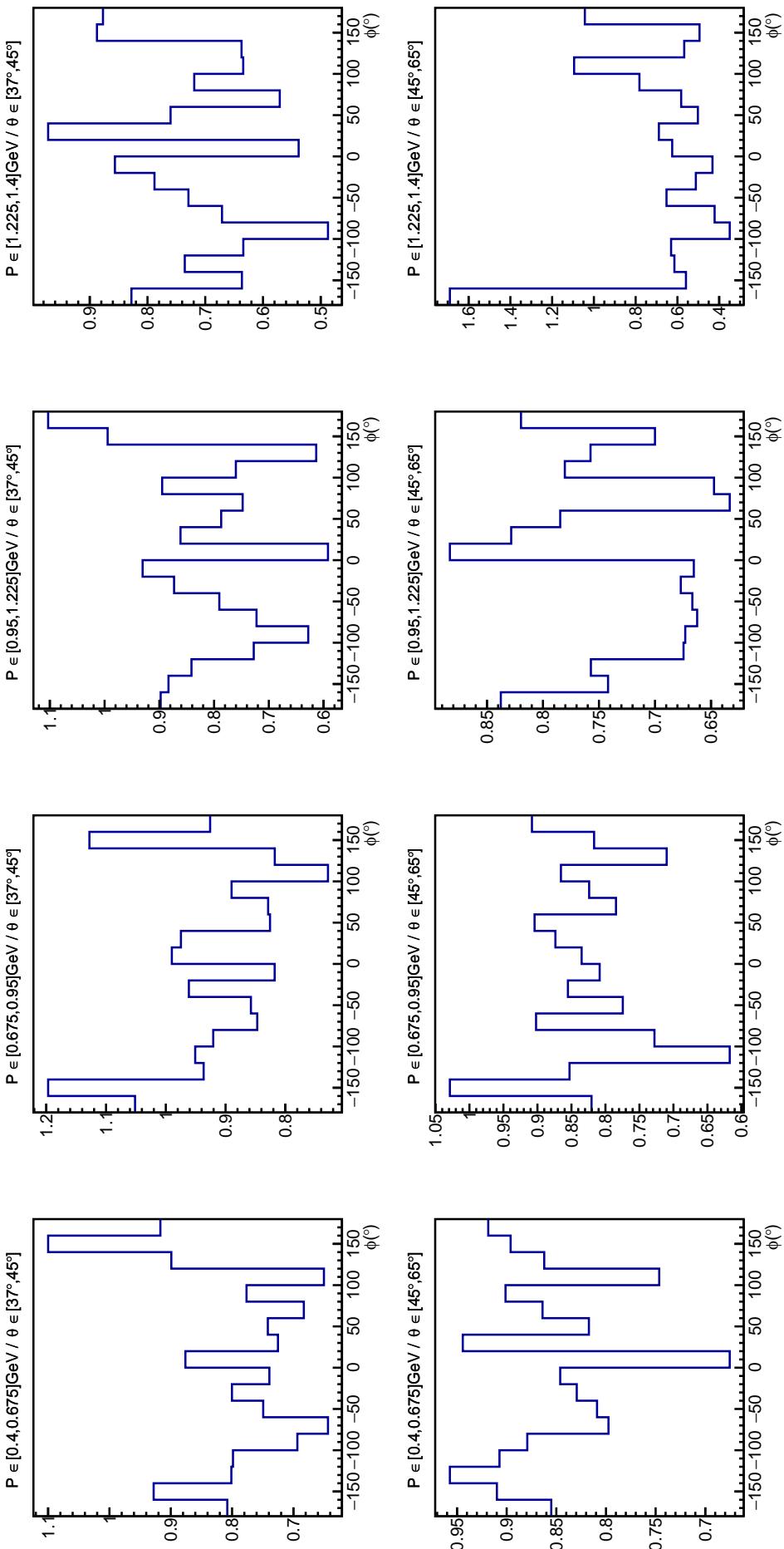


Figure 2.45: Efficiency correction in the CD for two bins in polar angle  $\theta$  and 4 momenta bins. The correction is given as a function of azimuthal angle of the proton  $\phi$ .

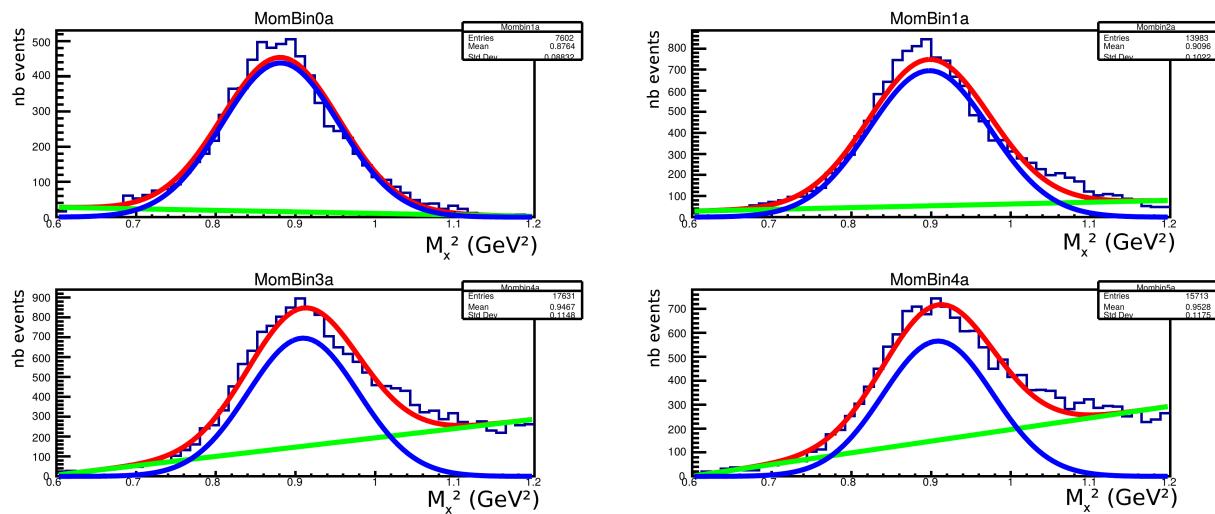


Figure 2.46: Examples of fits performed for the proton efficiency analysis in the FD. The missing mass peak is fitted with a gaussian and a linear background.

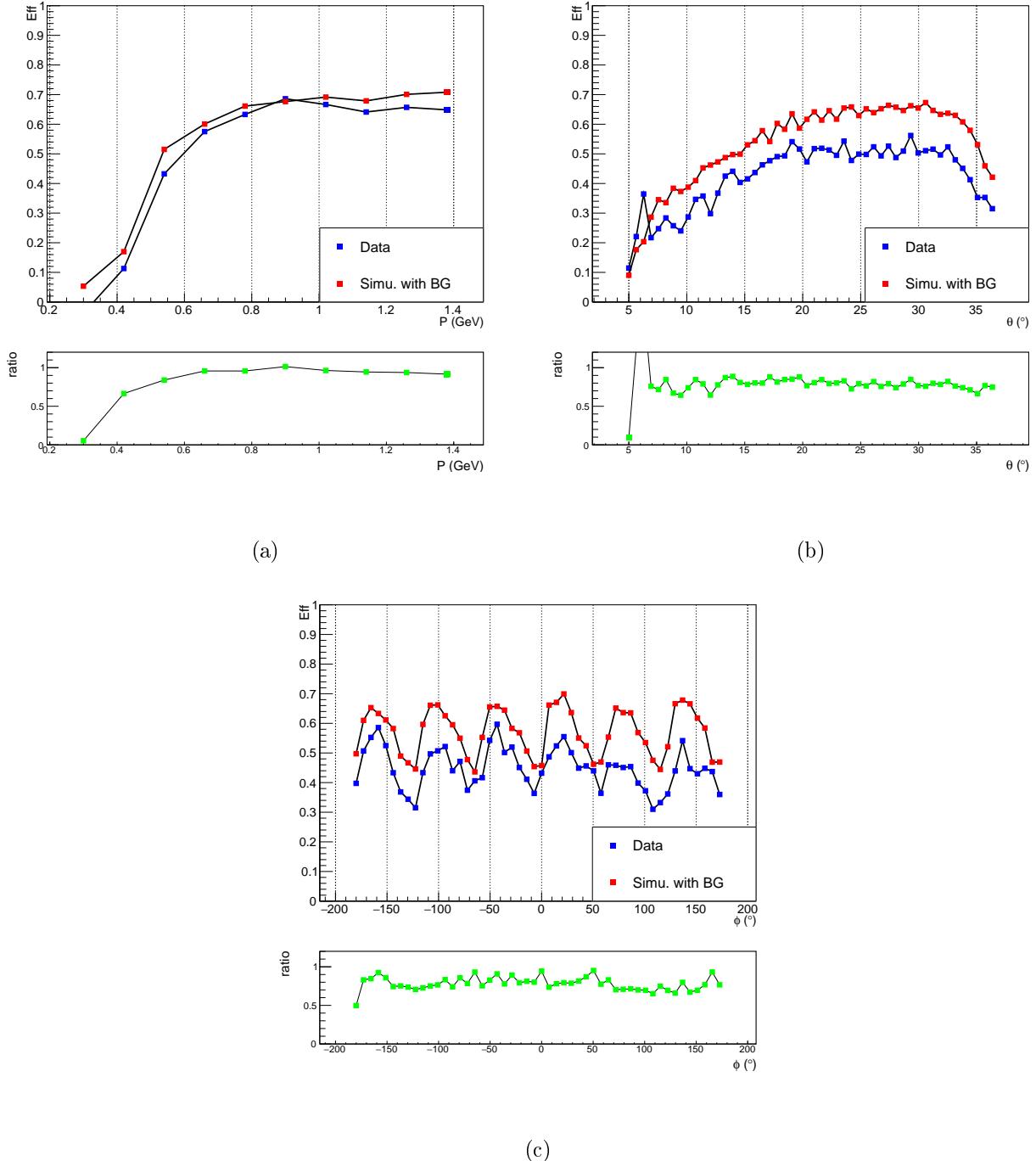


Figure 2.47: Proton efficiency in the FD, as a function of (a) the momentum, (b) the azimuthal angle and (c) the polar angle of the missing proton; for simulations in red and data in blue.



745 Chapter 3

746 747 Simulations and extraction of the TCS  
observables

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

756 3.1 TCS event generator and simulations

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

763 3.1.1 GRAPE

764 The *GRAPE* Generator [9] was developed for experiments running at HERA, at the DESY laboratory  
765 in Hamburg. It is a di-lepton generator to study the electromagnetic background mainly in  $J/\Psi$  and  $\Upsilon$   
766 measurements. It is a full generator including the kinematics of the beam electron producing the real  
767 photon. It also includes pair production from the incident electron as well as interferences between  
768 leptons in the final state. The diagrams which can be included in the generator are shown in Figure  
769 3.11. It is a non-weighted event generator.

770 3.1.2 TCSGen

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

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

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

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

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

789 The cross section  $\sigma$  can be either the BH one or include the TCS interference term. The formulae  
 790 used are those from [1].

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

## 796 3.2 Event selection

### 797 3.2.1 Final state selection

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

### 808 3.2.2 Exclusivity cuts

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

$$\gamma p \rightarrow e^- e^+ p'. \quad (3.3)$$

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

$$ep \rightarrow (e') \gamma p \rightarrow (e') e^- e^+ p'. \quad (3.4)$$

815 The corresponding conservation of momentum equation is:

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

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

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

818

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

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

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

823 where  $E_b$  is the energy of the electron beam,  $E_X$  is the energy of the undetected scattered electron,  
 824  $\theta_X$  is its scattering angle in the lab frame given by:

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

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

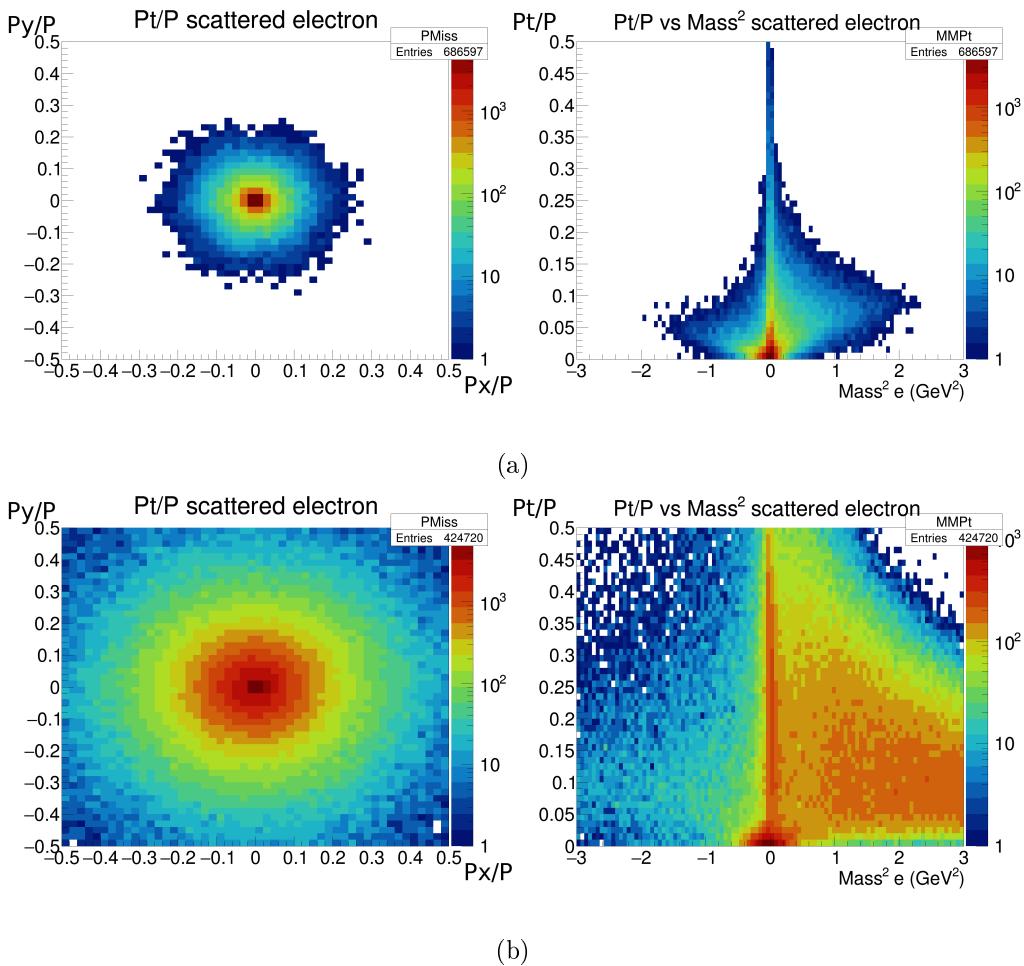


Figure 3.1: Distributions of the exclusivity variables for (a) TCS simulations and (b) the data.

831 The cuts used in the analysis are:

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

832 and

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

833 The resulting distribution of incoming photon virtuality, calculated using Equation (3.8), is shown in  
 834 Figure 3.2. The variation of the extracted quantities with different exclusivity cuts is investigated and  
 835 accounted for in the systematic uncertainties (see Section 3.12).

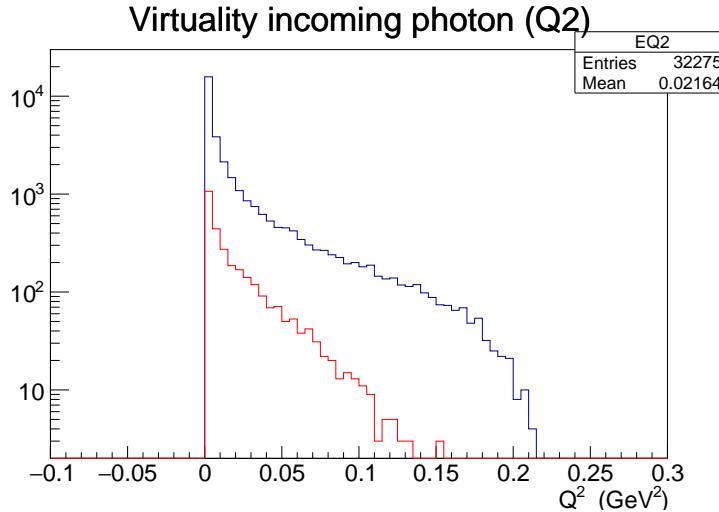


Figure 3.2: Virtuality of the incoming photon for all the events passing the exclusivity cuts (blue) and the events included in the analysis (red).

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

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

840 where  $E_\gamma$  is obtained from Equation (3.6).

841 Another constrained quantity that was investigated is the square of the transferred momentum to  
 842 the proton given by  $t = (p_{p'} - p_p)^2$ . Assuming the real incoming photon is along the beam direction  
 843 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  
 844 the measured momenta of the final state leptons and only the z-component of the measured proton as  
 845  $t = (p_{\gamma \text{ approx.}} - p_{e^+} - p_{e^-})^2$ . The difference between the two calculations given by:

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

846 allows to assess the exclusivity of the final events.

847 Figure 3.3 shows  $\Delta E$  and  $\Delta t$  for all events with the TCS final state, events that passed the  
 848 exclusivity cuts and final events that are included in the analysis after kinematic cuts (more details in  
 849 Section 3.11). One can see that the exclusivity cuts strongly constrain these quantities.

### 850 3.3 Phase space of interest and Simulations/Data comparison

851 After applying the exclusivity cuts detailed above, the invariant mass of the lepton pair is extracted.  
 852 Figure 3.4 shows the invariant mass spectrum obtained for the full RGA Fall 2018 dataset. Meson  
 853 resonances decaying into an electron-positron pair are visible.

854 The mass region between 1.5 GeV and the  $J/\psi$  mass (3096 MeV) is selected for the measurement of  
 855 TCS. The first reason for this choice is theoretical: the hard scale of TCS is given by the invariant mass  
 856 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  
 857 GPD factorization. As shown in the later analysis, a typical value of proton momentum transfer in this

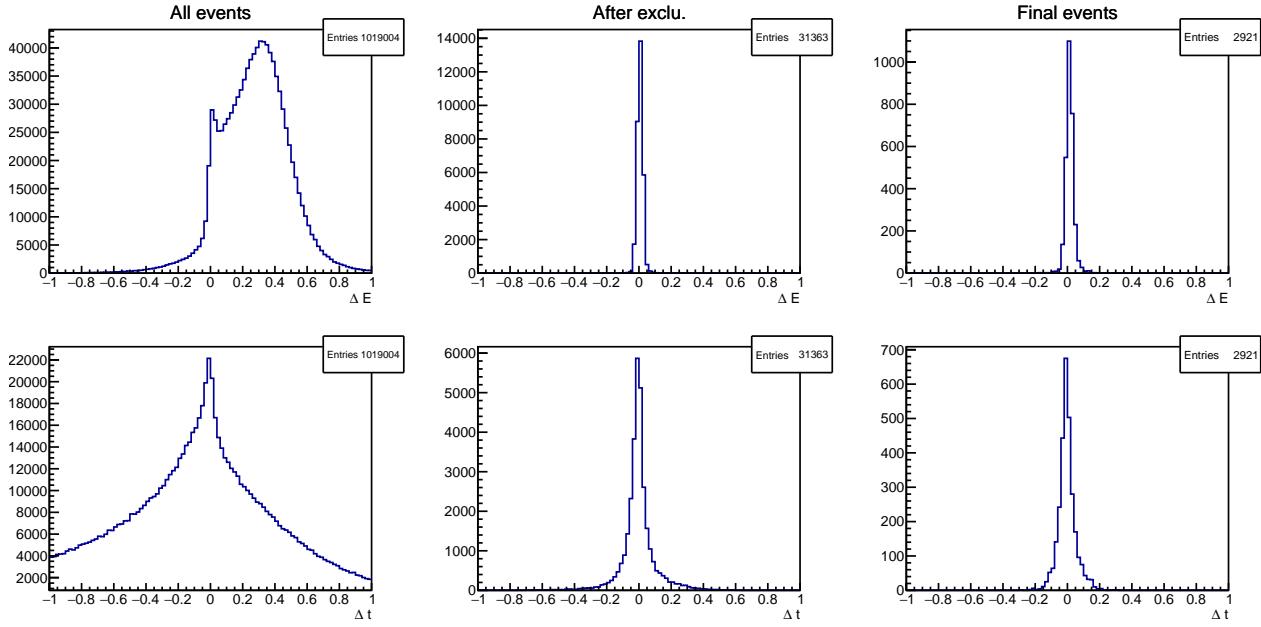


Figure 3.3:  $\Delta E$  and  $\Delta t$  distributions as defined in Equations (3.12) and (3.13), for all events with the final state of interest in the inbending data set (left), after exclusivity cuts (middle), and for events included in the final analysis after kinematic cuts (right).

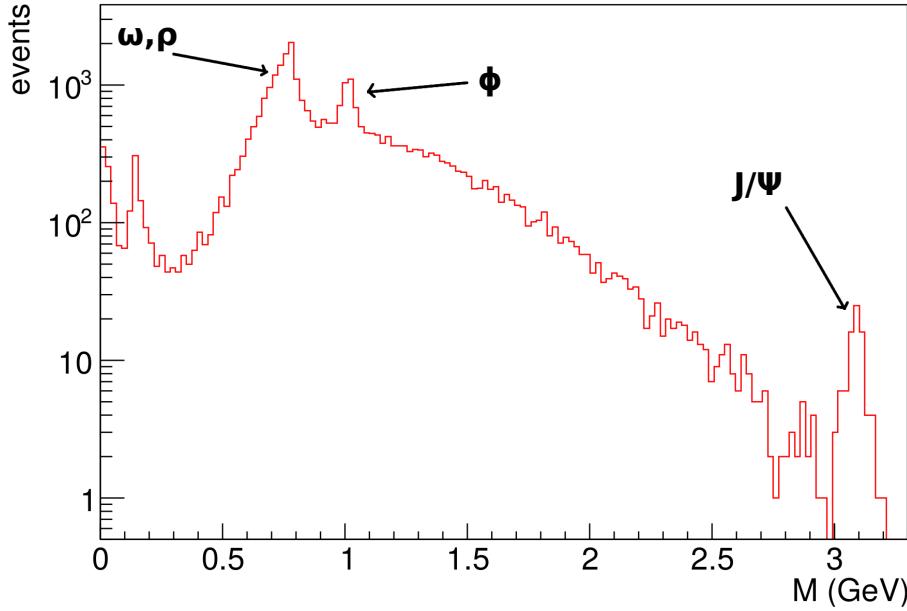


Figure 3.4: Lepton pair invariant mass spectrum after exclusivity cuts, extracted from data. Meson resonances ( $\omega/\rho$  (770/782 MeV);  $\phi$  (1020 MeV) and  $J/\psi$  (3096 MeV)) are visible. The peak at zero mass is due to photon conversion in the target material ( $\gamma \rightarrow e^+e^-$ ) and  $\pi^0$  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.

analysis is 0.3 GeV<sup>2</sup>, 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  $\rho$  (1450) and  $\rho$  (1700). In order to check the impact of these resonances, the mass spectrum obtained from the data is compared with

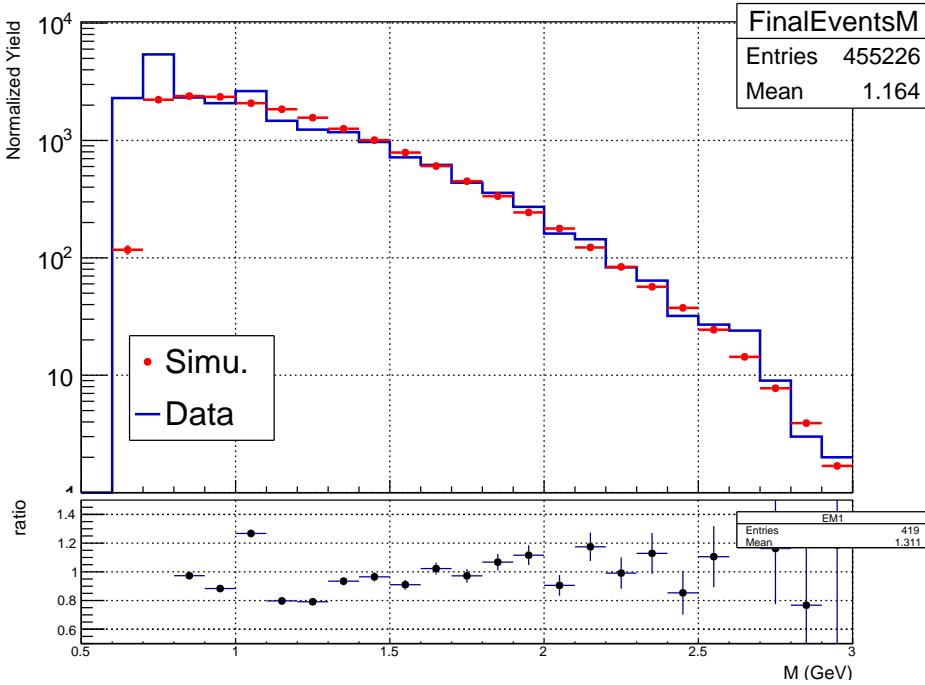


Figure 3.5: Comparison between data and simulations in the high-mass range. The ratio between the two plots is also shown. Cuts on the square of the transferred momentum ( $0.15 \text{ GeV}^2 < -t < 0.8 \text{ GeV}^2$ ) and the photon energy ( $2 \text{ GeV} < E_\gamma < 10 \text{ GeV}$ ) are applied.

862 BH-weighted simulations, between the  $\phi$  and the  $J/\psi$  mass, in Figure 3.5. In this figure the simulated  
 863 spectrum is normalized in order to have equal integral as the data spectrum, between 1.1 and 3 GeV.  
 864 According to this plot there is no obvious meson contamination in the mass range of interest. The  
 865 effect of this contamination on the observables is studied by extracting them in the two mass ranges  
 866 ([1.5 GeV – 2 GeV] and [2 GeV – 3 GeV]).

867 Eventually, the phase space selected for the analysis is:

- 868 •  $1.5 \text{ GeV} < M = \sqrt{Q'^2} < 3 \text{ GeV}$
- 869 •  $0.15 \text{ GeV}^2 < -t < 0.8 \text{ GeV}^2$
- 870 •  $4 \text{ GeV} < E_\gamma < 10.6 \text{ GeV}.$

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

872 The distributions of the three variables  $-t$ ,  $E_\gamma$  and  $M$  for data and simulations in the selected  
 873 phase space are compared in Figure 3.6. The simulation distributions are normalized to have the same  
 874 integral as data distributions. The simulations-data agreement for all three variable is good and no  
 875 large discrepancies are seen. The kinematics of the particles in the laboratory coordinates  $P$  and  $\theta_{Lab}$   
 876 are displayed in Appendix C.

### 877 3.4 Acceptance estimation

878 The large angular coverage of CLAS12 permits to detect most of the particles produced when an  
 879 electron interacts with the target. However, due to holes between detector subsystems, some particles  
 880 may escape without being detected. Furthermore, because of detection inefficiencies, the reconstructed  
 881 particle kinematics may differ from their initial values at the vertex. In order to correct for these effects,  
 882 the acceptance of CLAS12 for the  $\gamma p \rightarrow e^- e^+ p'$  reaction is estimated using simulations including  
 883 background merging. In the following, acceptance refers to the product of the geometrical acceptance

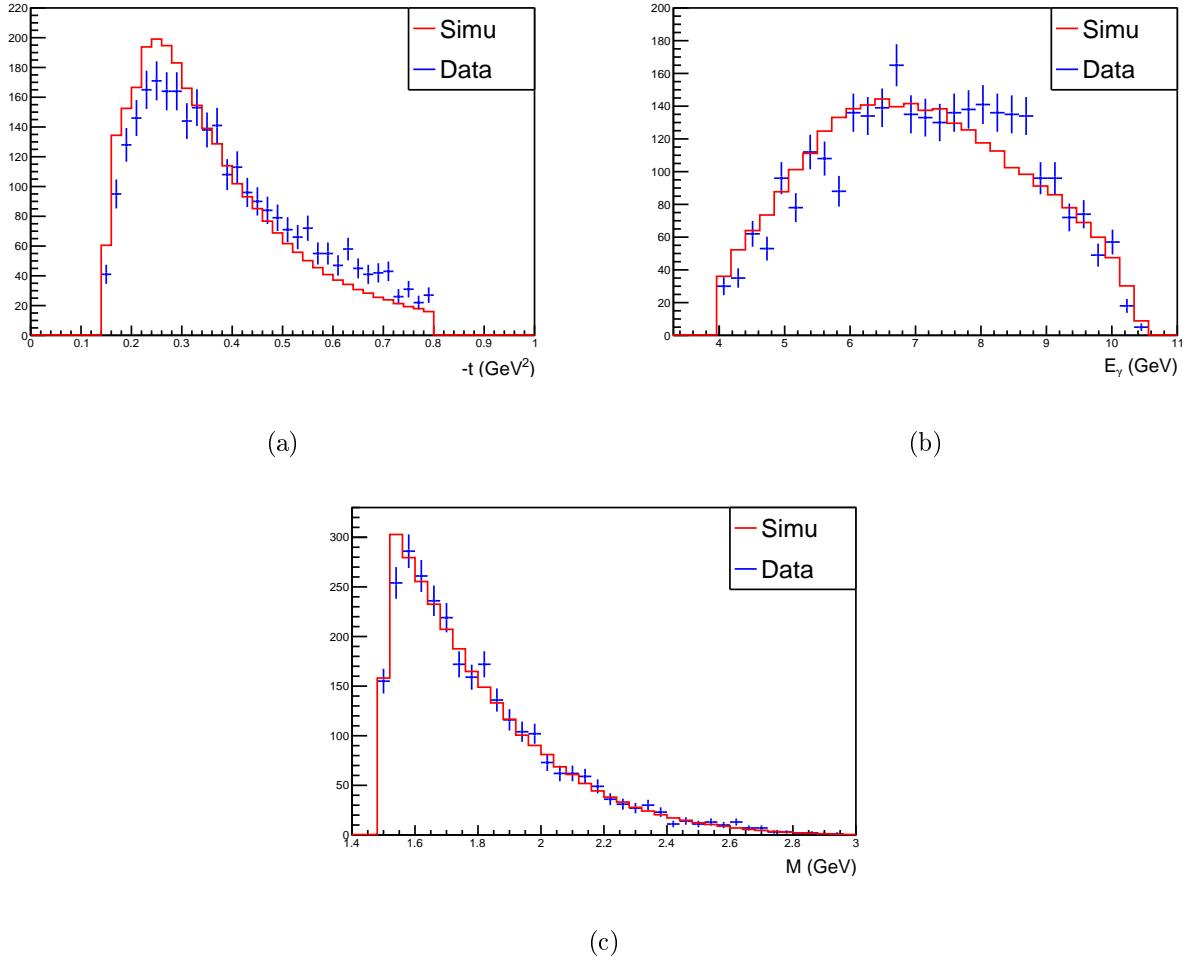


Figure 3.6: Simulations-data comparison for the distributions in (a)  $-t$ , (b)  $E_\gamma$  and (c)  $M$ . The simulation distributions are normalized to have equal integral as the data. Both the simulations and the data are in agreement, and no large discrepancies are seen.

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

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

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

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

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

896 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
$E_\gamma$	4.0 - 6.2 - 8.4 - 10.6
$Q'^2$	2.25 - 3.5 - 5.0 - 9.0
$\theta$	from $30^\circ$ to $160^\circ$ , $10^\circ$ bins
$\phi$	from $-180^\circ$ to $180^\circ$ , $10^\circ$ bins

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

897 For the results shown in Section 4.1, 36 million events were generated using *TCSGen*, passed  
 898 through GEMC and reconstructed using the CLAS12 reconstruction software. In order to remove bins  
 899 where there are too few events to perform the acceptance estimation, two additional cuts are used.  
 900 The bins with acceptance below 5% are removed from the analysis. A cut on the relative error of the  
 901 acceptance shown in Figure 3.7 is also applied:

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

902 The resulting acceptance distributions are shown in Figures 3.8, 3.9, and 3.10, as a function of  $\theta$  and  
 903  $\phi$  for each  $\{E_\gamma, -t, Q'^2\}$  bin.

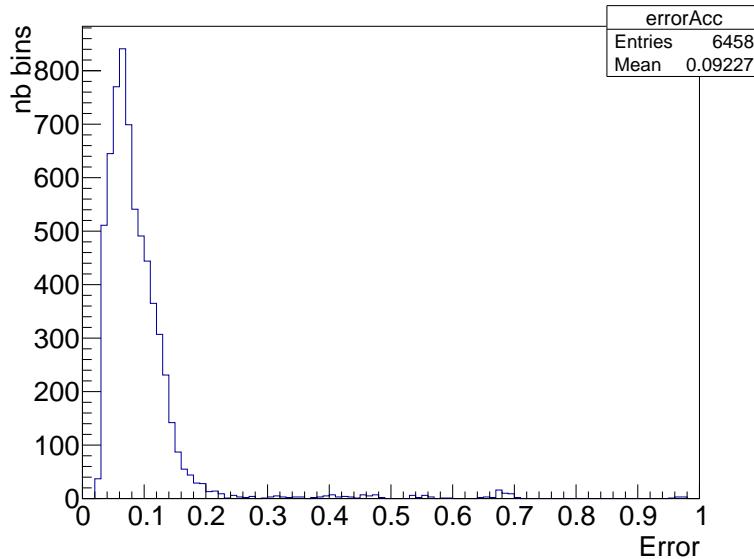
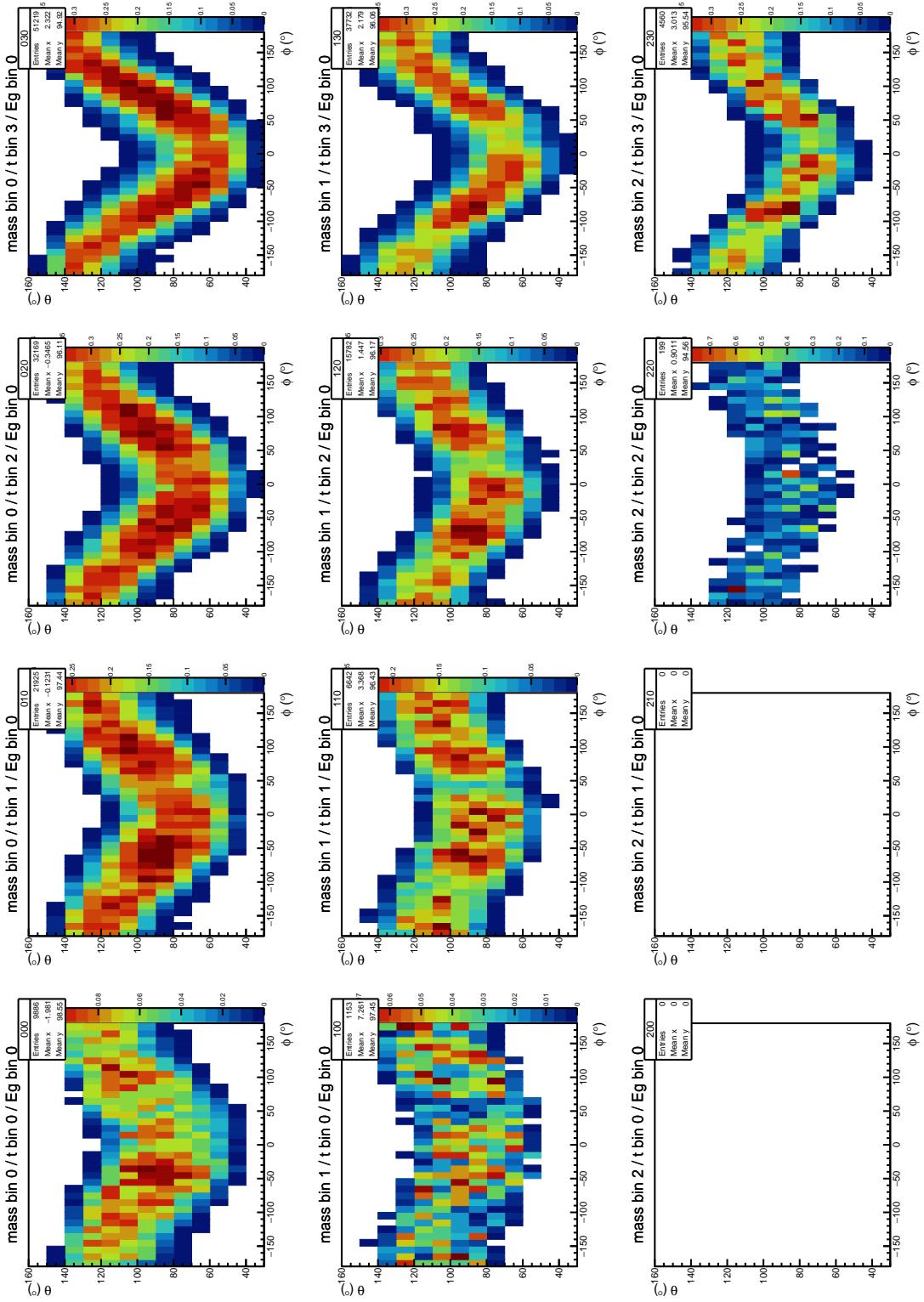


Figure 3.7: Relative error on the acceptance calculation, after bins with acceptance below 5% are removed. A cut is applied to remove bins with relative error higher than 50%.

904 Figures 3.8, 3.9 and 3.10 show the acceptance of CLAS12 for the TCS process. Each plot in the  
 905 grids corresponds to a single  $\{E_\gamma, -t, Q'^2\}$  bin, as defined in Table 3.1 in Chapter 3. The acceptance  
 906 plots shown here are made prior to applying the acceptance quality cuts described in Section 3.4.


 Figure 3.8: CLAS12 acceptance for TCS in the  $\phi/\theta$  plane for the first bin in  $E_\gamma$ .

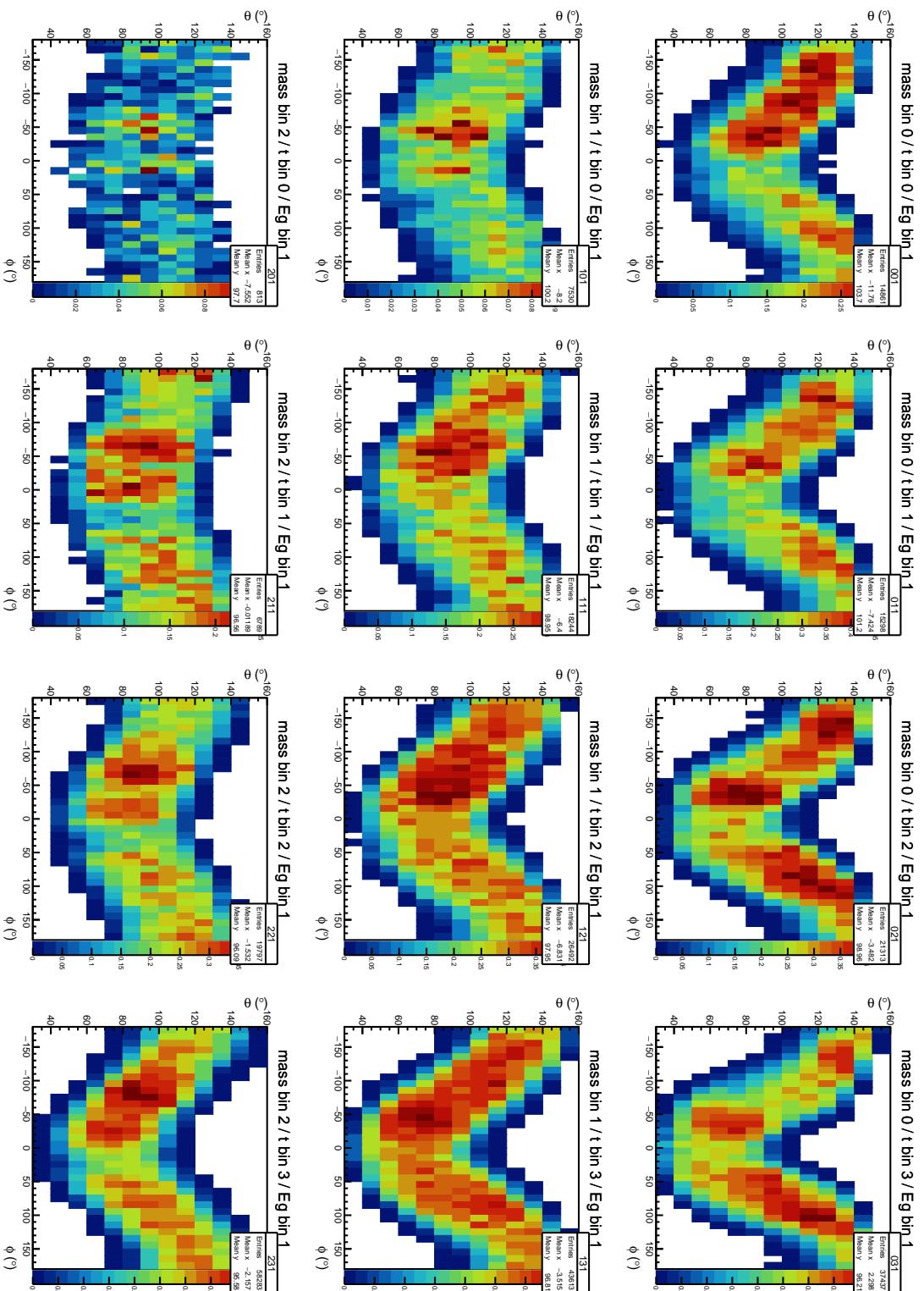


Figure 3.9: CLAS12 acceptance for TCS in the  $\phi/\theta$  plane for the second bin in  $E_\gamma$ .

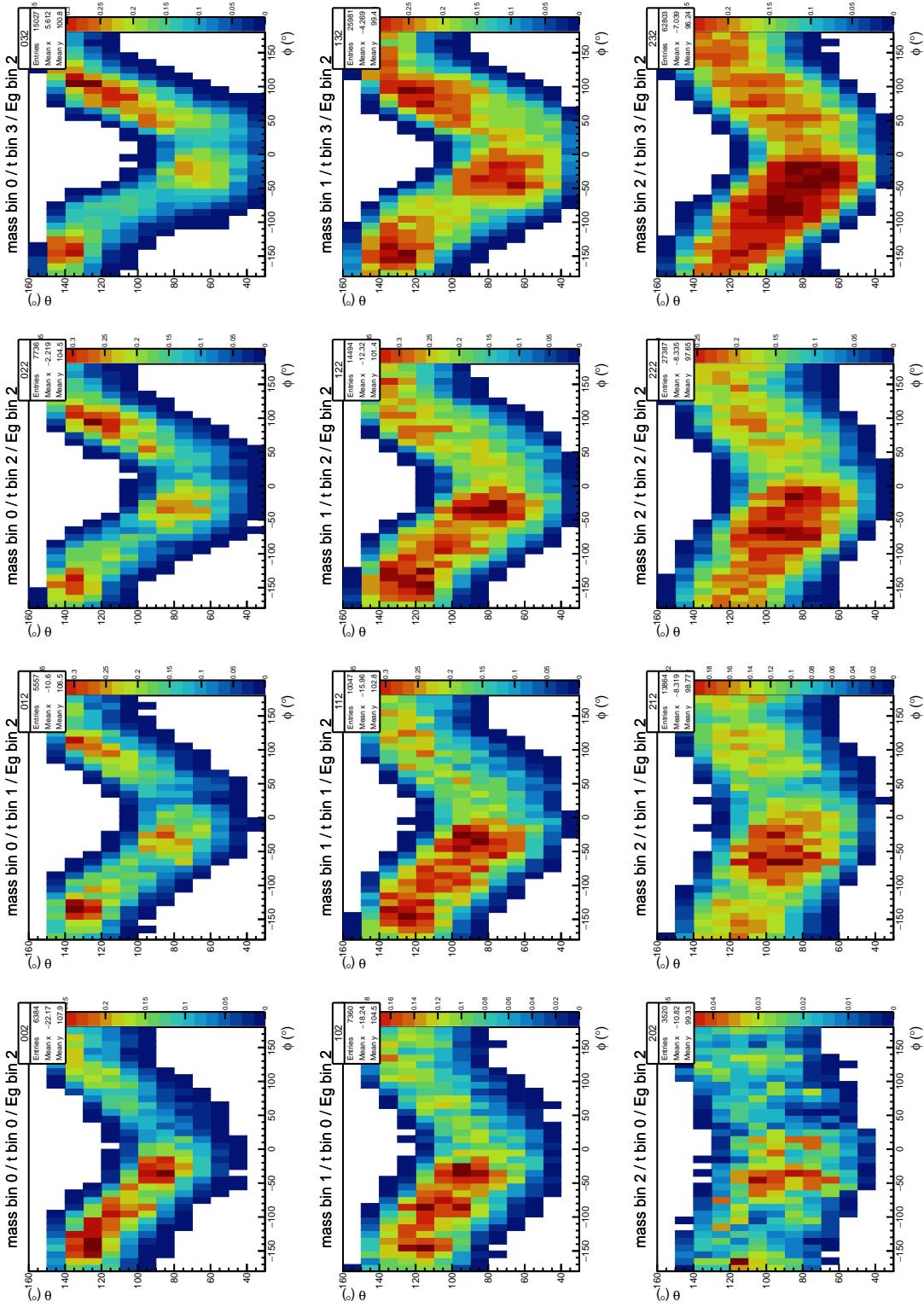


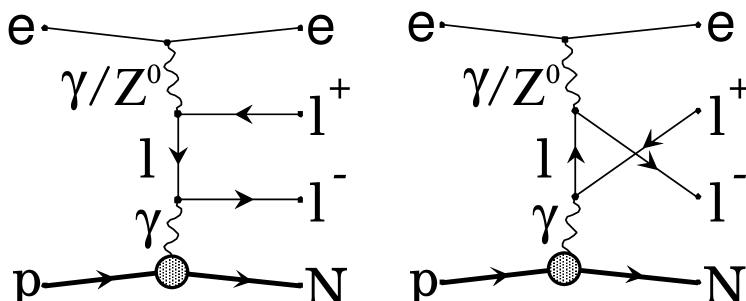
Figure 3.10: CLAS12 acceptance for TCS in the  $\phi/\theta$  plane for the third bin in  $E_\gamma$ .

## 3.5 Background estimations

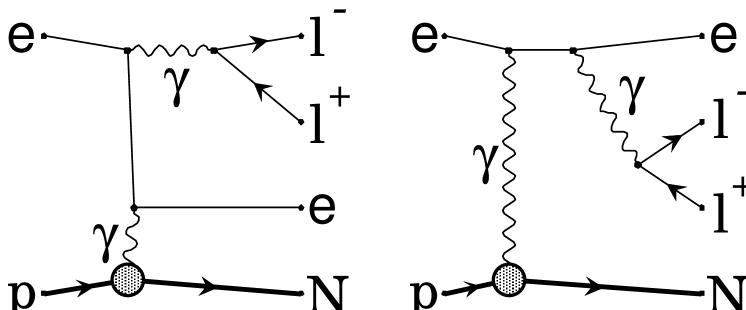
### 3.5.1 Electro-production of a lepton pair $ep \rightarrow pe^+(e^-)e_{scattered}$

#### 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 3.11), the pair production from radiated photons (bottom plots in Figure 3.11), 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 3.11: 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 3.12a 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  $Q_2'^2$  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 3.12b 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.

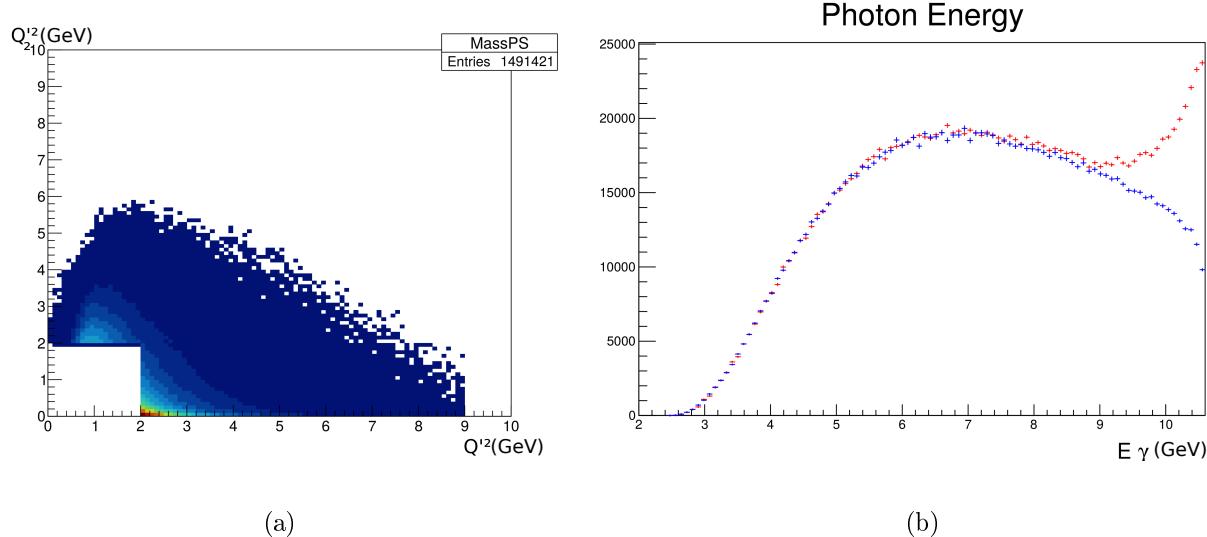


Figure 3.12: (a) Generated invariant mass phase space in the case where all possible processes leading to the final state  $pe^+(e^-)e_{\text{scattered}}$  are considered. (b) Generated photon energy distributions for BH-only events (blue) and including all the possible processes (red).

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 3.13. No large effect coming from the use of quasi-real photo-production is noticed.

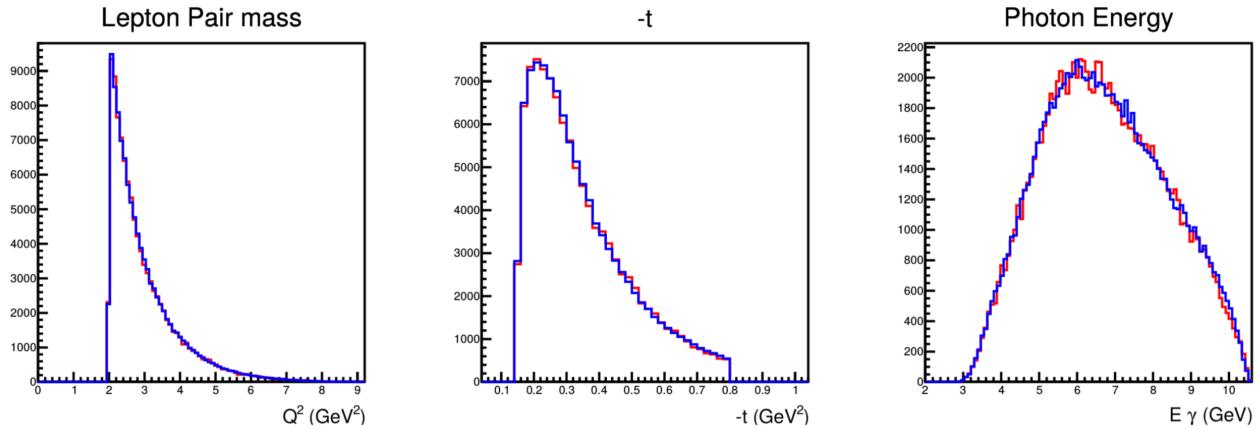
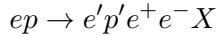


Figure 3.13: Measured invariant mass  $Q'^2$  (left), proton transferred momentum  $-t$  (middle) and photon energy  $E_\gamma$  (right) distributions, for the generated BH-only events (blue) and with all the processes included (red).

### 937 From data

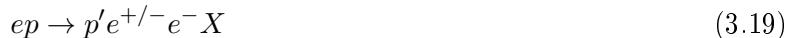
**Fraction of events with scattered electron** Events in electroproduction of  $(e^+e^-)$  pair contain two electrons in the final state. Final states with lepton pairs can come not only from a time-like photon or a vector meson decay but also bremsstrahlung photon conversions in the target or the detector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

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



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

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



Final states were identified using the even builder particle IDs with simple cuts developed for TCS and  $J/\psi$  analysis:

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

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

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

In Figures 3.14 and 3.15, the invariant mass distributions of lepton pairs for the "inbending" and "outbending" data sets are shown, respectively. In each figure, top graphs are the invariant mass distributions of  $e^+ e^-$  for  $e^+ e^- p$  final state. The lower graphs are the invariant mass distributions of two electrons in  $e^- e^- p$  final state. On the graphs, the total number events in each case and the number of events with  $M(ee) > 1.5$  GeV are shown as well.

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

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

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

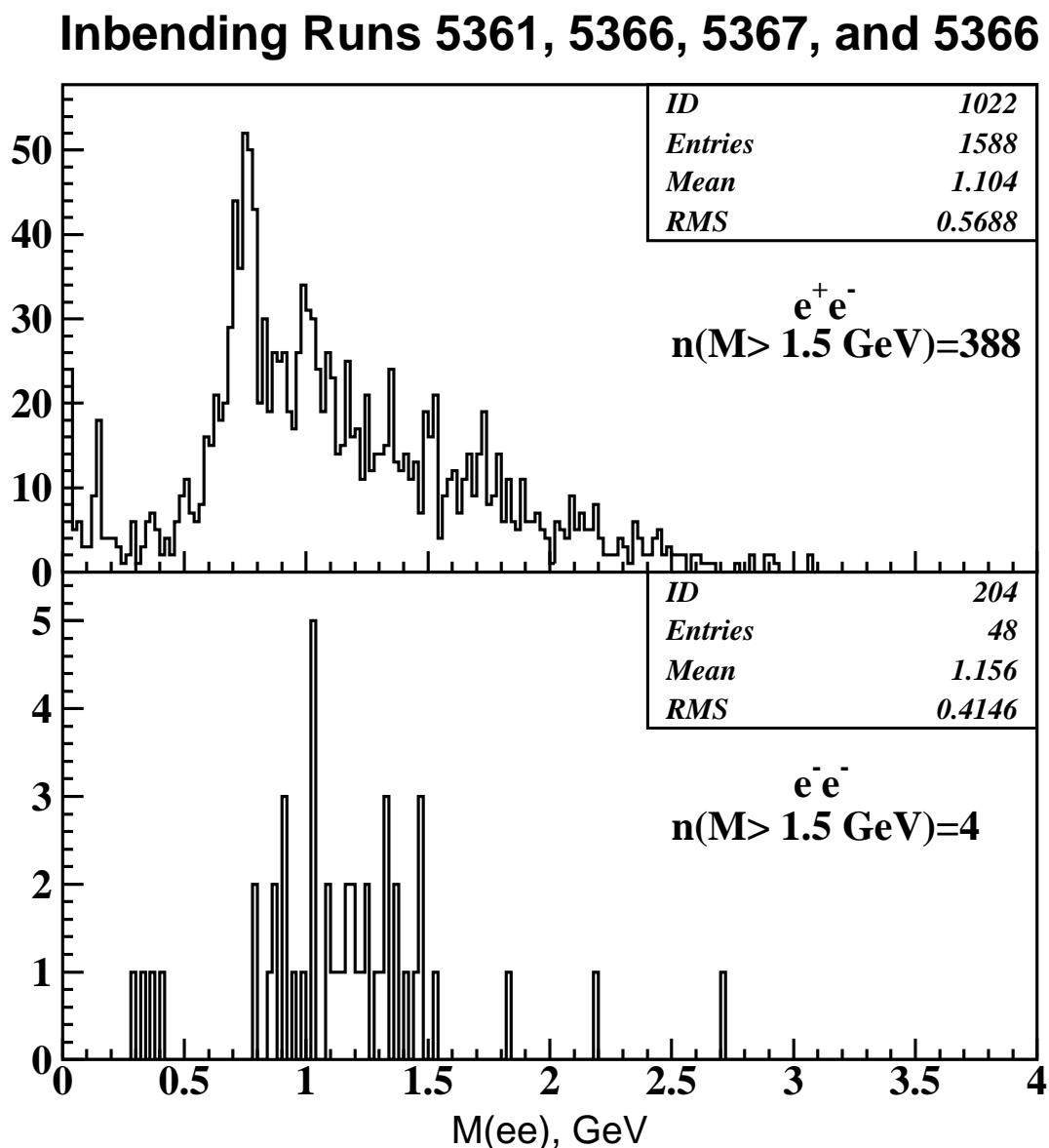


Figure 3.14: The invariant mass distributions for  $(e^+e^-)$ , top, and  $(e^-e^-)$ , bottom, pairs from "in-bending" data. Events in both samples were selected with the same conditions, final state  $eep$ , and exclusivity cuts.

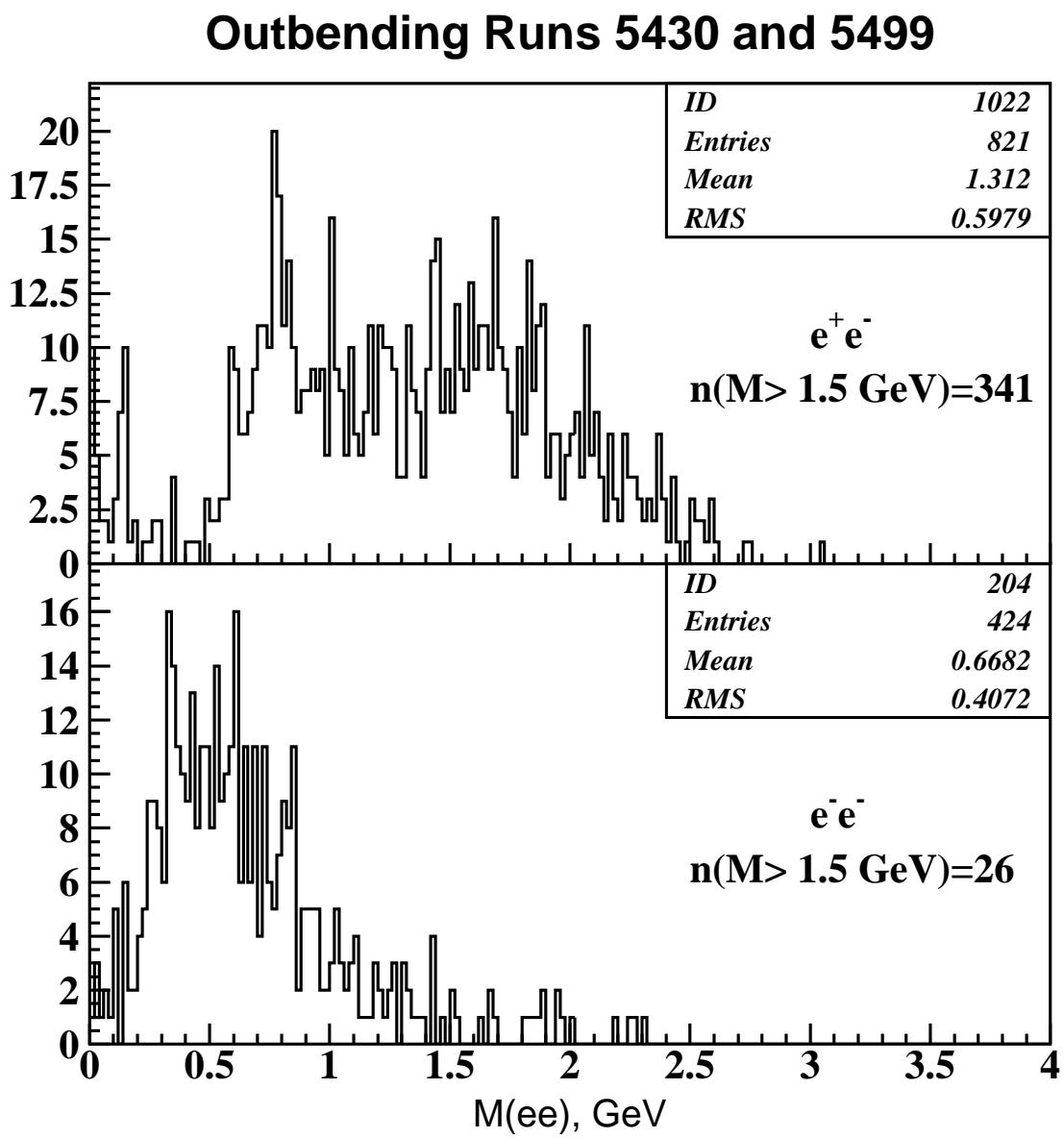


Figure 3.15: The invariant mass distributions for  $(e^+e^-)$ , top, and  $(e^-e^-)$ , bottom, pairs from "outbending" data. Events in both samples were selected with the same conditions, final state  $eep$ , and exclusivity cuts.

977 Then the super ratio will be -

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

$$\frac{a}{b} = 0.24$$

978 Now if we look the number events in each case with the invariant mass of letons  $M_{ee} > 1.5$  GeV, and  
 979 correct for the differences of detection efficiency for *in-bending* and *out-bending* tracks, the fraction of  
 980  $e^+e^-$  where  $e^- \equiv e'$  in "inbending" sample is  $1.0\% \pm 0.5\%$ , for the "outbending" data it is  $1.8\% \pm 0.3\%$ .  
 981 Within the measurements statistical errors two results are consistent, and we safely can conclude that  
 982 the fraction of such background is  $< 2\%$ .

### 983 3.5.2 Pion contamination

984 Pions can be mistakenly identified as leptons, especially at high momenta (above  $\sim 4.5$  GeV) where  
 985 the HTCC produces signals for leptons and pions. The pion background is addressed in Chapter 2.  
 986 The neural network PID is implemented in the TCS analysis with a cut at 0.5. We show in Chapter  
 987 2 that this method reduces the background ratio by a factor 10, to less than 5%. To account for  
 988 the remaining background and to estimate its impact on the measured observables, the background  
 989 rejection cut is varied. This latter point is addressed in the Section 3.12.

### 990 3.5.3 $J/\psi$ contamination

991 Another potential background of this analysis is  $J/\psi$  photo production events, where the invariant  
 992 mass of the  $J/\psi$  is falsely reconstructed below 3 GeV due to radiative effects. To assess this background,  
 993 Monte-Carlo simulations are used. JThese simulations were done by Joseph Newton and show that 8%  
 994 of the total number of generated  $J/\psi$  are reconstructed in the radiative tail below 3 GeV. With the  
 995 exclusivity cuts of this analysis, there are 82 events above 3GeV (all coming from Jpsi). Thus there  
 996 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  
 997 events used in the analysis). This is much lower than the potential contamination from  $\pi^+$  and can be  
 998 ignored.

## 999 3.6 Experimental cross-section ratio

1000 The theoretical  $R$  ratio is calculated over the full angular phase space of TCS. However CLAS12  
 1001 acceptance is limited and only covers a fraction of the angular phase space. The theoretical  $R$  ratio  
 1002 is thus inaccessible, but it can be calculated over the CLAS12 acceptance. Following the notations in  
 1003 [3], we call  $R'$  the ratio integrated over the CLAS12 acceptance. It is calculated as:

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

1004 where the sum over  $\phi$  is done in  $10^\circ$  bins and the  $\cos(\phi)$  factor is calculated at the center of each bin.  
 1005 The  $Y_{\phi}$  quantity is calculated for each  $\phi$  bin as:

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

1006 where the ratio  $\frac{L}{L_0}$  is calculated event-by-event (the factors  $L$  and  $L_0$  are given in [1]), respectively,  
 1007 and  $Acc$  is the acceptance in the kinematic bin of each event (given in Equation (3.14)). The statistical  
 1008 error of  $Y_{\phi}$  is calculated as:

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

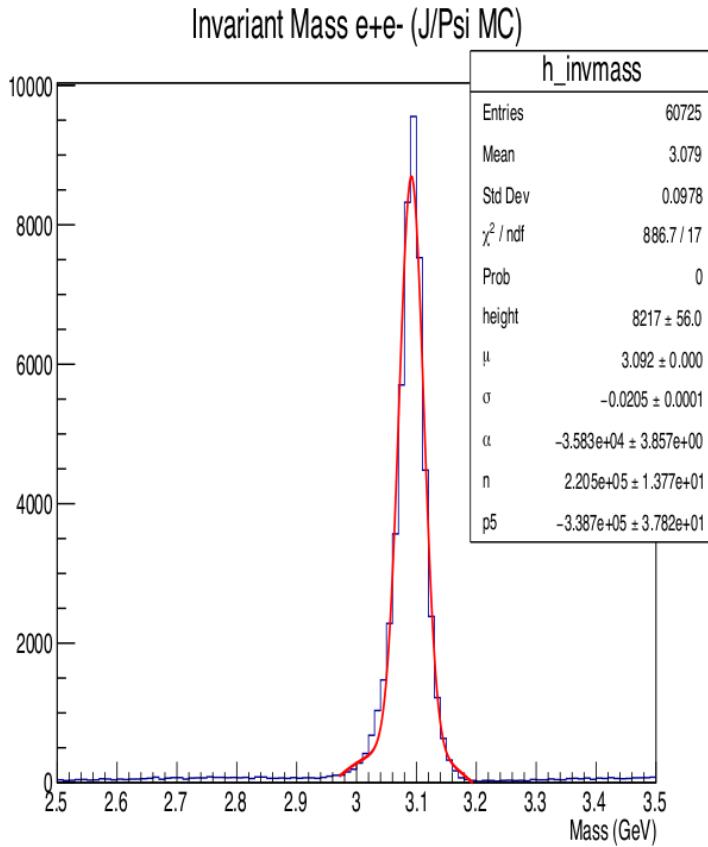


Figure 3.16:  $J/\psi$  simulation (done by J.Newton). 8% of the initially generated events are reconstructed below 3 GeV.

#### 1009 Statistical uncertainty using Monte Carlo method

1010 As the numerator and the denominator used in the computation of  $R'$  are correlated, a Monte Carlo  
1011 approach is used to compute the statistical uncertainty. It works as follows:

- 1012 •  $Y_\phi$  and  $E(Y_\phi)$  are computed as described above,
- 1013 • for each  $\phi$  bin, a value  $Y_\phi^R$  is randomly generated following a gaussian probability distribution of  
1014 mean  $Y_\phi$  and sigma  $E(Y_\phi)$ ,
- 1015 • the ratio is calculated using the randomly generated values:  $R^R = \frac{\sum_\phi Y_\phi^R \cos(\phi)}{\sum_\phi Y_\phi^R}$ .
- 1016 • The previous two steps are repeated 10000 times and the resulting distribution of  $R^R$  is fitted  
1017 with a gaussian,
- 1018 • the  $\sigma$  of the fitted gaussian is defined as the statistical uncertainty of the measurement. An  
1019 example of the fitted distribution is shown in Figure 3.17.

#### 1020 3.7 Phenomenological study of the TCS Forward-Backward asymme- 1021 try

1022 No predictions have yet been published for the Forward-Backward asymmetry ( $A_{FB}$ ) in TCS. In  
1023 this section the phenomenological work performed on the TCS  $A_{FB}$  is presented. The main goal of  
1024 this analysis is to estimate the size of the asymmetry as well as to gain an insight on its kinematic  
1025 dependencies in order to determine if CLAS12 can measure it.

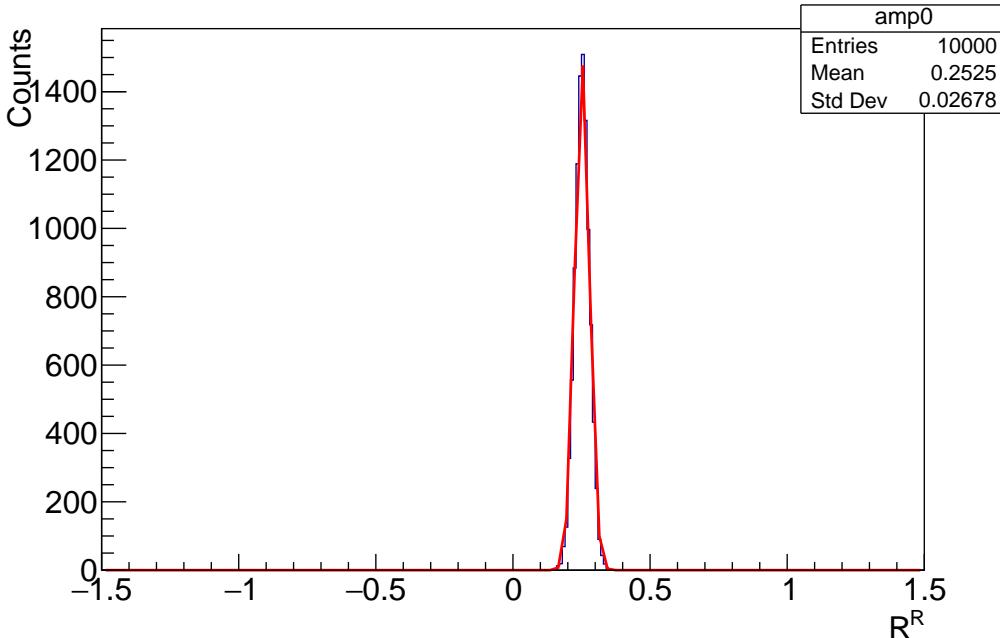


Figure 3.17: Distribution of  $R^R$  for the determination of the statistical error for first bin in  $\xi$ . The full results are shown in Figure 4.2. The gaussian fit is overlaid.

1026 This section is divided in three parts. In Subsection 3.7.1, the limitations of CLAS12 acceptance  
 1027 and the consequences on the measurement of the  $A_{FB}$  are presented. Initial results are compared to  
 1028 other predictions provided by independent groups and models. The results obtained according to the  
 1029 conclusion of the first section are then displayed in subsequent Subsections 3.7.2 and 3.7.3.

### 1030 3.7.1 Early considerations and comparison with other models

1031 To study the TCS  $A_{FB}$ , the TCS and BH processes and their interference cross sections have been  
 1032 calculated using the VGG code [13] provided by Michel Guidal. The formulas used in this code are  
 1033 described in [2]. The cross section is estimated in the forward and backward directions and the obtained  
 1034 results are then combined to produce the asymmetry curves shown in the following. The GPD model  
 1035 used is the VGG model described in [14]. The D-term contribution to the GPD  $H$  is included (unless  
 1036 mentioned otherwise). The  $\alpha'$  coefficient fixing the  $t$ -dependence of the GPDs is set to 1.098, the  $b_{val}$   
 1037 and  $b_{sea}$  parameters fixing the  $\xi$  dependence are set to 1 (unless specified otherwise).

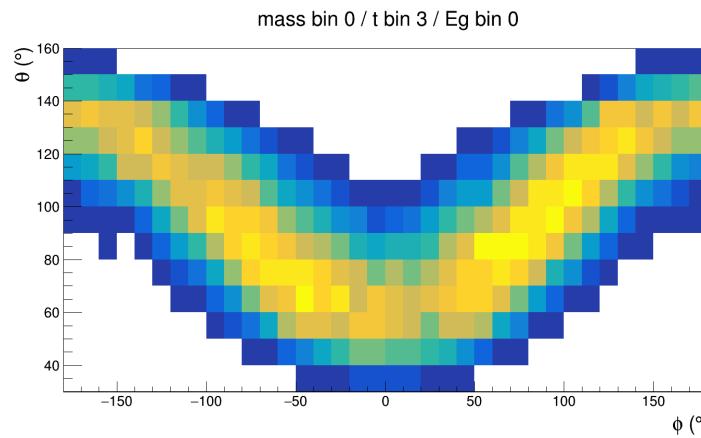


Figure 3.18: Acceptance of CLAS12 in the  $\theta/\phi$  plane. The region around  $\phi = 0^\circ$  and low polar angle, as well as  $\phi = \pm 180^\circ$  and high polar angle are well covered by CLAS12.

1038 The CLAS12 acceptance does not cover the full  $\phi - \theta$  plane of TCS. An example of the acceptance  
 1039 coverage, obtained from the CLAS12 acceptance study presented in Section 3.4, is shown in Figure  
 1040 3.18. The regions around  $\phi = 0^\circ$  and low  $\theta$ , and around  $\phi = \pm 180^\circ$  and high  $\theta$  are well covered by  
 1041 CLAS12. Following this observation, it is decided to study the FB asymmetry at  $\phi_0 = 0^\circ$ .

1042 A prediction for the asymmetry is shown in Figure 3.19a, where the  $t$ -dependence of the asymmetry  
 1043 is plotted for various values of  $\theta_0$ . These results were cross checked independently by M.Vanderhaeghen  
 1044 also using the VGG code. Figure 3.19b shows the results of this independent analysis. Both results  
 1045 are pointing toward an asymmetry with a positive value. Note that these results were obtained after  
 1046 correcting a missing minus sign in Equation (17) of [2]. The consequence of adding this minus sign is  
 1047 to flip the sign of the asymmetry.

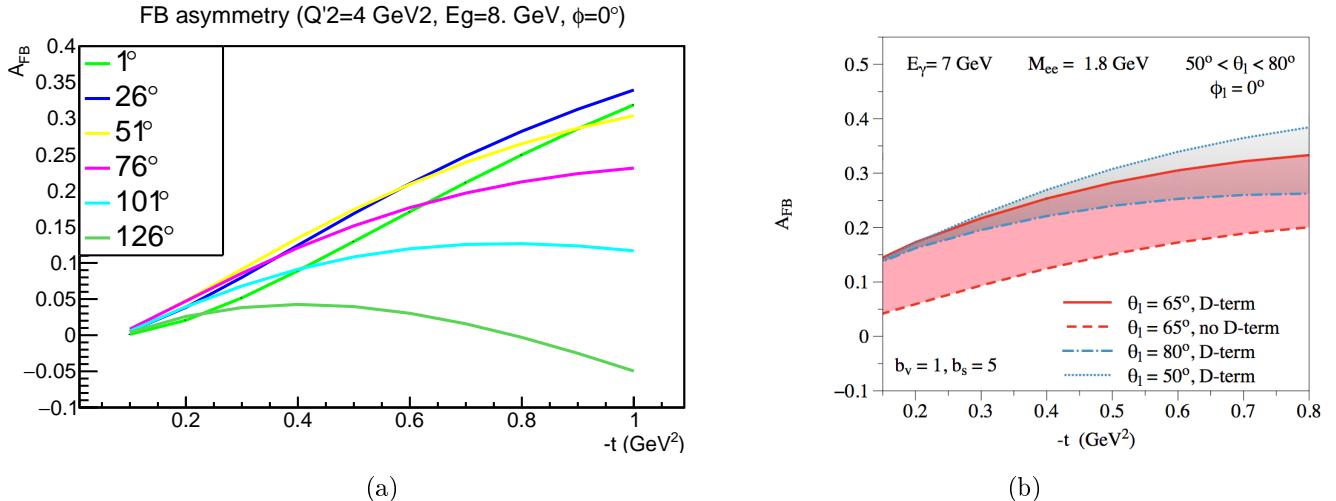


Figure 3.19: (a) FB asymmetry as a function of  $-t$  at fixed  $\phi_0 = 0^\circ$  for different values of  $\theta_0$ . (b) Asymmetry as a function of  $-t$  at fixed  $\phi_0 = 0^\circ$  for different values of  $\theta_0$  and different models: the impact of the D-term on the asymmetry is shown. Figure courtesy of M. Vanderhaeghen.

1048 Later predictions realized by Paweł Sznajder [15], using the *PARTONS* software [16], are shown in  
 1049 Figure 3.20. These predictions also point to a positive asymmetry, independently of the model used  
 1050 for the CFFs as well as high order corrections.

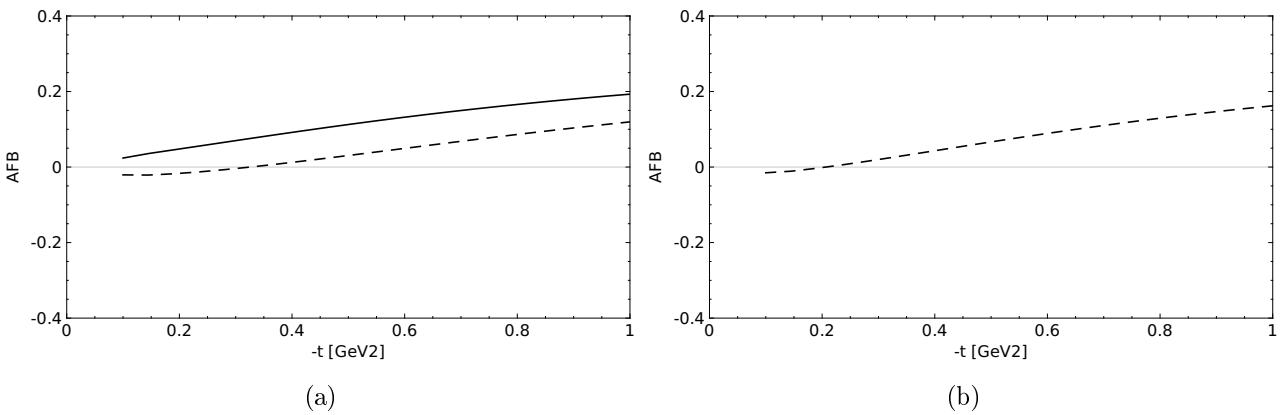


Figure 3.20: FB asymmetry as a function of  $-t$  at fixed  $\phi_0 = 0^\circ$ ,  $\theta_0 = 60^\circ$ ,  $E_\gamma = 7$  GeV and  $Q'^2 = 1.8$  GeV $^2$  (a) using the GK model at LO (dashed) and NLO (plain), (b) using the VGG model at LO. Figures courtesy of Paweł Sznajder.

1051 Furthermore, the analytical formulae for TCS derived in [1] were used to cross-check the asymmetry  
 1052 in VGG. This was done after making sure the CFF conventions are consistent. The results obtained  
 1053 for the analytical formulae using VGG CFFs are shown in Figure 3.21. The sign of the asymmetry  
 1054 obtained is consistent with the *PARTONS* results and with the corrected VGG asymmetries.

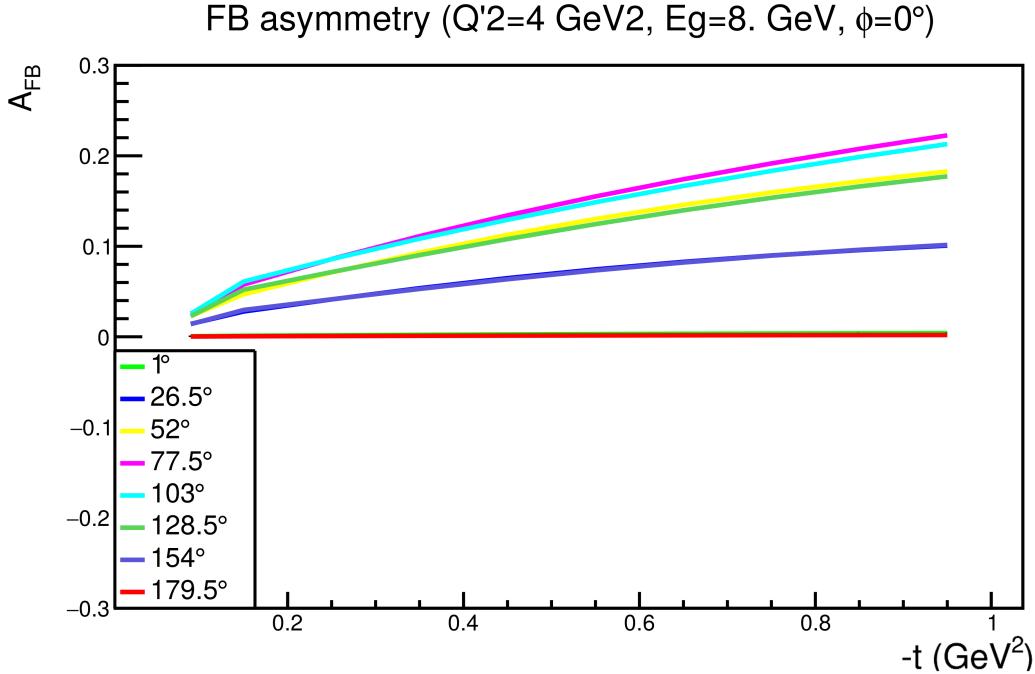


Figure 3.21: FB asymmetry as a function of  $-t$  at fixed  $\phi_0 = 0^\circ$  for different values of  $\theta_0$  using formulas derived in [1].

1055     A final consistency check was performed by calculating the asymmetry from the TCS only and BH  
1056     only cross sections. The absence of asymmetry in both cases is well verified.

### 1057     3.7.2 TCS $A_{FB}$ kinematic dependencies

1058     A main feature of the FB asymmetry is that it can be measured over a small portion of the TCS  
1059     angular phase space. This prevents any large detector-induced false asymmetry caused by holes in  
1060     the acceptance, which instead occurs in the case of the  $R$  ratio (see Section 3.6). On the contrary,  
1061     the statistics dramatically falls if the angular phase space of the measurement is too narrow. The  
1062     main goal of this analysis is to determine the angular range that can be used for the measurement in  
1063     CLAS12, in order to maximize statistics. One has to make sure that the FB asymmetry conserves its  
1064     sign over the angular integration domain to maximize the size of the measured asymmetry. Therefore,  
1065     the angular dependence of the FB asymmetry is studied. From the explicit expression of the  $A_{FB}$   
1066     derived following [1]:

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

1067     one can see that the  $\phi_0$  dependence is driven by the  $\cos(\phi_0)$  factor. The  $\phi_0$  behavior is shown in Figure  
1068     3.22 for fixed  $\theta = 80^\circ$ . The asymmetry changes sign at  $\phi_0 \approx \pm 90^\circ$ . This behavior is reproduced by  
1069     both the VGG model and the analytical model. The differences are attributed to the terms that are  
1070     neglected in the analytical model (see [1] for full details). Following this investigation, one can see that  
1071     the FB asymmetry in the region around  $\phi_0 = 0^\circ$  is maximal and that the sign is constant in its vicinity.  
1072     In Figure 3.19a, showing the asymmetry as a function of  $-t$  at  $\phi_0 = 0^\circ$  and for different values of  $\theta_0$ ,  
1073     the asymmetry has constant sign over the  $\theta_0$  range accessible in the CLAS12 acceptance (from  $\sim 40^\circ$   
1074     to  $\sim 100^\circ$ ).

1075     Finally, the asymmetry was studied as a function of the incoming photon energy  $E_\gamma$  and the  
1076     square of the invariant mass of the lepton pair  $Q'^2$ , aiming at increasing the statistics available for  
1077     the measurement. Figure 3.23 shows the  $Q'^2$  and  $E_\gamma$  dependencies of the FB asymmetry at  $\phi_0 = 0^\circ$   
1078     and for various values of  $\theta_0$ . These plots indicate that the sign of the asymmetry is constant over the

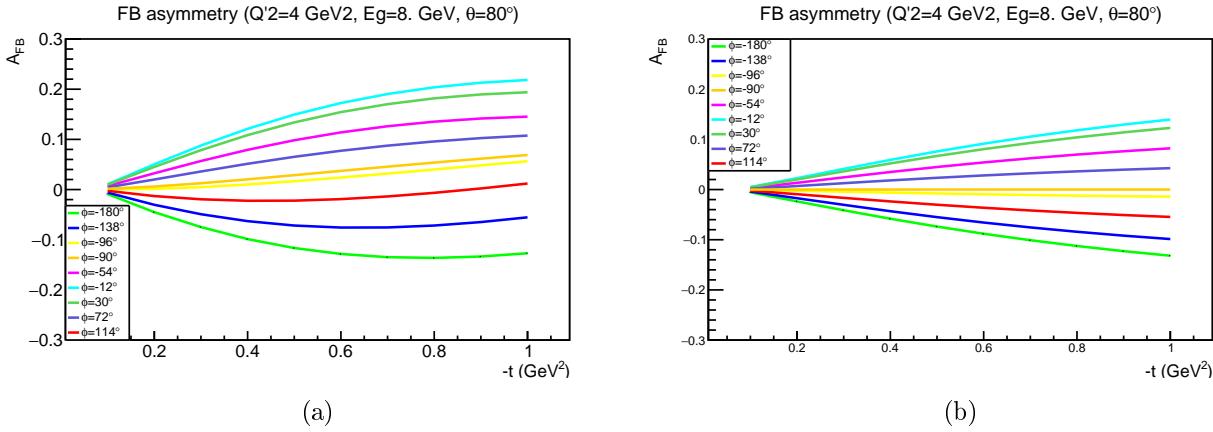


Figure 3.22: FB asymmetry as a function of  $-t$  at fixed  $\theta_0 = 80^\circ$  for different values of  $\phi_0$  computed (a) with the VGG code, (b) using the formulas derived in [1].

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.

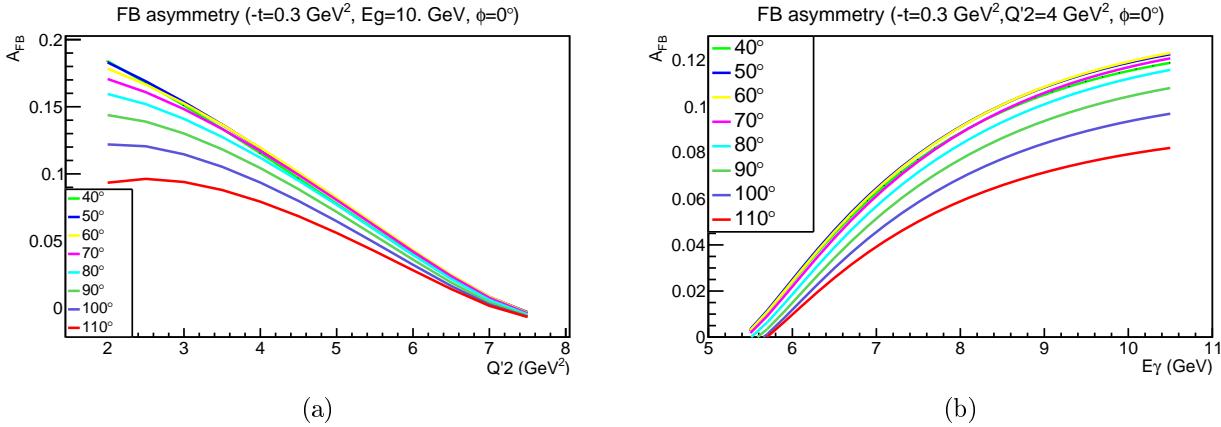


Figure 3.23: FB asymmetry calculated using the VGG code (a) as a function of  $Q'^2$  and (b) as a function of  $E_\gamma$  at fixed  $\phi_0 = 0^\circ$  and for different values of  $\theta_0$ .

### 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 3.24. 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.

#### $b_{sea}$ -dependence

The dependence of the asymmetry to the skewness parameter  $b_{sea}$  is also explored. Indeed in [18] it is suggested that DVCS data, sensitive to the imaginary part of the  $\mathcal{H}$ , are better described with skewness

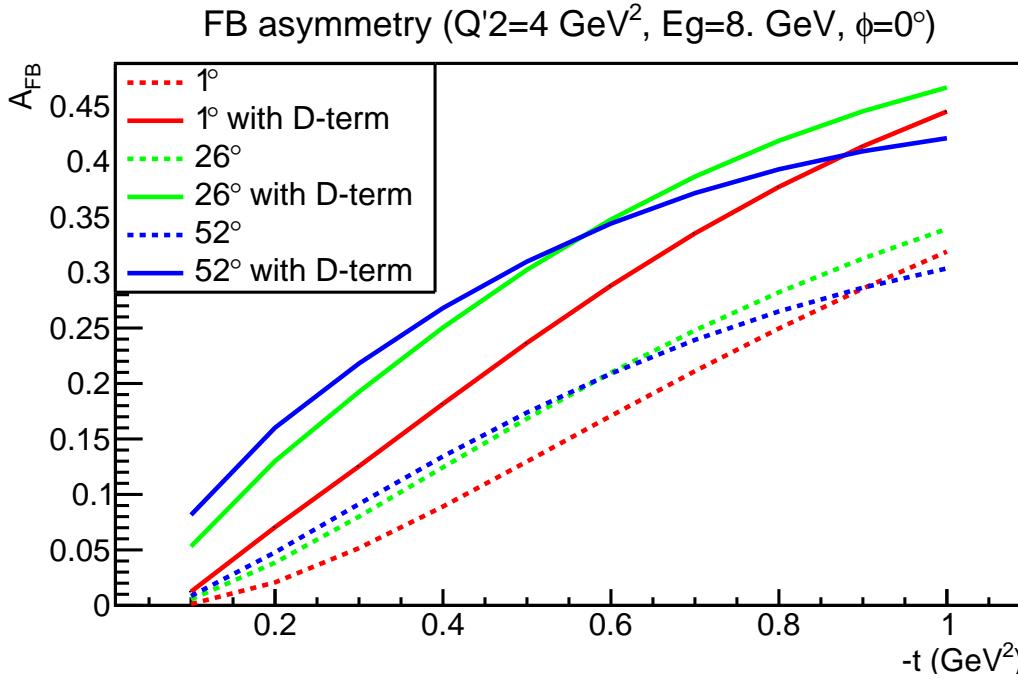


Figure 3.24: FB asymmetry as a function of  $-t$  at  $\phi_0 = 0^\circ$  for different values of  $\theta$ ; and with (solid lines) or without (dashed lines) the D-term contributions to the GPD H.

parameters  $b_{val} = 1$  and  $b_{sea} = 5$ . Following this observation, the value of the sea parameter is varied in the asymmetry calculation, from 1 (strong skewness dependence) to 8 (low skewness dependence), to verify if the FB asymmetry can help in the determination of this parameter. The results are shown in Figure 3.25, where the asymmetry dependencies on  $-t$  and  $Q'^2$  are plotted for various values of the sea parameter. No large effect is seen when varying the parameter  $b_{sea}$ .

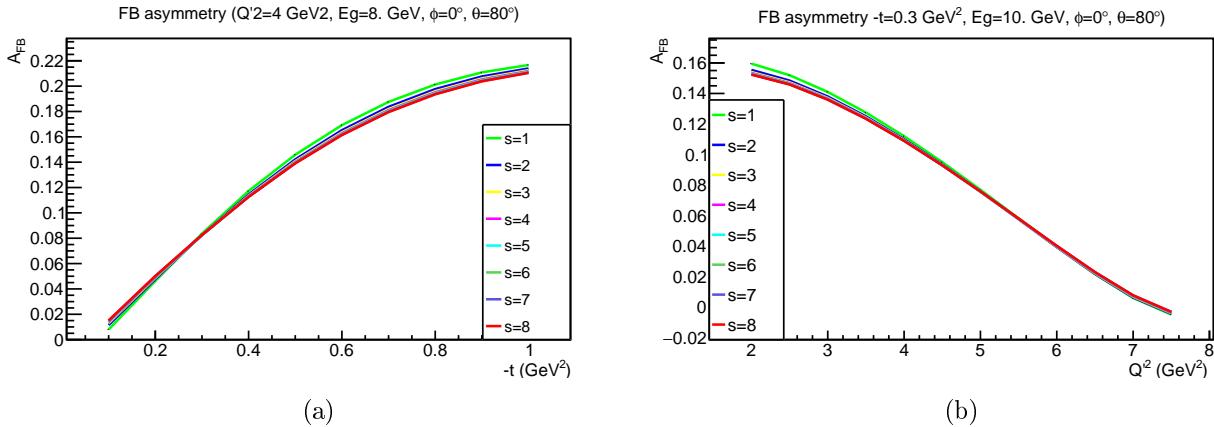


Figure 3.25: (a)  $-t$  dependence of  $A_{FB}$  for different values of the  $b_{sea}$  parameter. (b)  $Q'^2$  dependence of  $A_{FB}$  for different values of the  $b_{sea}$  parameter.

## 3.8 Experimental Forward-Backward asymmetry

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

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

- 1105 •  $50^\circ < \theta_0 < 80^\circ$ .

1106 The corresponding backward bin limits are:

- 1107 •  $140^\circ < 180^\circ + \phi_0 < 220^\circ$
- 1108 •  $100^\circ < 180^\circ - \theta_0 < 130^\circ$ .

1109 **Bin volume correction**

1110 Figure 3.26 shows the acceptance of CLAS12 in the  $\theta$  -  $\phi$  plane for a given  $-t$ ,  $E_\gamma$  and  $Q'^2$  bin. The  
 1111 limits of the forward and backward bins are shown in green and red, respectively. Although the limits  
 1112 of the angular bins have been chosen to be covered by the CLAS12 acceptance, some  $\{E_\gamma, -t, Q'^2\}$   
 1113 acceptance bins do not fully cover the forward or backward angular bins. The difference in coverage  
 1114 between the forward and the backward direction can be the source of fake asymmetries. To correct  
 1115 for this, a bin volume correction is applied during the calculation of the FB asymmetry. For each  
 1116 acceptance bin  $\{E_\gamma, -t, Q'^2\}$ , the fraction of the angular bins covered by the acceptance,  $CorrVol_{F/B}$ ,  
 1117 is calculated. Each event detected in the forward (resp. backward) bin is then assigned a correction  
 1118 weight equal to the inverse of the fraction of the volume covered by the acceptance in this bin. This  
 1119 correction assumes that the cross section of the TCS reaction is relatively constant within the volume  
 1120 of the forward (resp. backward) bin and that it can be estimated only by measuring it in the volume  
 1121 covered by the acceptance of CLAS12. This hypothesis is verified by extracting the FB asymmetry with  
 1122 BH-weighted simulation events and the difference between the expected value for BH (null asymmetry)  
 1123 and the extracted value is assigned as a systematic uncertainty (see Section 3.12). Two sets of volume  
 correction coefficients are obtained, one for the forward and one for the backward angular bins.

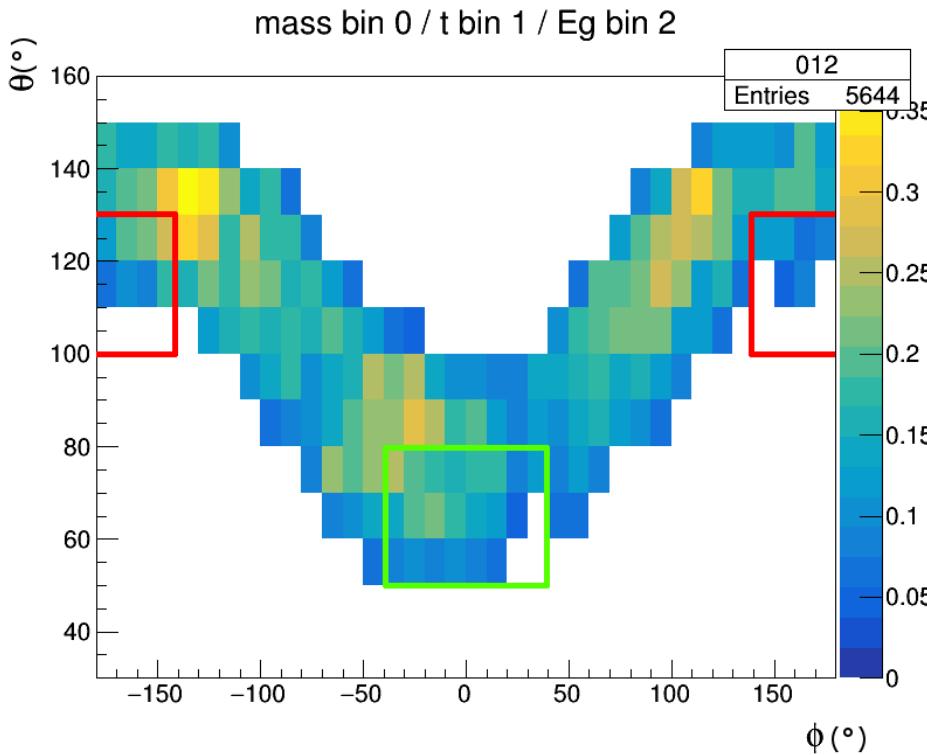


Figure 3.26: Volume correction for the  $A_{FB}$  calculation. The forward bin is represented by the green square. The red square delimits the backward bin. One can see that CLAS12 does not fully cover the angular bins for some  $\{E_\gamma, -t, Q'^2\}$  bins.

1124

1125 For a given bin, the value of the FB asymmetry is calculated as the ratio:

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

1126 where  $N_{F/B}$  is the number of events in the forward (resp. backward) angular bin, corrected by the  
 1127 acceptance and the bin volume, as:

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

1128 The reported statistical error bars are calculated by propagating the weighted error on  $N_{F/B}$ :

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

### 1129 3.9 Experimental beam-spin asymmetry

1130 As mentioned previously, this analysis is done on quasi-real photoproduction events, where a real  
 1131 photon is radiated by the initial electron beam. In this configuration, the circular polarization of the  
 1132 photon can be inferred from the initial longitudinal polarization of the electron beam. An electron  
 1133 polarized in the direction (resp. opposite) of the beam emits a right-(resp. left-) handed circularly  
 1134 polarized photon, with a polarization transfer  $Pol_{transf}$ . fully calculable analytically in QED (see  
 1135 [19, 20] for details and in Section 3.10 for the detailed formulas used to calculate the polarization  
 1136 transfer).

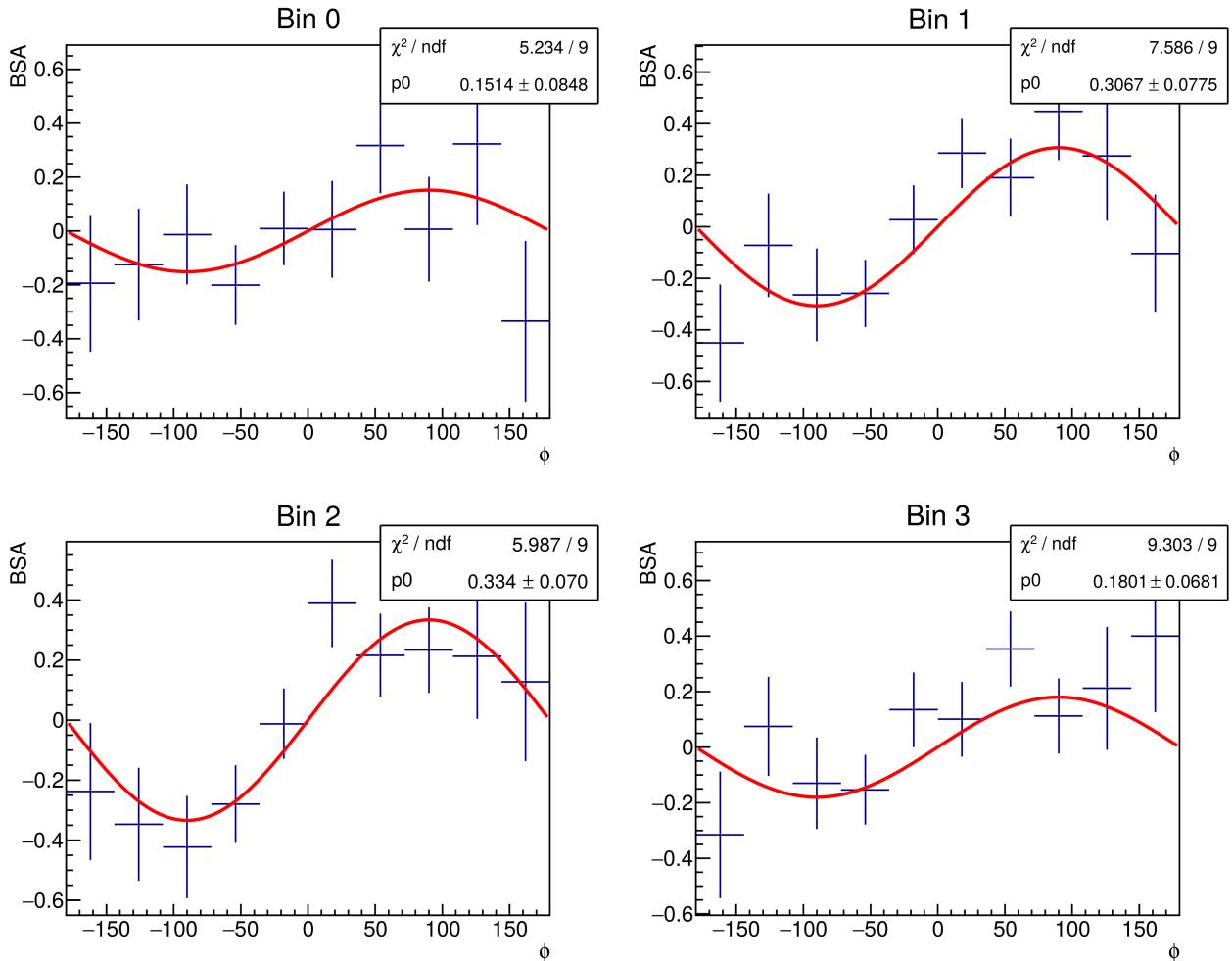


Figure 3.27: Distributions of the photon helicity asymmetry as a function of  $\phi$  for four bins in  $-t$  (as described in Section 3.11). The fit used to extract the asymmetry is overlaid. The fit amplitudes are displayed in Figure 4.11.

1137 The photon polarization asymmetry is calculated as a function of the azimuthal angle  $\phi$  as:

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

1138 where the number of events with reported positive  $N^+$  (resp. negative  $N^-$ ) electron helicity in each  
1139 bin is corrected for the acceptance and the polarization transfer as:

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

1140  $Pol_{eff}$  is the effective polarization of the CEBAF electron beam, which is estimated to be 85% on  
1141 average. The  $\phi$ -distribution is then fitted with a sinus function:

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

1142 as shown in Figure 3.27. Following the definition of [2], the amplitude of the sinus function given by  
1143 the fit (BSA at  $\phi = \pi/2$ ) is extracted. The reported statistical error bars are given by the fit error on  
1144 the amplitude of the function.

## 1145 3.10 Photon polarization transfer

1146 The circular polarization  $P$  of the incoming quasi-real photon in the  $ep \rightarrow (e')\gamma p \rightarrow (e')e^-e^+p'$  reaction  
1147 can be fully calculated in QED. It is given by the helicity  $S$  of the electron from the beam that emitted  
1148 the photon multiplied by a polarization transfer coefficient  $L$ . If the electron has positive helicity  
1149 the photon has right-handed polarization, while a negative helicity electron will emit a left-handed  
1150 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:

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

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

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

1151 where  $L$  is the polarization transfer factor,  $E_1$  and  $E_2$  are respectively the energy of the incident and  
1152 the scattered electrons and  $k = (E_1 - E_2)$  is the energy of the photon. The Coulomb screening factor  
1153  $\Gamma$  is given by:

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

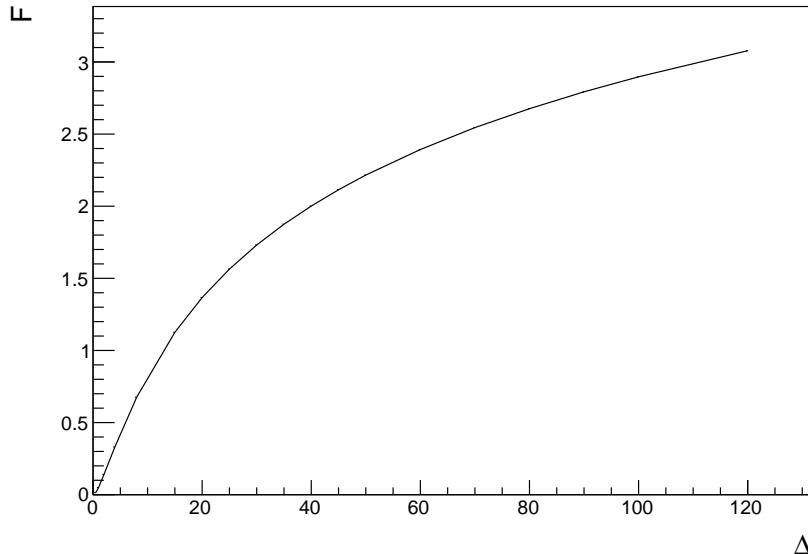
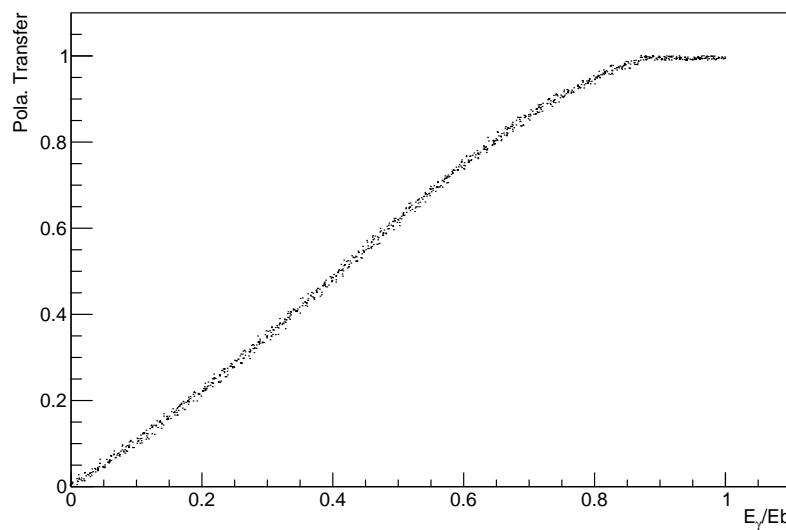
1154 where  $\delta = k/2E_1E_2$ . The factor  $\xi$  is calculated as  $\xi = 1/(1 + u^2)$ , with  $u = E_1 \sin \theta_\gamma$ , where  $\theta_\gamma$  is  
1155 the angle between the incoming electron and the radiated photon. The function  $\mathcal{F}$  accounts for the  
1156 screening effects. Tabulated values for  $\mathcal{F}$  are given in Table 3.2. The screening function is also plotted  
1157 in Figure 3.28. Finally the Coulomb correction function  $f(Z)$  is given by:

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

1158 where  $a = \alpha Z$ ,  $\alpha$  is the electromagnetic coupling constant and  $Z$  is the atomic number of the target  
1159 material where the radiation of the photon occurs.

1160 The polarization transfer function  $L$  used for the TCS analysis is plotted as a function of the ratio  
1161 between the energy of the photon and the energy of the beam in Figure 3.29.

$\Delta$	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
$\Delta$	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  $\Delta = (6Z^{-1/3}/121)(\xi/\delta)$ .Figure 3.28: Screening function  $\mathcal{F}$  as a function of  $\Delta$ .Figure 3.29: 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.

### 3.11 Binning of the data

The phase space used for the analysis is binned to have similar number of events in each bin for the  $t$ ,  $E_\gamma$  and  $\xi$  variables. This aims at achieving similar error bars in each bins. There are four bins in  $-t$ , three in  $E_\gamma$  and  $\xi$ . For the lepton invariant mass, the ranges from 1.5 GeV to 2 GeV and from 2 GeV to 3 GeV are divided in two bins each, in order to study the possible effect of vector-meson contamination in the low-mass bin. Besides, the binning in  $-t$  is different in the two mass ranges. The bin limits are summarized in Table 3.3 and superimposed on the kinematic distributions of the data in Figure 3.30.

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
$E_\gamma$	4.0 - 6.4 - 8.0 - 10.6
$M = \sqrt{Q^2}$	1.5 - 1.7 - 2.0 - 2.5 - 3.0
$\xi$	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.

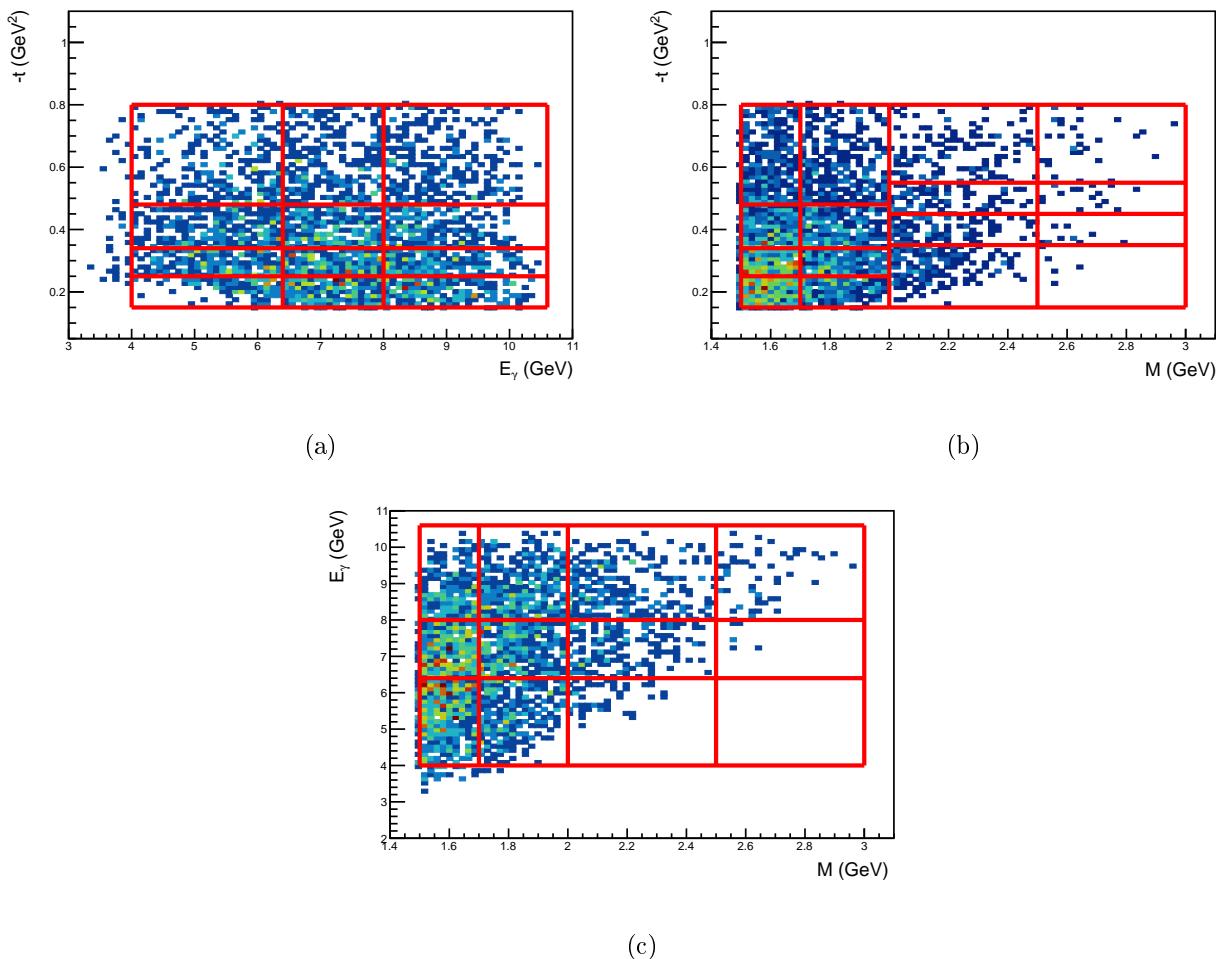


Figure 3.30: Binning grid for the TCS observables in (a) the  $E_\gamma$  -  $(-t)$  plane (the binning in  $-t$  used in this figure is the one used for the [1.5 GeV – 2 GeV] mass range), (b) the  $M$  -  $(-t)$  plane and (c) the  $M$  -  $E_\gamma$  plane.

## 3.12 Systematic errors estimation

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

Source of uncertainty	Standard value	Alternative values
Method	Generated events	Full chain simulations
Positron ID	0.5	0.4/0.6
$\chi^2$ proton	No cuts	3- $\sigma$ 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
Exclu. 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.

### Combining systematic uncertainties

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

### Method

The method used to calculate the observable involves binned acceptance corrections and bin volume corrections for the FB asymmetry. To estimate the impact of these corrections on the extracted values, the method systematic error is computed using simulations. First the observables are calculated using a sample of generated BH-weighted events, without passing them through GEMC and with no corrections applied. This corresponds to the green point in Figure 3.32a. In the case of the  $R'$  ratio, only events within a non-zero acceptance bin are kept. In the case of the FB asymmetry, the observable is calculated within the full angular bin defined in Section 3.8. Finally, in the case of the BSA, a random beam helicity is associated to each event. Then the observables are calculated a second time with the full-chain simulation events, applying all the corrections, except that the acceptance is not corrected for efficiency and data-driven momentum corrections are not applied. Two samples of simulation events are used, one using weighted events from *TCSGen* and one generated with *GRAPE* (blue points and red points on Figure 3.32a). The systematic uncertainty associated with the extraction method is the difference between the ideal case (no corrections) and the value extracted after the full analysis procedure. The systematic error is set between 0 and the difference between the ideal case and the extracted case for the two simulation samples. This procedure can thus result in an asymmetric error in order to reflect the shift induced by the measurement method. This systematics mostly affects observables binned in  $\xi$  (for example the  $R'$  ratio in Figure 4.2), as the acceptance is not binned in this variable. The maximum error induced on the extracted observables is at most 0.1 for the  $A_{FB}$  binned in  $\xi$  in Figure 4.7.

---

1207 **Efficiency / Background merging**

1208 The CLAS12 background merging procedure, aiming at reproducing the detection efficiencies, is  
 1209 used in this analysis. A second method, a method to estimate the proton efficiency is described in  
 1210 Section 2.8. The systematic error associated with these corrections is estimated by the difference of the  
 1211 values of the observables obtained with both methods,  $\Delta_{Eff}$ . The systematic error bar is then defined  
 1212 as  $\pm\Delta_{Eff}/2$ . Figure 3.32d illustrates the determination of the efficiency systematic error, with blue  
 1213 points produced using efficiency corrected acceptance and red points without the correction. As shown  
 1214 in Section 4.1, the efficiency systematics is most of the time dominated by other sources of systematic  
 1215 errors for most of the observables. The maximum observed shift on the extracted observables is 0.1  
 1216 for the  $A_{FB}$  as a function of  $-t$  (Figure 4.9) and the BSA as a function of  $M$  (Figure 4.10).

1217 **Positron identification**

1218 In this analysis the positron identification algorithm plays an important role. To estimate the  
 1219 impact of the remaining pion contamination, the cut applied on the output of the neural network  
 1220 is varied around the chosen value,  $0.5 \pm 0.1$ . The acceptance is also recalculated accordingly. The  
 1221 difference between the observables extracted with the standard and the shifted cuts is assigned as  
 1222 the positron identification systematics. This systematics can be asymmetric as the variation of the  
 1223 extracted observables with the shifted positron cuts can be different in each case. As illustrated in  
 1224 Figure 3.32c, the variation is small for most of the observables. The induced shift is at most 0.1 for  
 1225 the  $A_{FB}$  as a function of  $-t$  in the high-mass region in Figure 4.9, but usually limited to 0.03 for most  
 1226 of the observables.

1227  **$\chi^2$  proton**

1228 In the analysis, no cut on the  $\chi^2$  of the proton is applied. This quantity provided by the CLAS12  
 1229 event builder is related to the match of the measured time-of-flight of the proton to the expected one  
 1230 calculated from momentum. In order to assess the quality of the proton identification and its impact  
 1231 on the results of the analysis, the latter is done applying a  $3\sigma$  cut on the proton as:

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

1232 where the mean  $\mu$  and width  $\sigma$  are obtained from data and simulation via the fits shown in Figure 3.31.  
 1233 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	$\mu$	$\sigma$
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  $\mu$  and widths  $\sigma$  of the  $\chi^2$  of the proton in the CD and the FD for data and simulation.

1234  
 1235 Figure 3.32b illustrates how this systematics is calculated. This systematic uncertainty can produce  
 1236 a shift up to 0.05, but is most of the time dominated by the other sources of uncertainties.

1237 **Acceptance model**

1238 The dependence of the extracted results on the model used in the acceptance is also studied. The  
 1239 acceptance is calculated with BH-weighted events and events without weights (i.e. phase-space gener-  
 1240 ator). The difference between the two methods  $\Delta_{Acc}$  is defined as the acceptance model systematics,  
 1241 and the associated error bar is set to  $\pm\Delta_{Acc}/2$ . The method is illustrated in Figure 3.32f. This source  
 1242 of systematics becomes large for observables with low statistics, such as the  $A_{FB}$ , for which the induced  
 1243 absolute shift can reach values up to 0.05 (as in Figure 4.4).

**1244 Exclusivity cuts**

1245 Finally, the last source of systematic uncertainty studied is the values chosen for the exclusivity cuts.  
1246 To estimate this uncertainty, the analysis is performed with tighter/wider cuts than those presented  
1247 in Section 3.2:

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

1248 OR

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

1249 The systematic uncertainty is determined as for the positron ID. The maximum deviation above  
1250 and under the standard value is assigned as the uncertainty, thus this systematic can be asymmetric.  
1251 This systematic dominates in bins where the statistics is limited, especially for the  $A_{FB}$  and the  
1252 BSA at high mass in Figure 4.10. This shows that the exclusivity cut error is largely correlated with  
1253 the statistical error in these bins. However, the observed systematic shift always remains within the  
1254 statistical error bars.

**1255 Statistics of the simulation sample**

1256 The impact of the number of events in the simulation sample used to produce the acceptance is  
1257 also verified. The results obtained with a smaller simulation set (16M events, instead of 36M events)  
1258 are compared to the results obtained with the standard procedure. Figure 3.33 shows the difference  
1259 between these results. One can see that the size of the simulation sample has a very small impact on  
1260 the final results. For this reason, we do not include this difference in the systematic uncertainties.

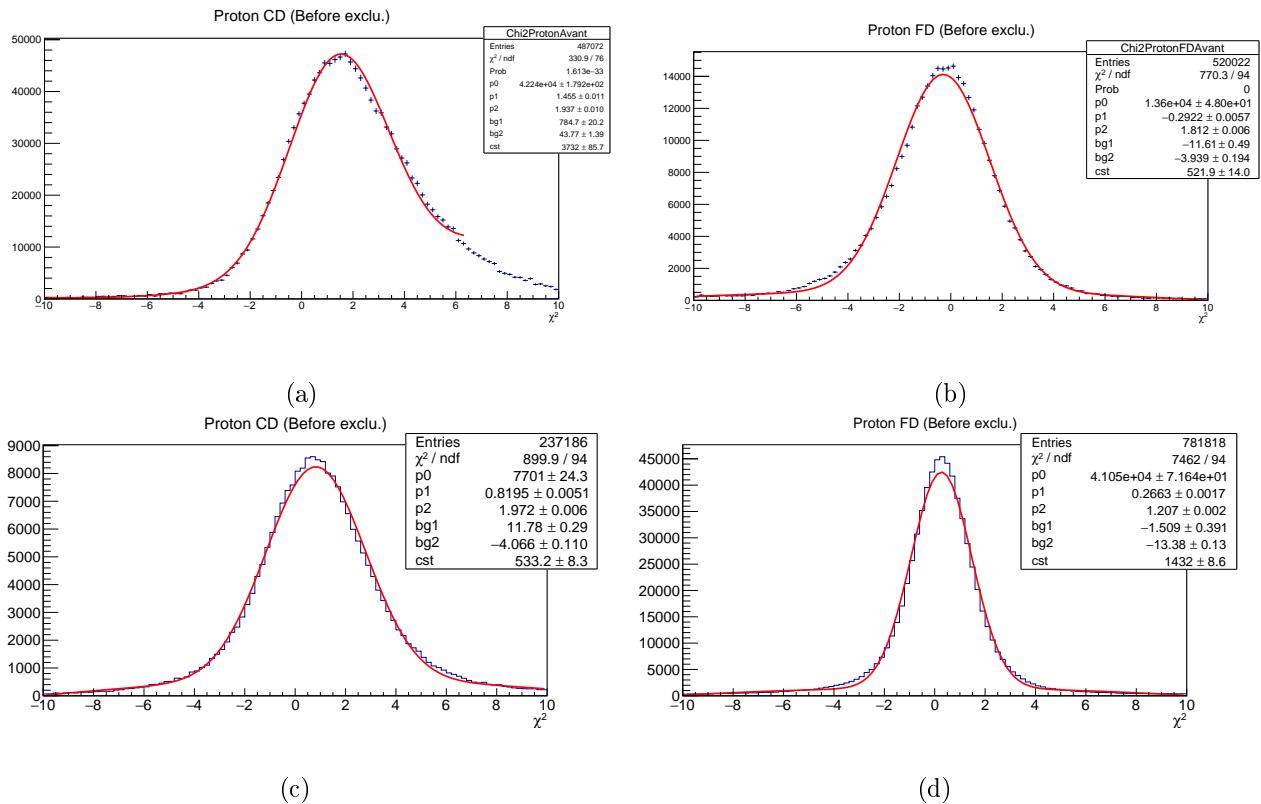


Figure 3.31: Fits used to determine the values shown in Table 3.5. A gaussian plus second order polynomial is used. (a) and (b) show the fits from simulation for the CD and FD respectively. (c) and (d) show the fits from data for the CD and FD respectively.

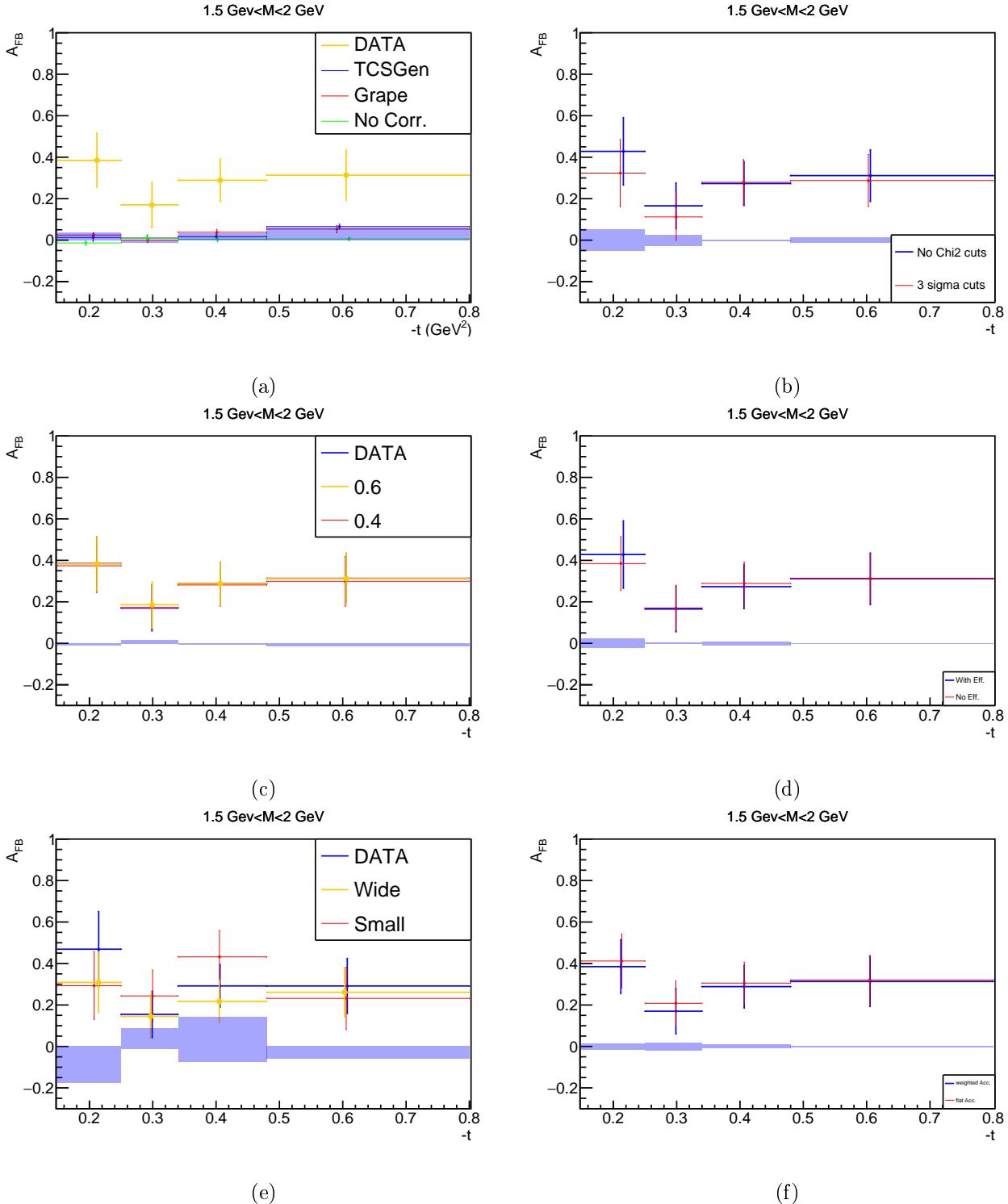


Figure 3.32: Systematic errors for the FB asymmetry in the  $1.5 \text{ GeV} < M < 2 \text{ GeV}$  mass range, as blue bands with respect to a reference value of 0. The vertical error bars correspond to the statistical errors and horizontal error bars expand along the bin size. The same study is done 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  $\chi^2$  cut applied on the proton calculated as the difference of the observable calculated without cut (blue) and with a  $3\sigma$  cut (red). (c) Positron ID systematic error. The cut on the output of the neural network is varied from 0.5 (blue) by  $\pm 0.1$  (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).

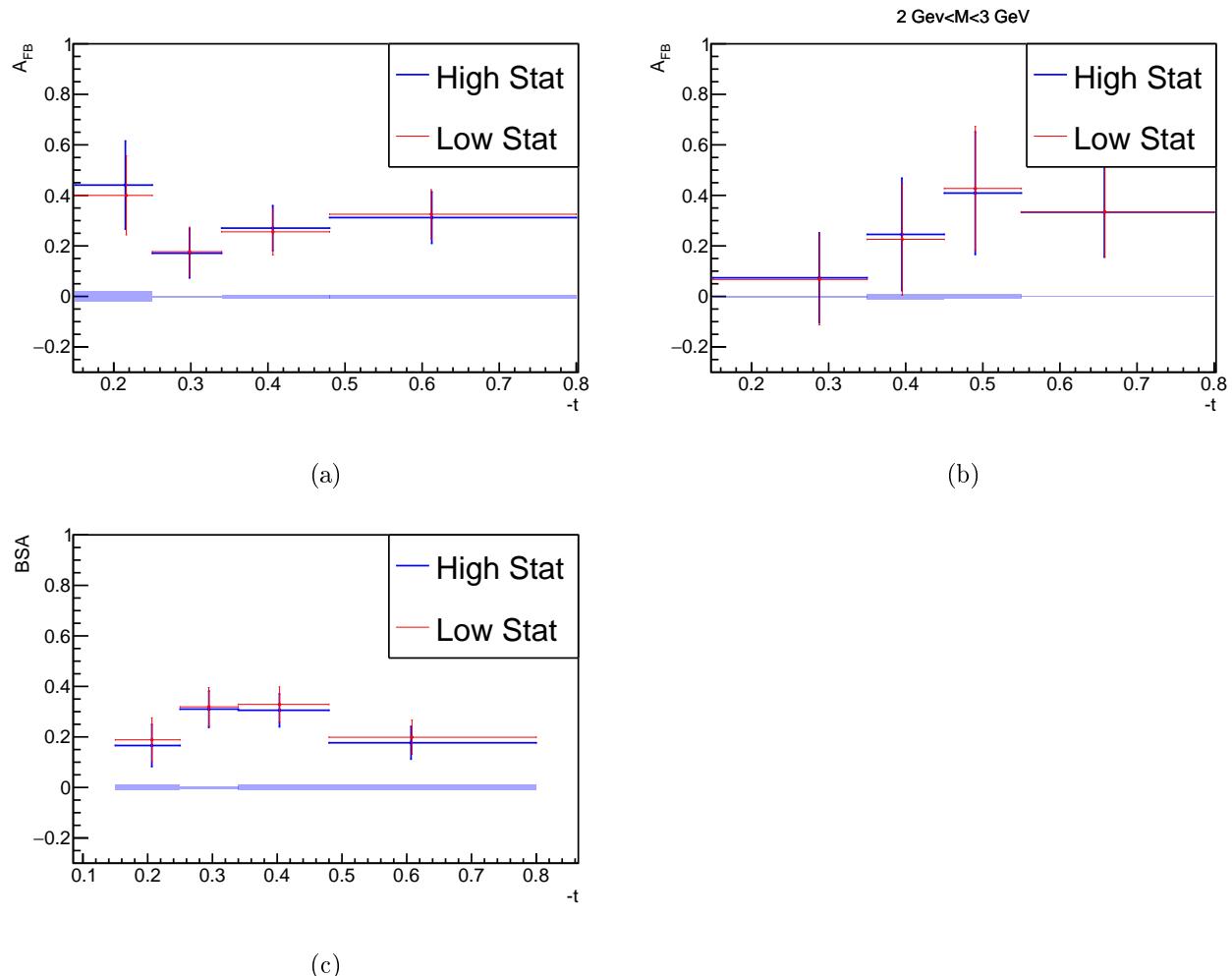


Figure 3.33: Difference of the values of the observables obtained with simulation sample of 16M events (in red) and 36M events (in blue) (a) for the integrated  $A_{FB}$ , (b) for the  $A_{FB}$  at high mass only, (c) for the integrated  $BSA$

# 1261 Chapter 4

## 1262 Results and comparison with model 1263 predictions

1264 This chapter presents the results obtained with the Fall 2018 CLAS12 dataset in the inbending con-  
1265 figuration. The three TCS observables ( $R'$  ratio,  $A_{FB}$  and BSA) are extracted for different kinematic  
1266 bins as a function of different variables. First, all the extracted values of the TCS observables and their  
1267 systematic errors are shown in Section 4.1. A comparison with the CLAS data is shown in Section 4.3.  
1268 Finally some results are compared with model predictions and discussed in Section 4.4.

### 1269 4.1 Complete CLAS12 results for the TCS observables

1270 In this section all the data points obtained in the analysis are displayed. For each observable and for  
1271 each bin, the statistical error bar is shown as a vertical error bar. The horizontal error bar corresponds  
1272 to the size of the bin. The horizontal position of the data points is the average value of the variable in  
1273 the bin, corrected by the acceptance (and in the case of the  $A_{FB}$ , also by the bin volume correction).  
1274 The grey bands show the total systematic uncertainty defined as the quadratic sum of all of the  
1275 systematic contributions described in Section 3.12. The decomposition of the systematic uncertainty  
1276 is shown under each plot. The red points reported on each plots correspond to the expected values  
1277 of the observables in each bin for BH-only events. These points are calculated using BH-weighted  
1278 simulation events and, within the acceptance of CLAS12 for the  $R'$  ratio and the BSA, and within  
1279 the experimental forward and backward bin for the FB asymmetry. Finally, the mean values of the  
1280 integrated kinematic variables, calculated using the same simulations, are given above each plot.

#### 1281 4.1.1 $R'$ ratio

1282 The  $R'$  ratio is extracted as a function of  $-t$  (Figure 4.1) and  $\xi$  (Figure 4.2), with all the other  
1283 variables integrated. The dependence on  $-t$  is further explored by plotting the ratio as a function of  
1284  $-t$  in the invariant mass bin [2 GeV – 3 GeV] (Figure 4.3).

1285 The measured values of the  $R'$  ratio are always bigger than the values expected if only the BH  
1286 process was contributing to the  $\gamma p \rightarrow p' e^+ e^-$  cross section (red points in the following figures). This  
1287 behavior is also seen in the high-mass region [2 GeV – 3 GeV] in Figure 4.3, where no vector-meson  
1288 contamination is expected. This observation validates the fact that the CLAS12 data are sensitive to  
1289 the BH-TCS interference cross section.

#### 1290 4.1.2 $A_{FB}$

1291 The FB asymmetry is extracted as a function of  $E_\gamma$  (Figure 4.4),  $M$  (Figure 4.5),  $-t$  (Figure 4.6) and  
1292  $\xi$  (Figure 4.7), with all the other variables integrated. In order to explore the invariant mass dependence  
1293 further, and especially to investigate the effect of low-mass vector-meson resonances, it is also measured  
1294 as a function of  $-t$  for the two additional mass bins, in the low-mass bin [1.5 GeV – 2 GeV] (Figure  
1295 4.8) and in the high-mass bin [2 GeV – 3 GeV] (Figure 4.9).

1296 As in the case of the  $R'$  ratio discussed above, the extracted values of the  $A_{FB}$  shown in Figures  
1297 4.4, 4.5 and 4.6 are not compatible with zero, confirming that the BH process does not contributes  
1298 alone to the  $\gamma p \rightarrow p' e^+ e^-$  cross section. A none-zero asymmetry is seen in both the low-mass and  
1299 high-mass bins. The signal seen in the high-mass bin, in Figure 4.9, can be attributed to the BH-TCS  
1300 interference cross section, as there is no known vector-meson resonance in this mass range.

1301 **4.1.3 BSA**

1302 The BSA is extracted in CLAS12 acceptance, as a function of  $M$  (Figure 4.10),  $-t$  (Figure 4.11)  
1303 and  $\xi$  (Figure 4.12), with all the other variables integrated.

1304 In each of these figures a clear photon polarization asymmetry is reported. This is a further confirmation  
1305 that we observe the BH-TCS interference in the CLAS12 dataset, as the expected asymmetry  
1306 for the BH process only is zero.

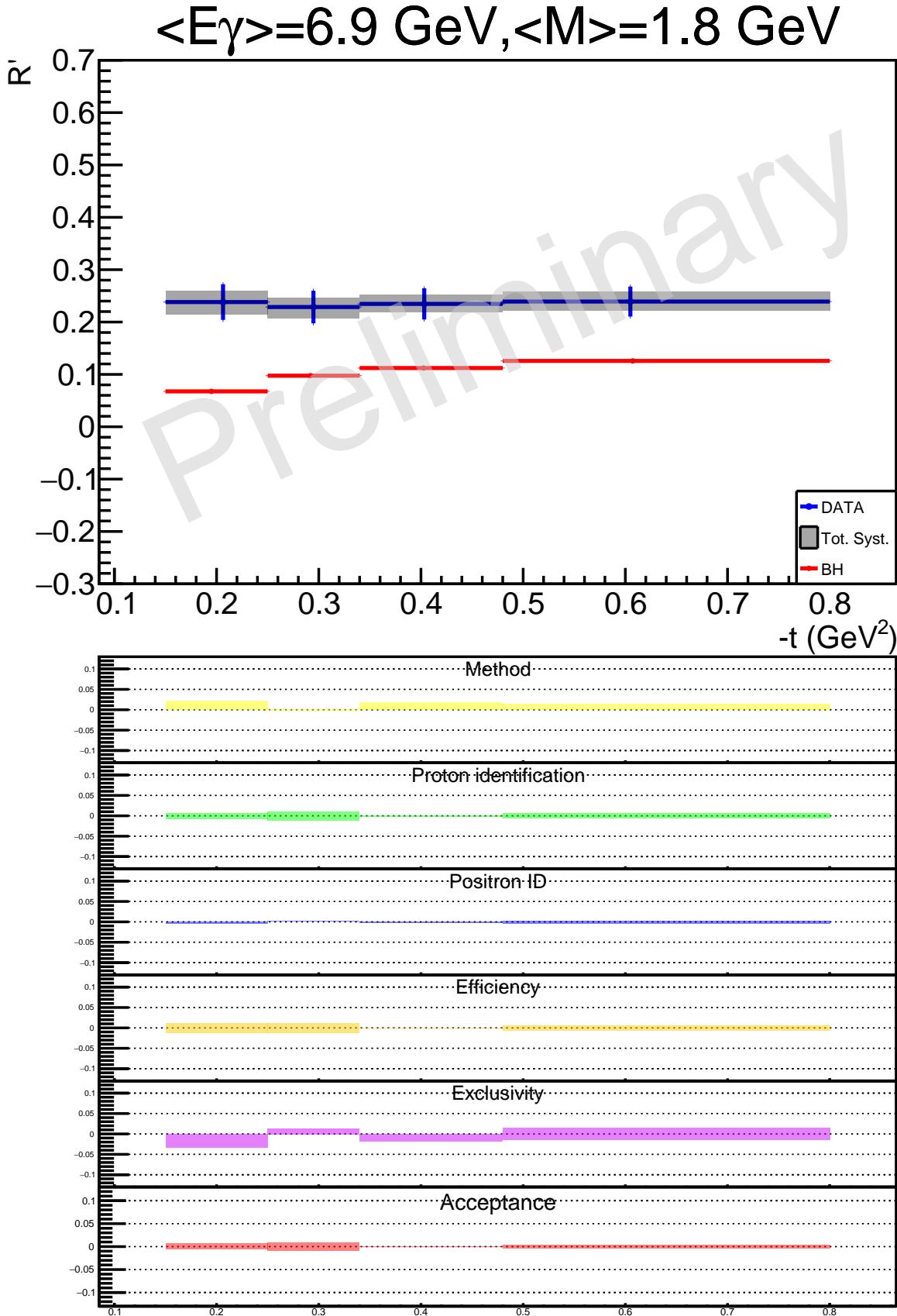


Figure 4.1:  $R'$  ratio as a function of  $-t$ , integrated over all the other variables. The horizontal error bars denote the bin size, the vertical error bars are statistical errors and the grey bands display the total systematic uncertainty. The red points are expected values for BH-only events, obtained with simulations. Tabulated values in Table 4.1 in Appendix 4.2.

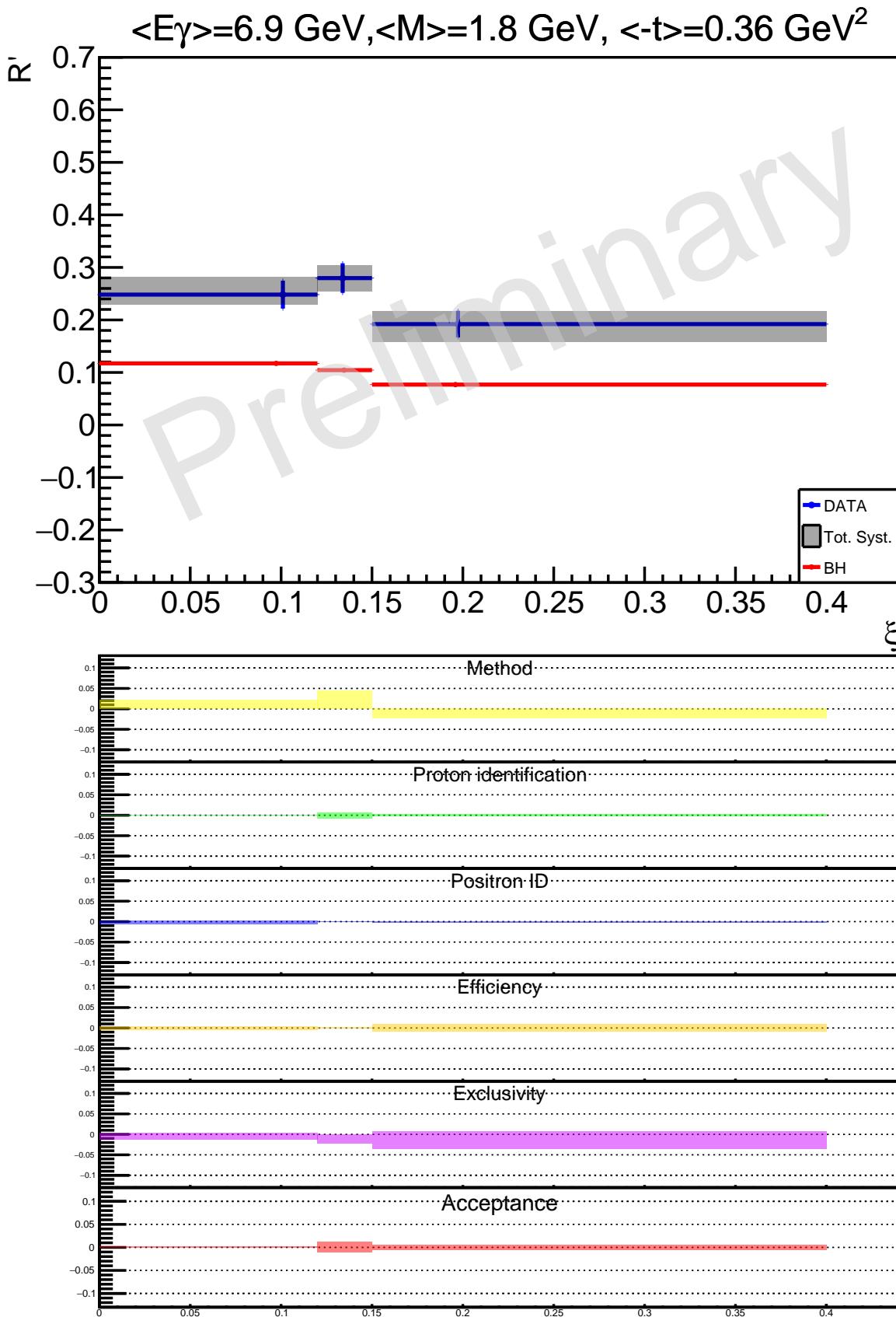


Figure 4.2:  $R'$  ratio as a function of  $\xi$ , integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.2 in Appendix 4.2.

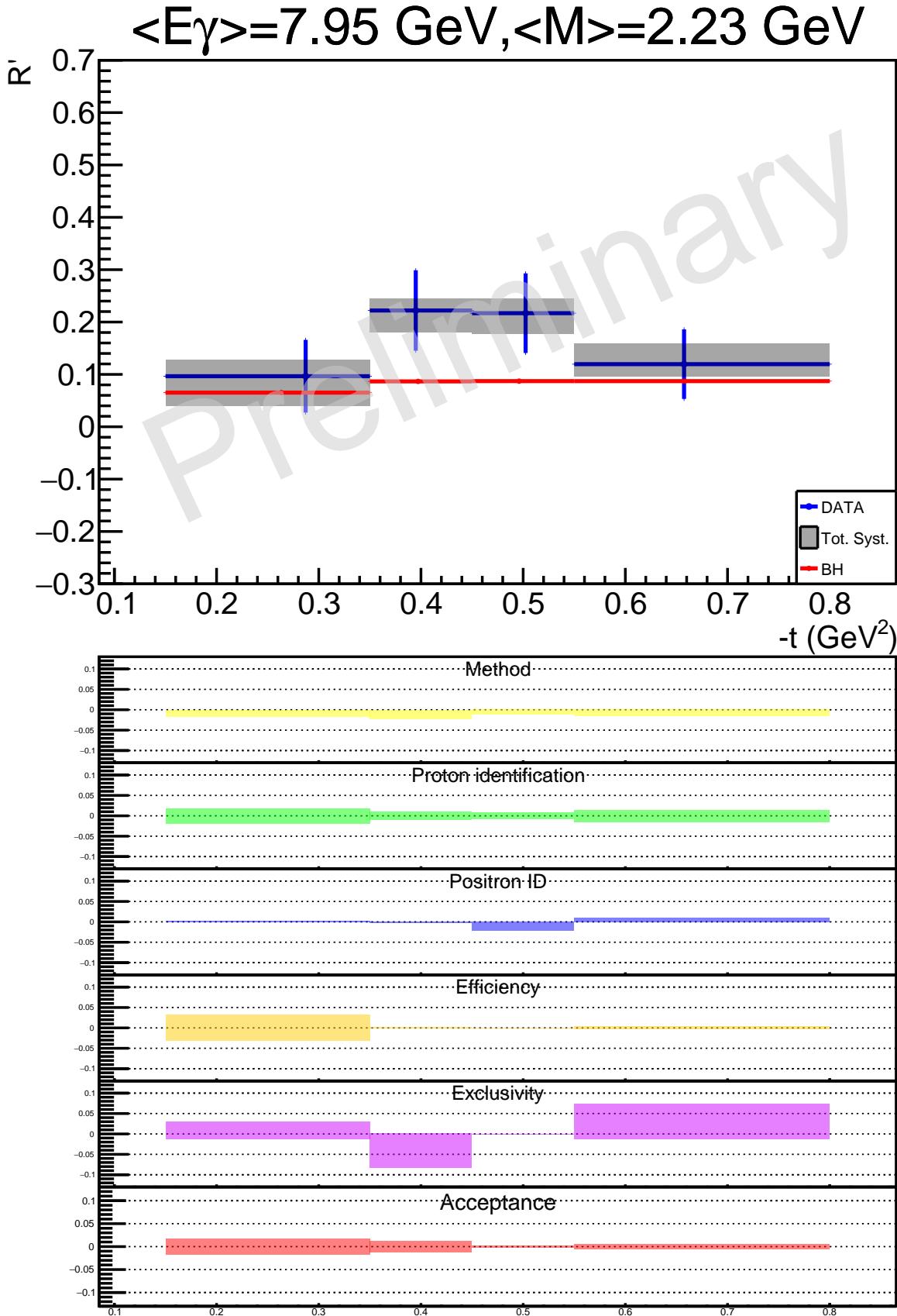


Figure 4.3:  $R'$  ratio as a function of  $-t$ , for the mass range [2 GeV – 3 GeV], integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.3 in Appendix 4.2.

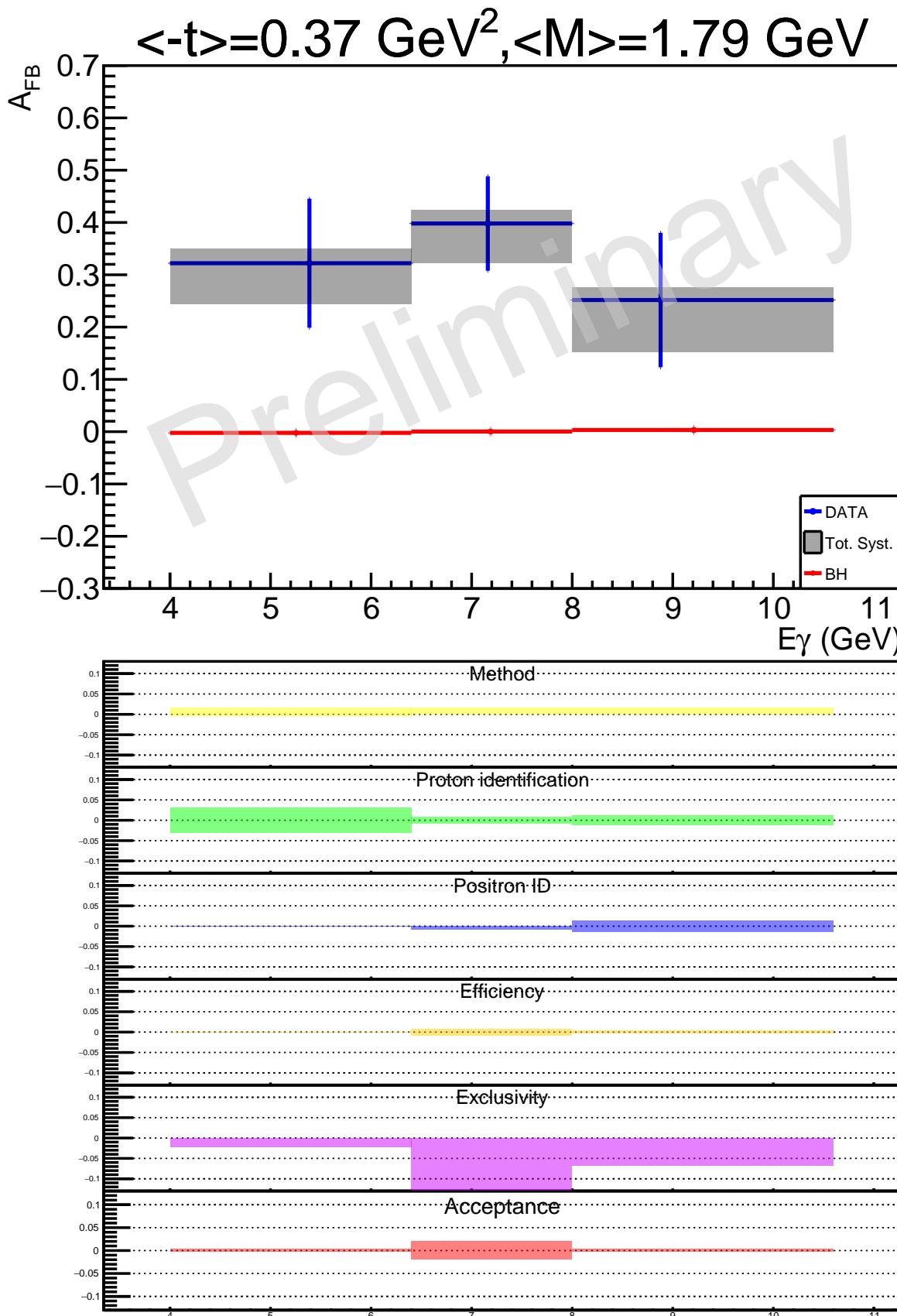


Figure 4.4:  $A_{FB}$  as a function of  $E_\gamma$ , integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.4 in Appendix 4.2.

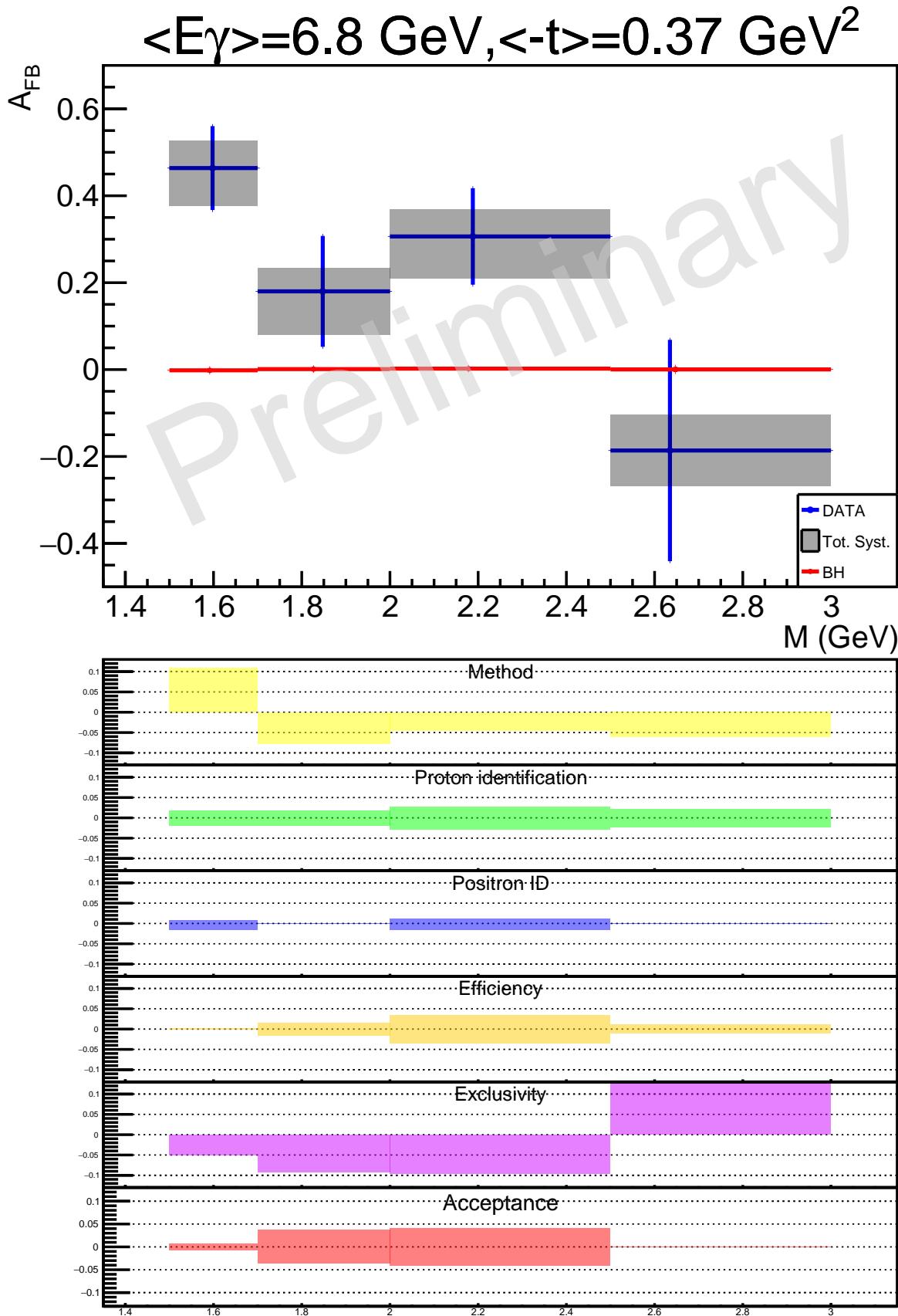


Figure 4.5:  $A_{FB}$  as a function of  $M$ , integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.5 in Appendix 4.2.

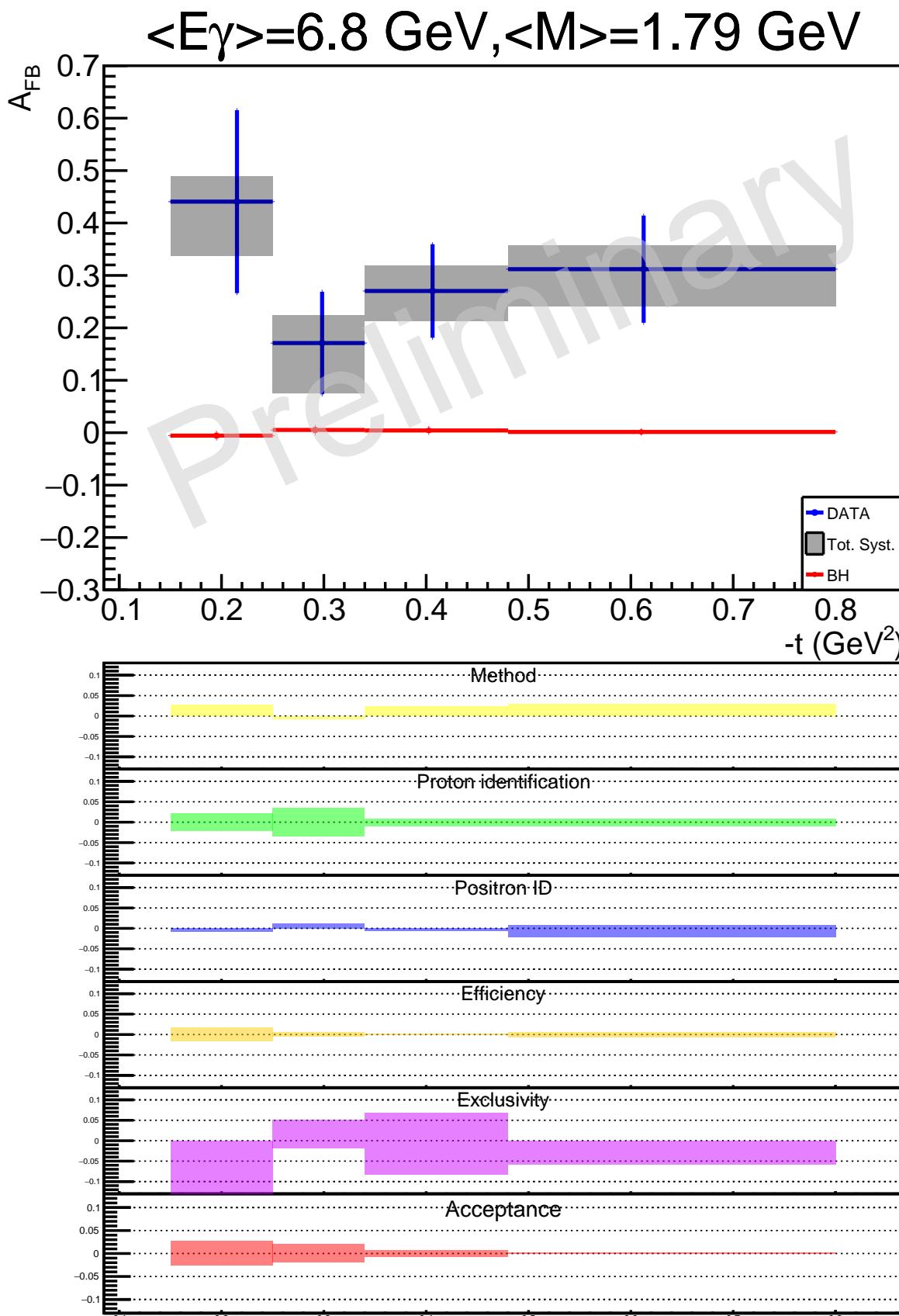


Figure 4.6:  $A_{FB}$  as a function of  $-t$ , integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.6 in Appendix 4.2.

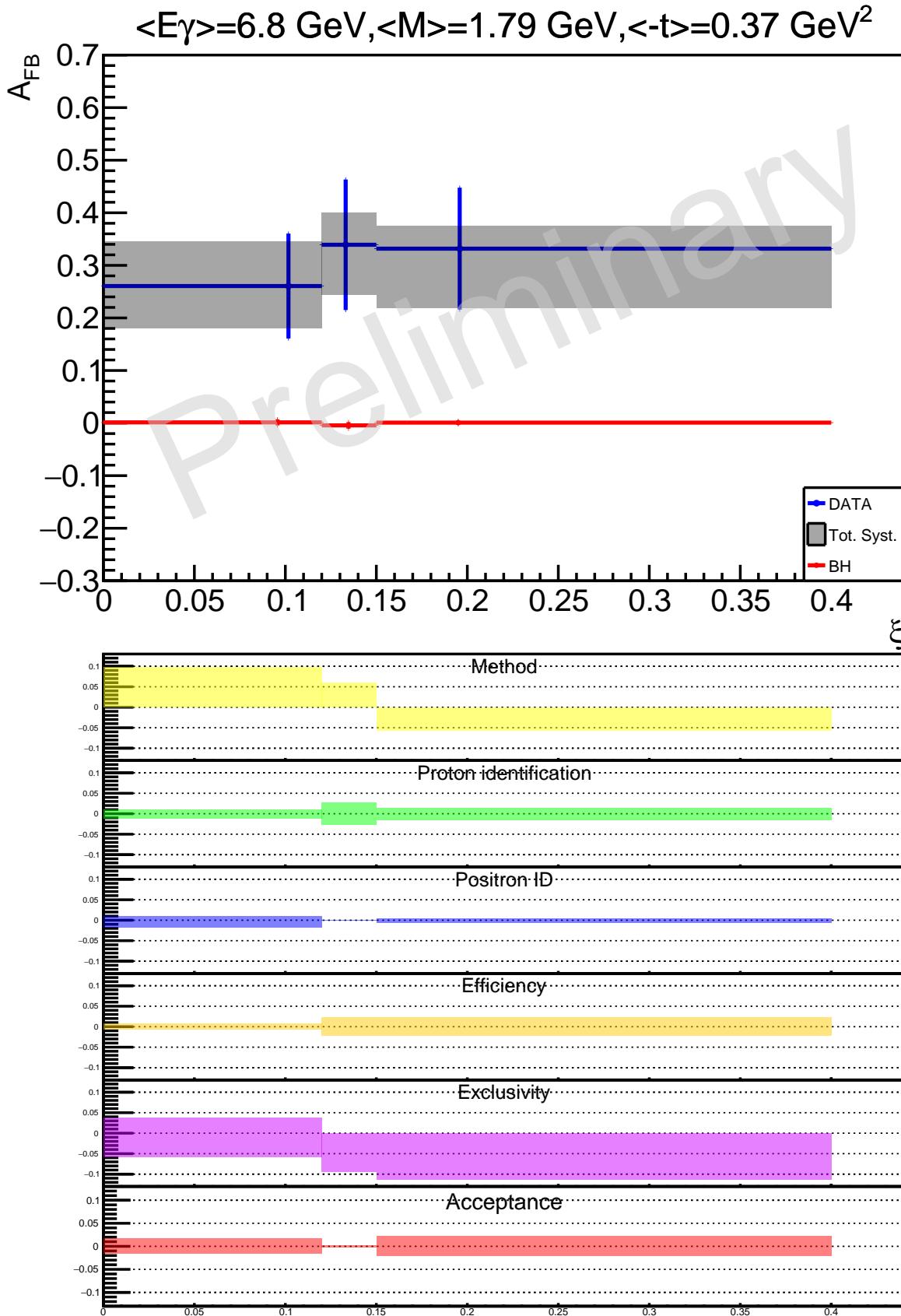


Figure 4.7:  $A_{FB}$  as a function of  $\xi$ , integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.7 in Appendix 4.2.

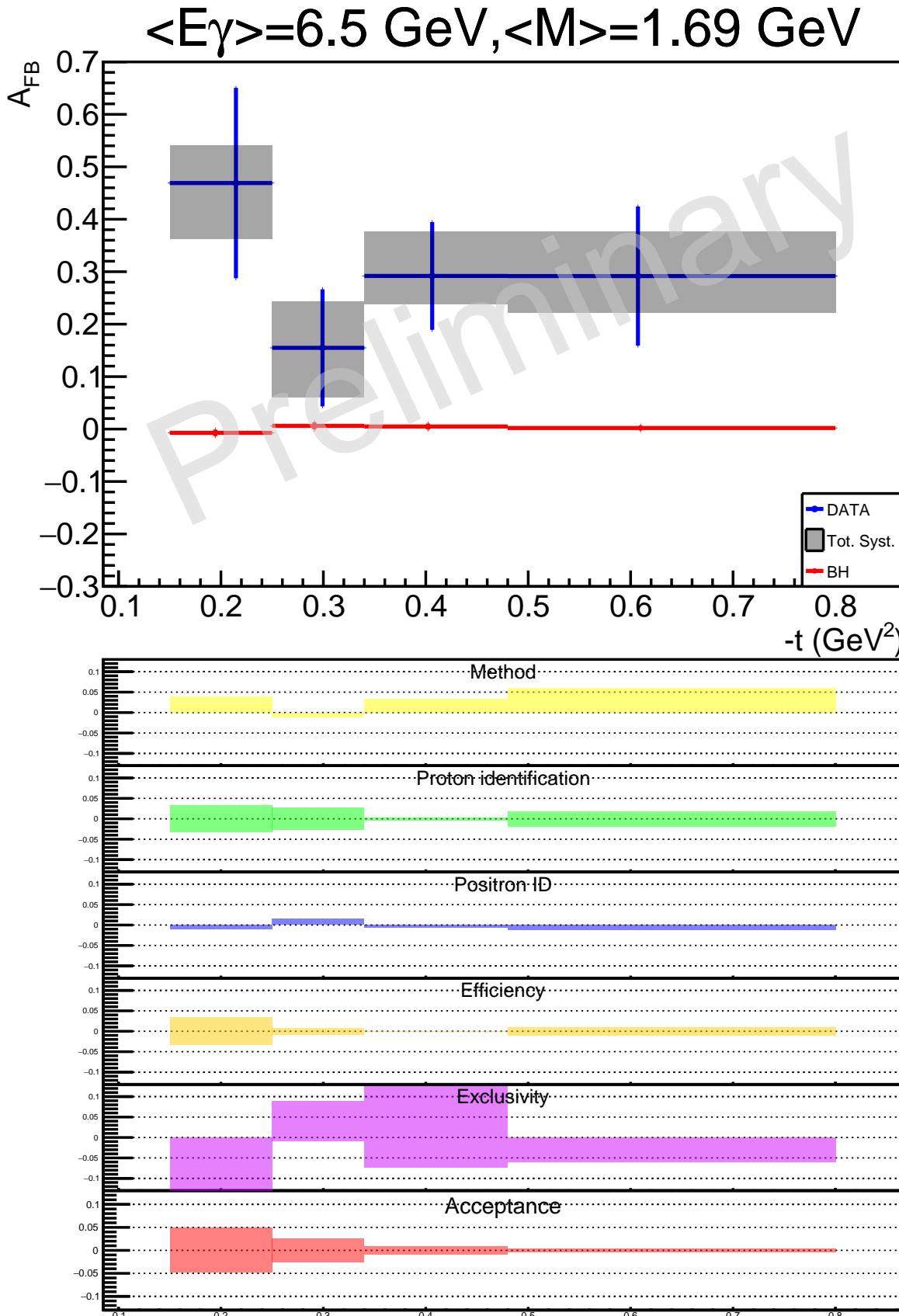


Figure 4.8:  $A_{FB}$  as a function of  $-t$ , in the [1.5 GeV – 2 GeV] mass range, integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.8 in Appendix 4.2.

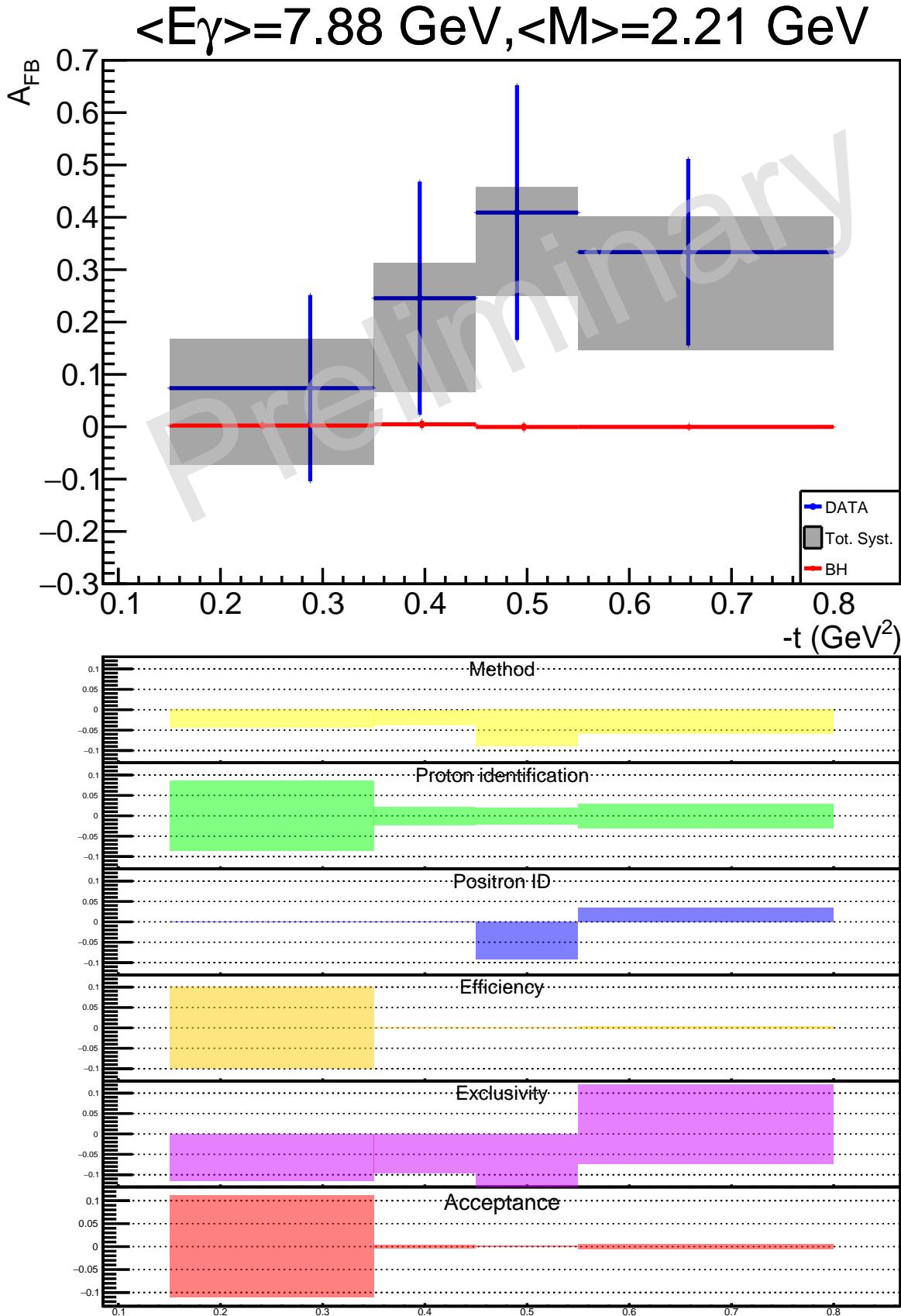


Figure 4.9:  $A_{FB}$  as a function of  $-t$ , in the [2 GeV – 3 GeV] mass range, integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.9 in Appendix 4.2.

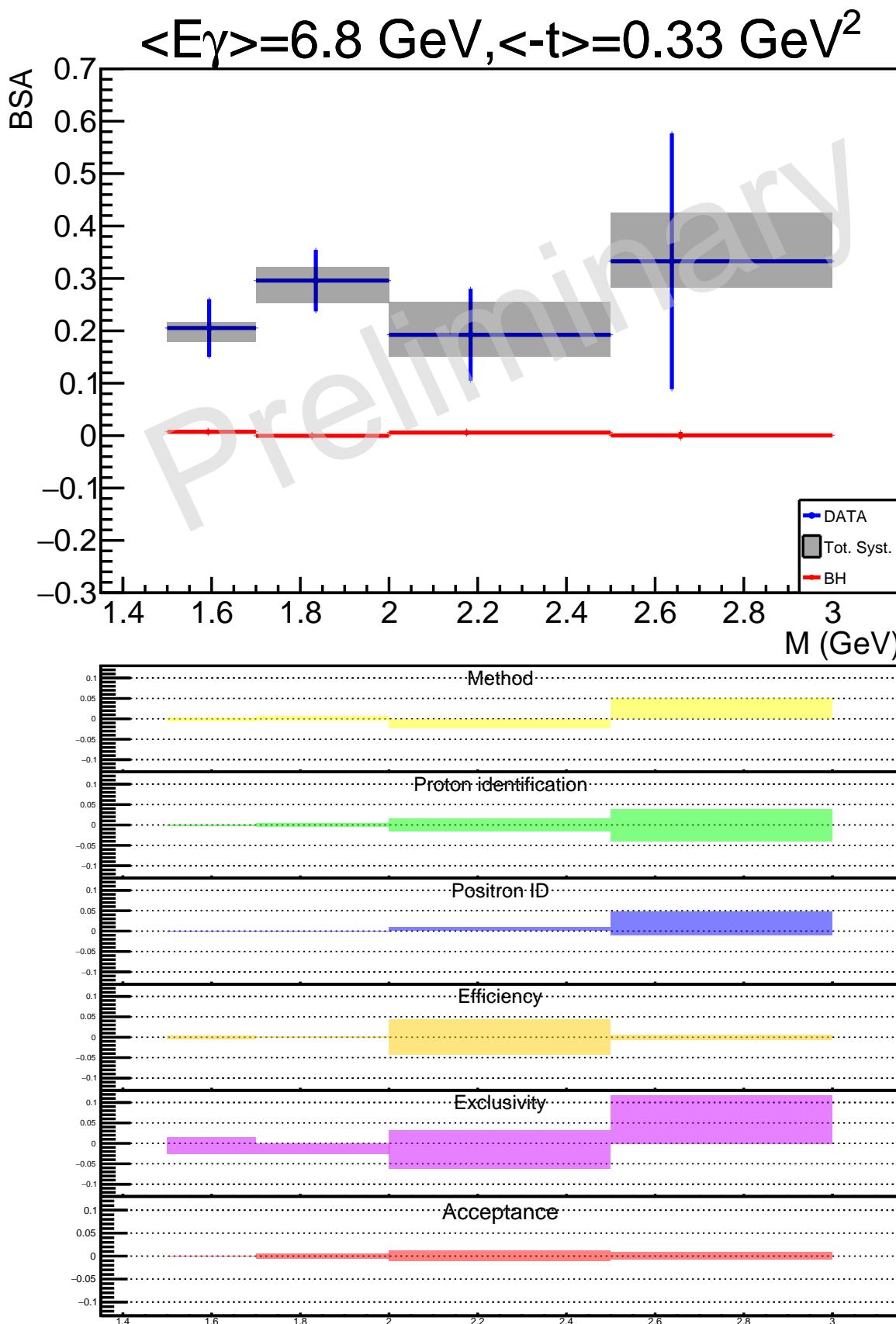


Figure 4.10: BSA as a function of  $M$ , integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.10 in Appendix 4.2.

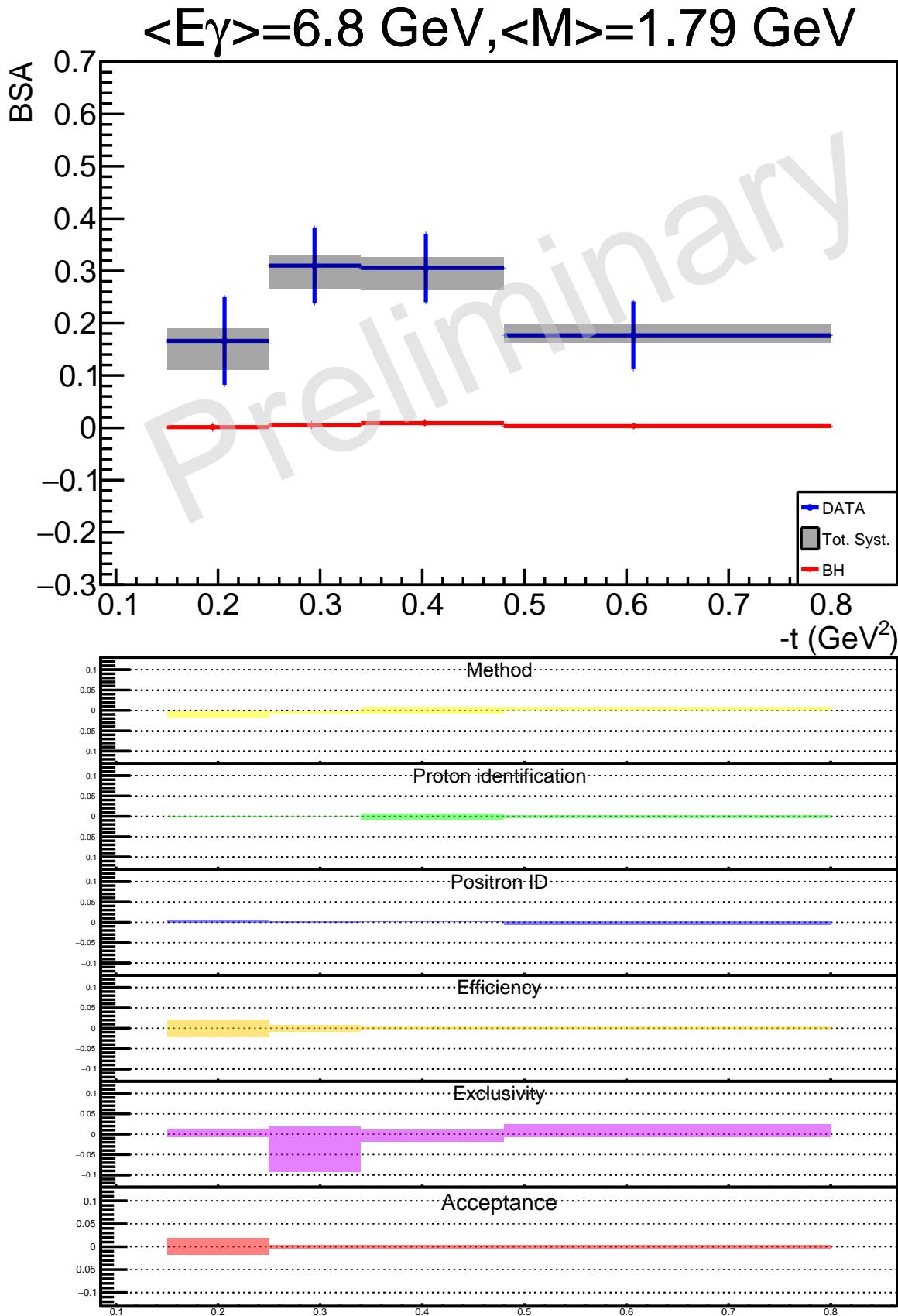


Figure 4.11: BSA as a function of  $-t$ , integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.11 in Appendix 4.2.

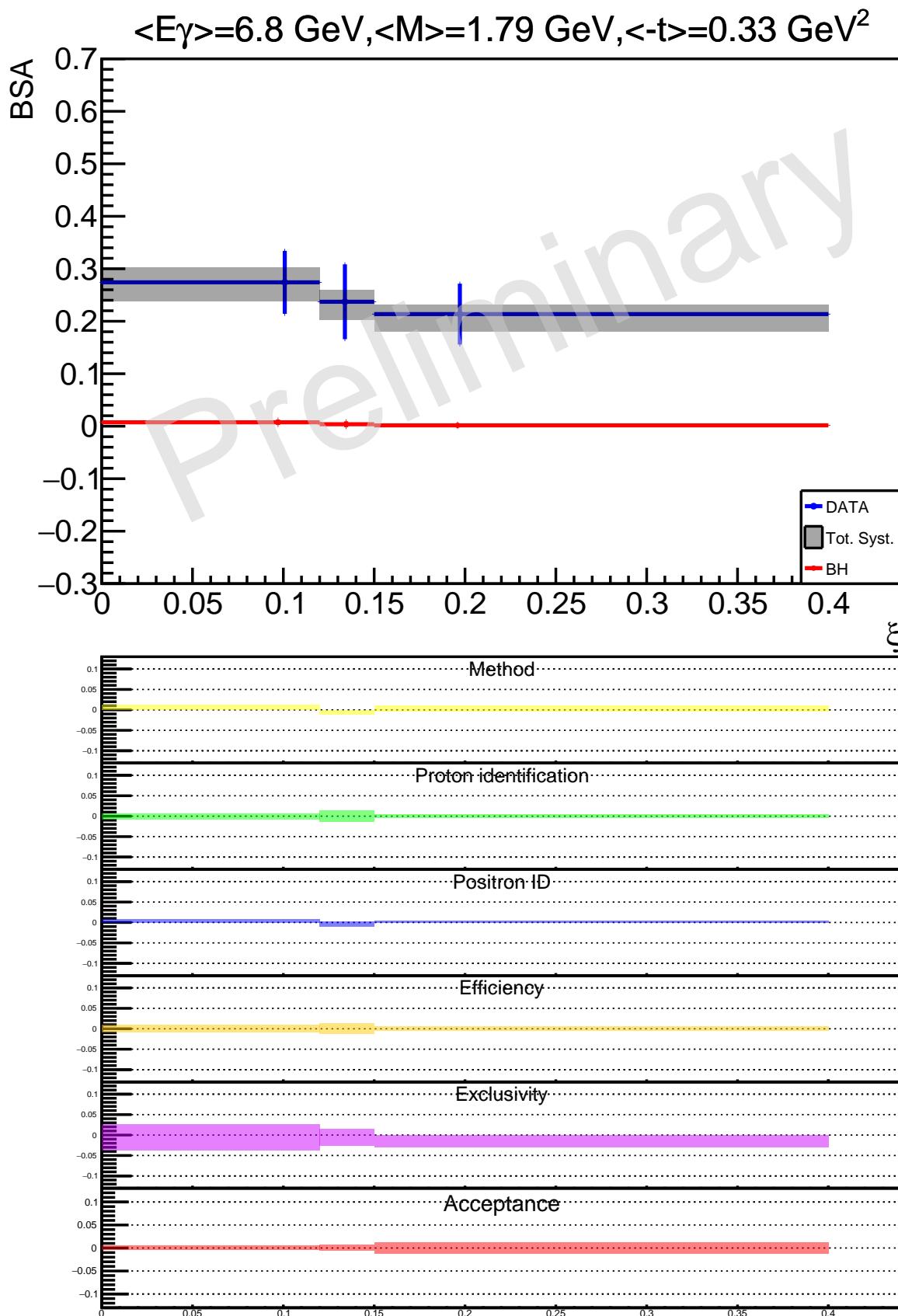


Figure 4.12: BSA as a function of  $\xi$ , integrated over all the other variables, using the same plotting conventions as in Figure 4.1. Tabulated values in Table 4.12 in Appendix 4.2.

---

## 4.2 Tabulated results

1307 The results for the three TCS observables obtained from Fall 2018 CLAS12 dataset are tabulated in  
 1309 this section, with the corresponding statistical and systematic uncertainties. Tables 4.1, 4.2 and 4.3  
 1310 contain results for the  $R'$  ratio, Tables 4.4 to 4.9 display the results for the FB asymmetry and Tables  
 1311 4.10, 4.11 and 4.12 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 4.1:  $R'$  ratio as a function of  $-t$  (see Figure 4.1).

$\xi$	$R'$	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 4.2:  $R'$  ratio as a function of  $\xi$  (see Figure 4.2).

$-t(GeV^2)$	$R'$	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 4.3:  $R'$  ratio as a function of  $-t$  in the mass range [2 GeV – 3 GeV] (see Figure 4.3).

$E\gamma(GeV)$	$A_{FB}$	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 4.4:  $A_{FB}$  as a function of  $E_\gamma$  (see 4.4).

$M(GeV)$	$A_{FB}$	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 4.5:  $A_{FB}$  as a function of  $M$  (see 4.5).

$-t(GeV^2)$	$A_{FB}$	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 4.6:  $A_{FB}$  as a function of  $-t$  (see 4.6).

$\xi$	$A_{FB}$	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 4.7:  $A_{FB}$  as a function of  $\xi$  (see Figure 4.7).

$-t(GeV^2)$	$A_{FB}$	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 4.8:  $A_{FB}$  as a function of  $-t$  in the mass range [1.5 GeV – 2 GeV] (see 4.8).

$-t(GeV^2)$	$A_{FB}$	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 4.9:  $A_{FB}$  as a function of  $-t$  in the mass range [2 GeV – 3 GeV] (see 4.9).

$M(GeV)$	$BSA$	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 4.10: BSA as a function of  $M$  (see Figure 4.10).

$-t(GeV^2)$	<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 4.11: BSA as a function of  $-t$  (see Figure 4.11).

$\xi$	<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 4.12: BSA as a function of  $\xi$  (see Figure 4.12).

### 4.3 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.

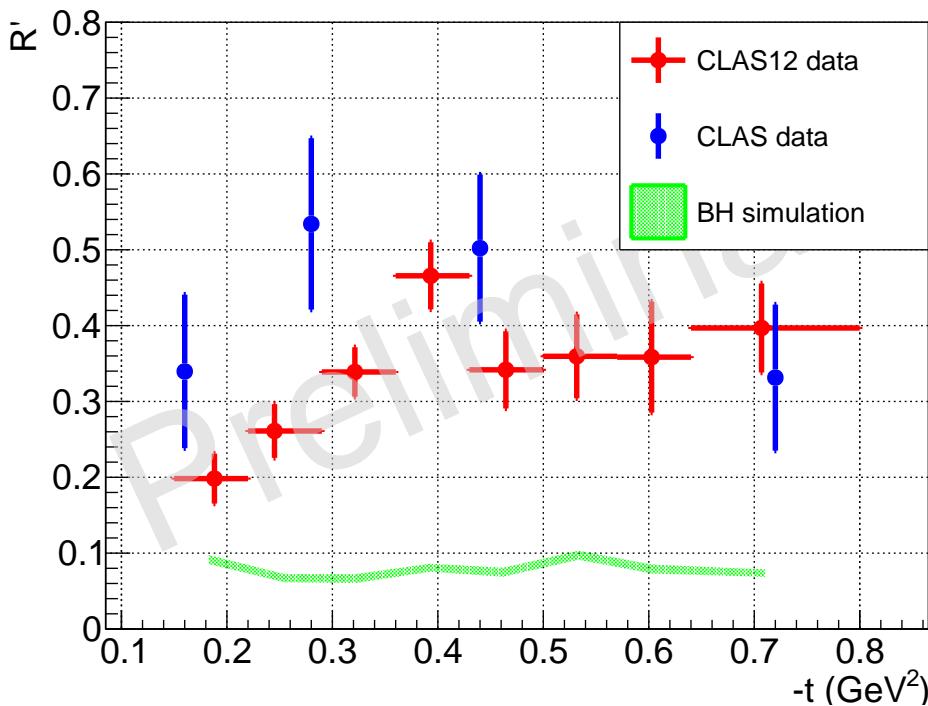


Figure 4.13: CLAS12 results for the  $R'$  ratio as a function of  $-t$  in the kinematic region accessible by CLAS, together with the CLAS results obtained in [3] and BH-only simulations for comparison. The error bars correspond to the statistical error only.

Figure 4.13 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.

## 4.4 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.

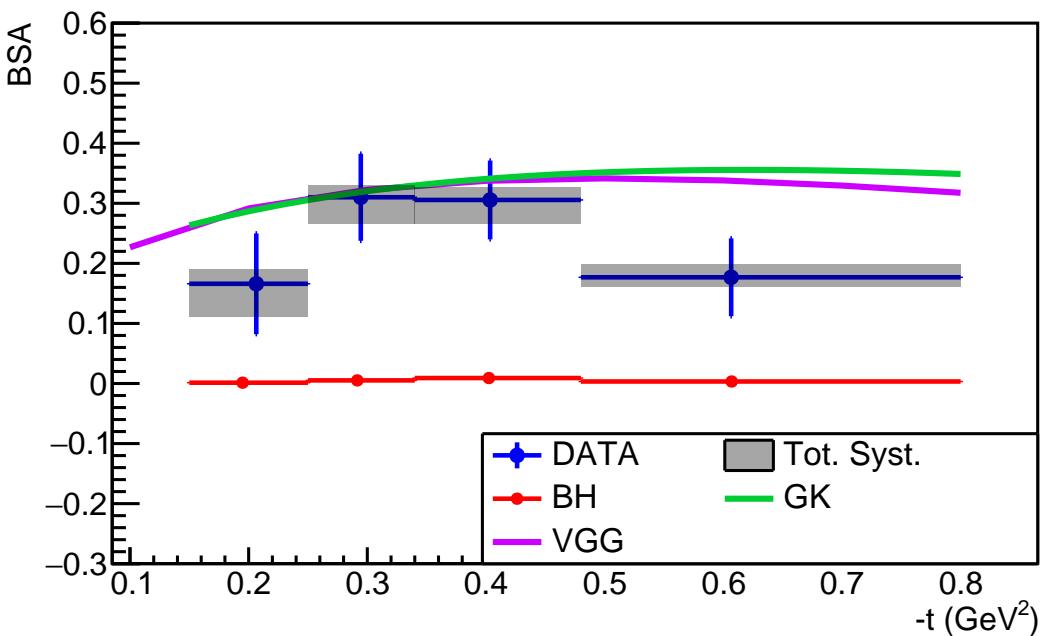


Figure 4.14: CLAS12 data points for the TCS BSA as a function of  $-t$ , evaluated at  $\phi = 90^\circ$ , integrated over CLAS12 acceptance and over all the other variables. The vertical blue error bars are statistical uncertainties while the grey bands correspond to systematic uncertainties. Three 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 4.14 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  $\theta$  integrated from  $\pi/4$  to  $3\pi/4$ . The other variables,  $E_\gamma$  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

1359 and integration are identical to those of VGG.

1360 The values of the BSA extracted from the CLAS12 data are in agreement, within error bars, with  
 1361 the three theoretical predictions. This observation tends to validate the use of the GPD formalism to  
 1362 describe TCS data and is a hint for the universality of the GPDs, as the VGG model also describes  
 1363 well the DVCS data [24]. However, our data points do not strongly favor any of the VGG hypothesis.  
 1364 Further studies on the dependence of the TCS BSA with the  $b_{sea}$  parameter should be made in order  
 1365 to identify the kinematic regions where one could discern between both values.

1366 Figure 4.15 shows the measured BSA as a function of the invariant mass of the lepton pair,  $M$ . The  
 1367 prediction obtained with the GK model, for the mean kinematic point specified above the plot and for  
 1368 angular dependencies identical to the ones in Figure 4.14, is superimposed (orange line). The GK curve  
 1369 is only displayed in the mass range between 1.5 GeV to 2.3 GeV, as values of the mass higher than 2.3  
 1370 GeV are not kinematically allowed at the mean kinematic point of this plot. Nevertheless, the data  
 1371 points are in agreement, within error bars, with the theoretical prediction in the [1.5 GeV – 2.3 GeV]  
 1372 mass range. This is an indication that the GPD factorization seems to apply to TCS in a large part  
 1373 of the mass range accessible by CLAS12.

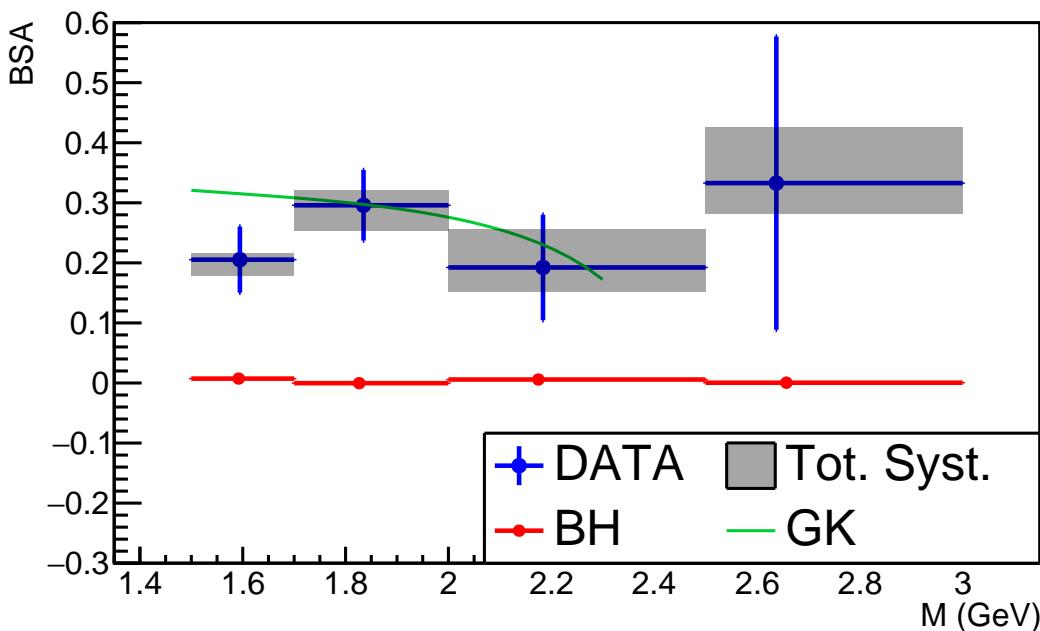


Figure 4.15: CLAS12 data points for the BSA as a function of  $M$ . The blue error bars are statistical uncertainties while the grey bands represent the systematic errors. The orange curve is the GK model prediction at the mean kinematic point of the data. The red points are the expected values for BH only (from simulations).

#### 1374 $A_{FB}$ interpretation

1375 The  $A_{FB}$  asymmetry has the advantage that it can be easily compared to theory, as it does not  
 1376 involve acceptance limits. Also, as shown in Section 3.7, it has a large sensitivity to the D-term,  
 1377 making it a valuable observable to extract this quantity.

1378 The  $t$ -dependence of the  $A_{FB}$  extracted from data is compared to theoretical predictions in three  
 1379 cases. In all three cases the GK predictions are calculated for the average  $E_\gamma$  and the invariant mass  
 1380 of the lepton pair given above each plot. For the VGG predictions the mean kinematics are given in  
 1381 the corresponding figure captions.

1382 In Figure 4.16 the  $A_{FB}$  asymmetry data points are plotted against  $-t$ , and all the other variables  
 1383 are integrated over the phase space detailed in Section 3.11. They are compared to predictions obtained

1384 using the VGG model for  $M = 1.8$  GeV (red lines) and for different values of  $\theta_0$  (at the lower edge  
 1385 of the angular bin  $\theta_0 = 50^\circ$  (dashed), at its center  $\theta_0 = 65^\circ$  (solid), and at its upper edge  $\theta_0 = 80^\circ$   
 1386 (dash-double-dotted)). The effect of changing the average mass is also illustrated with the green curve  
 1387 calculated for  $M = 1.5$  GeV. The cyan curve is calculated without the contribution of the D-term to  
 1388 the GPD  $H$  (the D-term contribution used for this calculation is the one described in [25]). Finally the  
 1389 GK prediction (orange) shown here is obtained by integrating the BH-TCS cross section in the forward  
 1390 and backward angular bin defined in Section 3.8. Also it has to be noticed that the GK prediction  
 1391 does not include the contribution of the D-term.

1392 As already mentionned in Section 3.7, the VGG curves produced with and without the D-term  
 1393 clearly indicate that the D-term contribution to the GPD  $H$  has a large effect on the value of FB  
 1394 asymmetry. The data points are better described by the VGG model when the D-term is included,  
 1395 although error bars are still too large to completely rule out the case without the D-term. The GK  
 1396 model prediction seems to largely underestimate the asymmetry. This could be explained by the  
 1397 absence of D-term in this prediction. In order to ensure that this interpretation is valid for the whole  
 1398 mass range studied, the same comparison is done in the low-mass region [1.5 GeV – 2 GeV] and in  
 1399 the high-mass region [2 GeV – 3 GeV]. Indeed low mass events dominate when the asymmetry is  
 1400 integrated over the full mass range [1.5 GeV – 3 GeV], therefore low-mass vector-meson resonances  
 1401 (e.g.  $\rho(1450)$ ) could be the origin of the observed FB asymmetry.

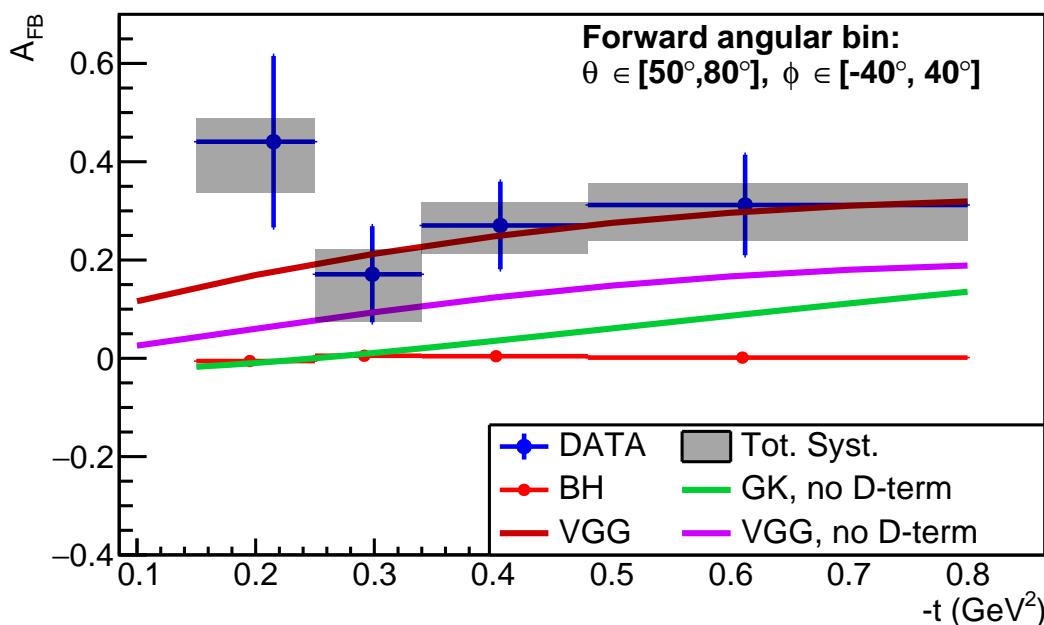


Figure 4.16: CLAS12  $A_{FB}$  as a function of  $-t$ , integrated over all other variables. The data points are compared with theory predictions realized using the VGG and GK models. The cyan line is calculated with VGG and without the D-term contribution to  $H$ . The red lines correspond to three different  $\theta_0$  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).

1402 Figure 4.17 shows the data points extracted in the [1.5 GeV – 2 GeV] low-mass region. They are  
 1403 compared with the same VGG predictions as in the full mass range case, as the average mass and the  
 1404 average photon energy do not change dramatically. The GK prediction is however recalculated at the  
 1405 mean kinematic point of this plot. The data points do not change substantially when the mass range  
 1406 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 3.6c. Therefore the conclusions drawn for Figure 4.16 also apply to Figure 4.17. Again the data points are better described when the D-term contribution is taken into account.

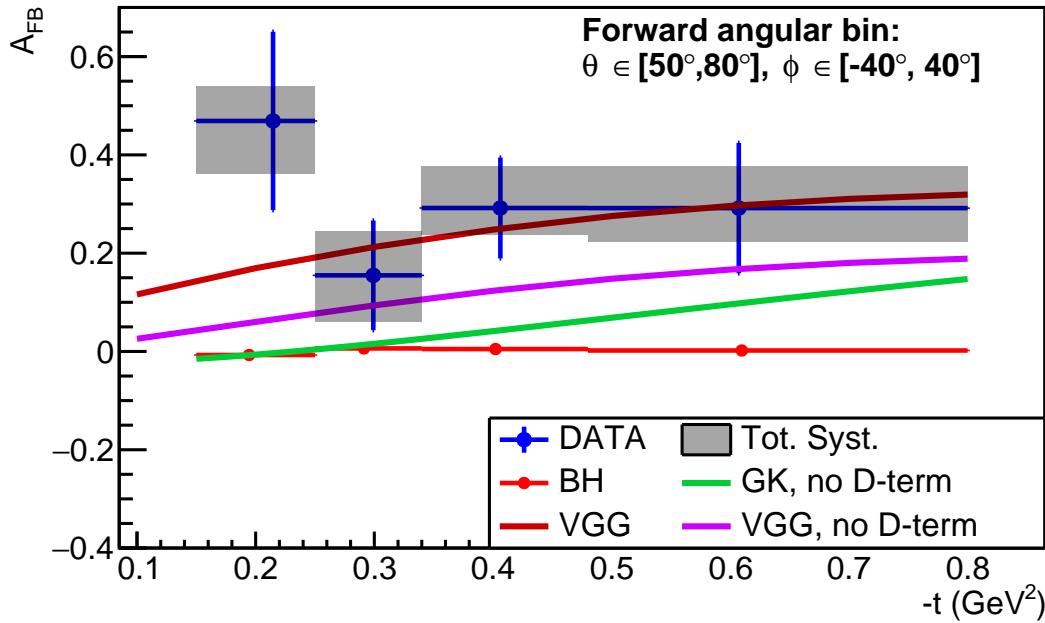


Figure 4.17: CLAS12  $A_{FB}$  as a function of  $-t$ , integrated over all the other variables, with the invariant mass of the lepton pair integrated in the range [1.5 GeV – 2 GeV]. The VGG curves are the same as in Figure 4.16. The GK prediction is calculated at the mean kinematic point. The red points 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 4.18. 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 4.19 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 4.16 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 4.20 displays the  $A_{FB}$  data points as a function of the photon energy,  $E_\gamma$ . The GK prediction for this observable is also shown. As already observed in Figures 4.16 and 4.19, the GK predictions largely underestimate the measured FB asymmetry.

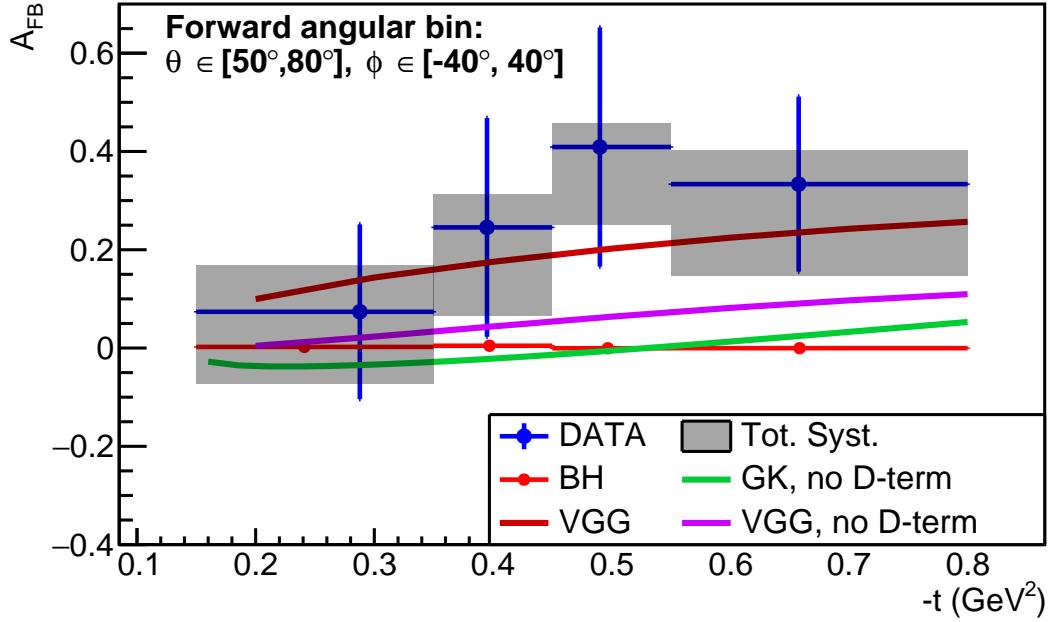


Figure 4.18: CLAS12  $A_{FB}$  as a function of  $-t$ , integrated over all the other variables and with the invariant mass integrated in the range [2 GeV – 3 GeV]. The VGG model cyan curve does not take into account the D-term contribution. The plain red curve is calculated using VGG with the invariant 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).

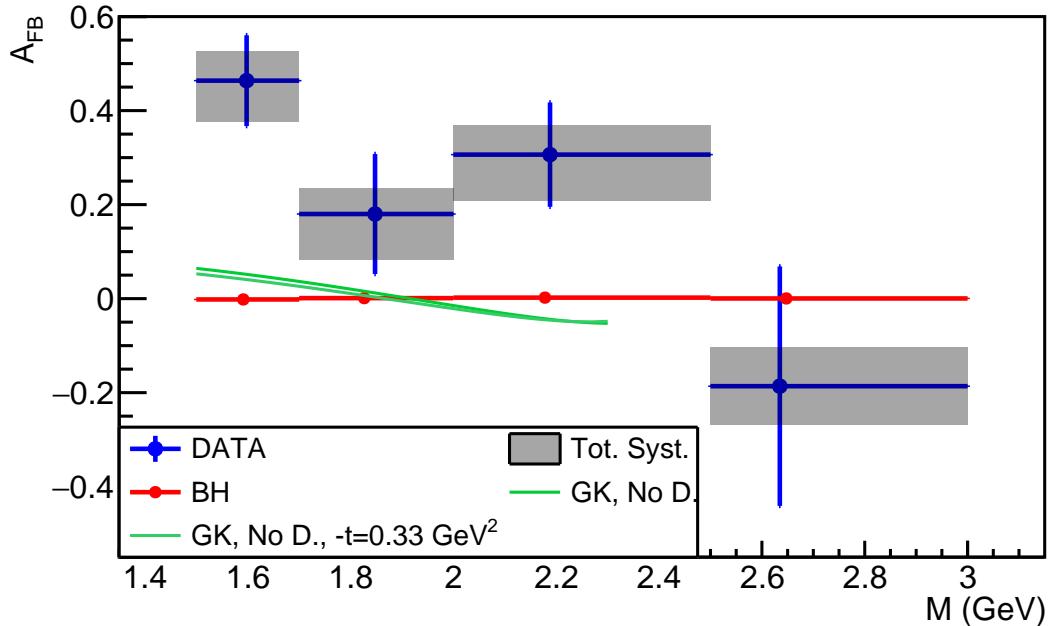


Figure 4.19: CLAS12  $A_{FB}$  as a function of  $M$ , integrated over all the other variables. The vertical blue error bars are statistical errors and the grey bands are systematic uncertainties. The orange and brown curves are two predictions obtained with the GK model, for  $-t = 0.37 \text{ GeV}^2$  and  $-t = 0.33 \text{ GeV}^2$ , respectively. The red points are the expected values for BH only (from simulations).

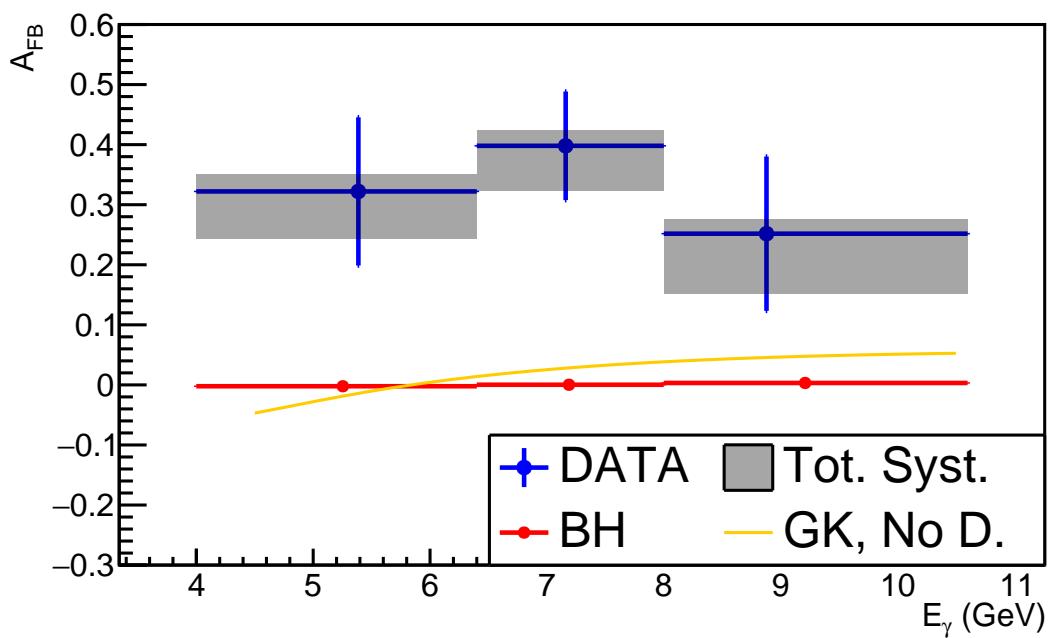


Figure 4.20: CLAS12  $A_{FB}$  as a function of  $E_\gamma$ , integrated over all the other variables. The vertical blue error bars are statistical errors and the grey bands are systematic uncertainties. The orange curve is obtained with the GK model. The red points are the expected value for BH only (from simulations).

## Appendices



<sup>1435</sup> **Appendix A**

<sup>1436</sup> **Derivation of the background/signal ratio**

<sup>1437</sup> Let  $B(x)$  and  $S(x)$  respectively be the number of background (mis-id. pions) and signal (true positron)  
<sup>1438</sup> events in the TCS sample for a background strength  $x$  ( $x \in [0, 1]$ ).

<sup>1439</sup> The number of background events in the TCS sample is linear with the background strength  $x$ :

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

<sup>1440</sup> where  $\beta$  is the number of background events in the TCS sample when no cut is applied.

<sup>1441</sup> The function  $S(x)$  is unknown but we assume it is increasing with  $x$  (when background is removed,  
<sup>1442</sup> signal events might also be removed by mistake), and does not vary much with  $x$  (signal events should  
<sup>1443</sup> not be removed by the classifier).

<sup>1444</sup> The normalized number of TCS events can then be described in the linear region with the following  
<sup>1445</sup> expression:

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

<sup>1446</sup> Although this formula is only applicable in the linear region, we can extrapolate it to  $x = 0$ :

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

<sup>1447</sup> Assume the chosen cut yields to a normalized background strength  $x_0$  in the linear region. The  
<sup>1448</sup> corresponding normalized number of TCS events is:

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

<sup>1449</sup> We want to estimate the background/signal ratio for a given normalized background strength  $x_0$ ,  
<sup>1450</sup>  $\frac{B(x_0)}{S(x_0)}$ . Solving  $B(x_0)$  from Equations (A.1) and (A.4) one can write for the B/S ratio at  $x_0$ :

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

<sup>1451</sup> 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  
<sup>1452</sup> seen in Figure 2.20 of Chapter 2. Therefore one finds:

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

<sup>1453</sup> and the quantity  $\frac{y(x_0)}{y(0)} - 1$  gives a good estimate of the ratio  $\frac{B(x_0)}{S(x_0)}$ .



1454 Appendix B

1455 Generator checks: comparison between  
1456 *GRAPE* and *TCSGen*

1457 Bethe-Heitler events have been generated using *TCSGen* and *GRAPE* within the following phase  
1458 space:

- 1459 • Lepton momenta bigger than 1 GeV
- 1460 • Lepton polar angle in the lab in CLAS12 acceptance  $5^\circ < \theta_{\text{Lab e}^{+/-}} < 45^\circ$
- 1461 •  $0.01 \text{ GeV}^2 < -t < 1 \text{ GeV}^2$
- 1462 •  $1.7 \text{ GeV} < M < 3.08 \text{ GeV}$
- 1463 •  $2 \text{ GeV} < E_\gamma < 10.6 \text{ GeV}$
- 1464 •  $40^\circ < \theta_{\text{COM}} < 140^\circ$ ,

1465 and the results of both generators have been compared.

1466 The following plots in Figures B.1 to B.5 show the distributions of all five relevant TCS variables,  
1467 obtained for both generators as well as their ratio. Each distribution is normalized by the integral of  
1468 the  $E_\gamma$  distribution obtained for each generator.

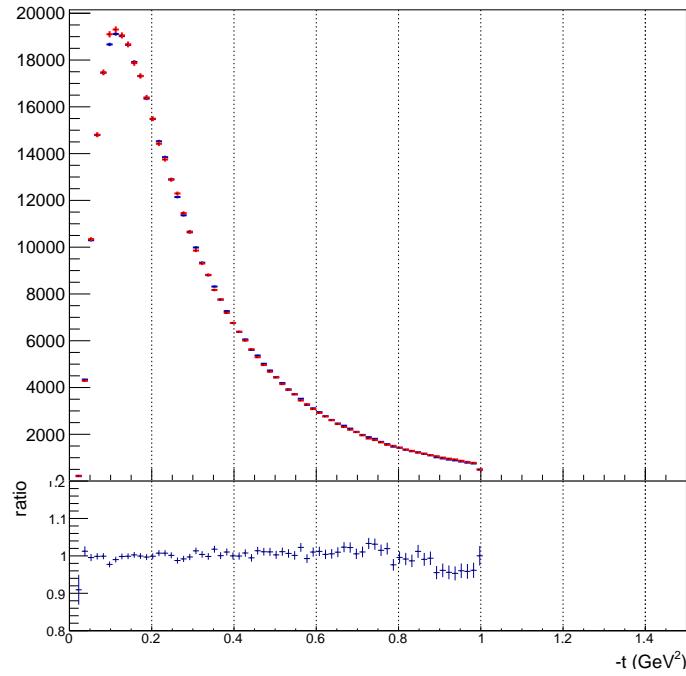


Figure B.1: Comparison of the generated proton momentum transfer squared  $-t$  distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

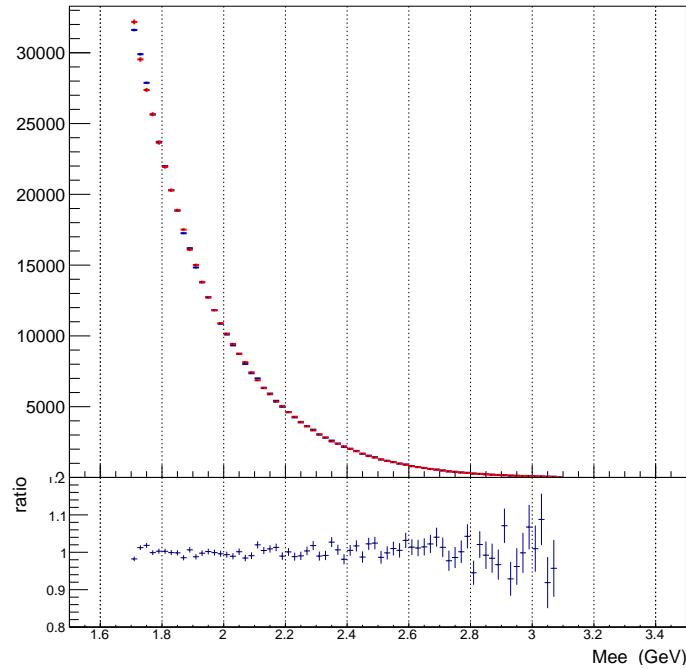


Figure B.2: Comparison of the generated invariant mass of the lepton pair  $M$  distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

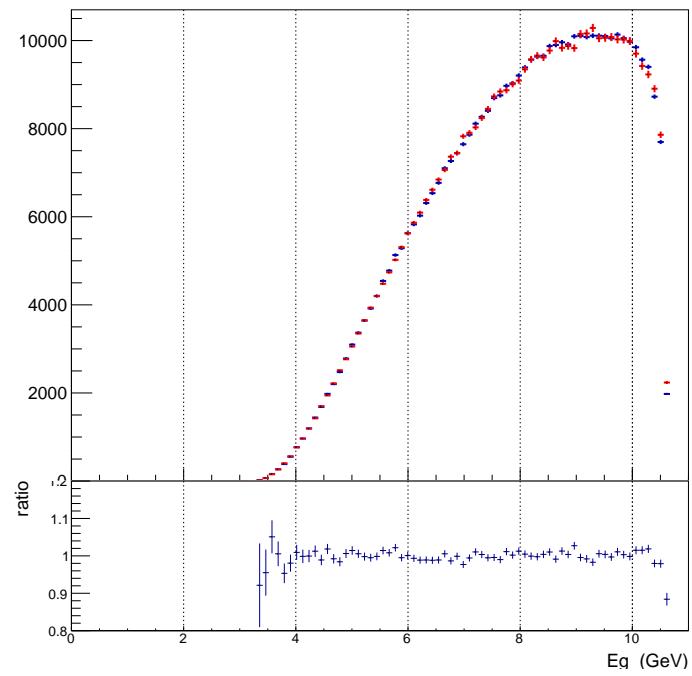


Figure B.3: Comparison of the generated photon energy  $E_\gamma$  distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

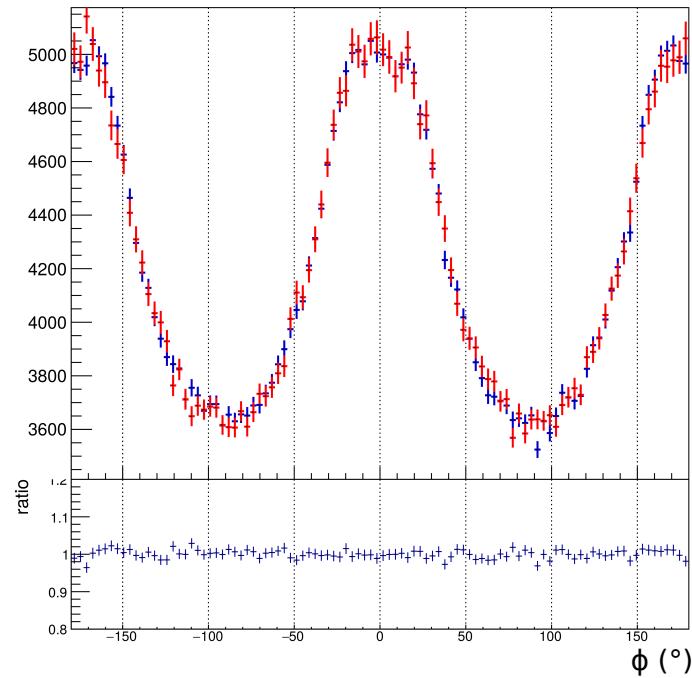


Figure B.4: Comparison of the generated COM azimuthal angle  $\phi$  distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

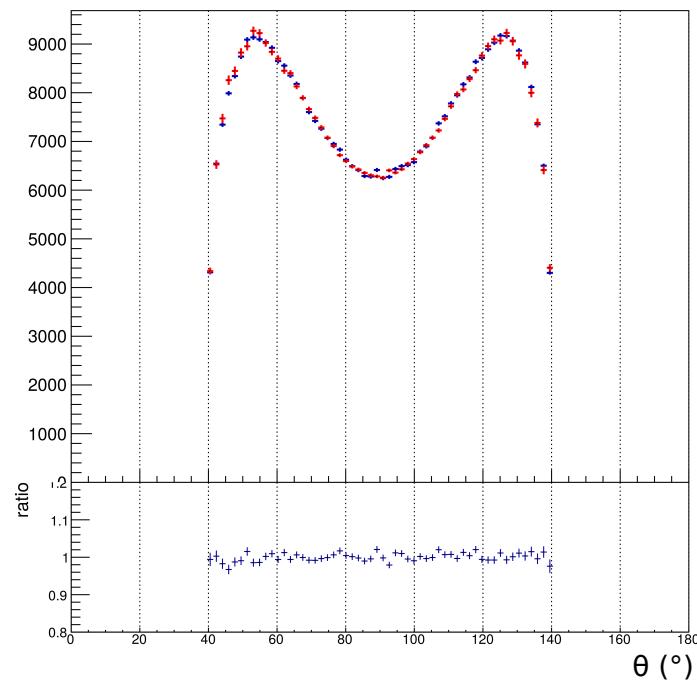


Figure B.5: Comparison of the generated COM polar angle  $\theta$  distributions obtained with *TCSGen* (in red) and *GRAPE* (in blue), as well as their ratio.

<sup>1469</sup> **Appendix C**

<sup>1470</sup> **Final state particle kinematics**

<sup>1471</sup> Figures C.1 and C.2 in this appendix show the kinematics of the three final state particles of the TCS  
<sup>1472</sup> reaction in the CLAS12 data. The top three plots of Figure C.1 show these kinematics for all events  
<sup>1473</sup> in CLAS12 data in which a proton, an electron and a positron are detected. The three bottom plots  
<sup>1474</sup> in the same figure, show the kinematic distributions once the exclusivity cuts presented in Section 3.2  
<sup>1475</sup> are applied. The plots displayed in Figure C.2 show the kinematics of the final state particles for the  
<sup>1476</sup> events selected in the analysis of this manuscript.

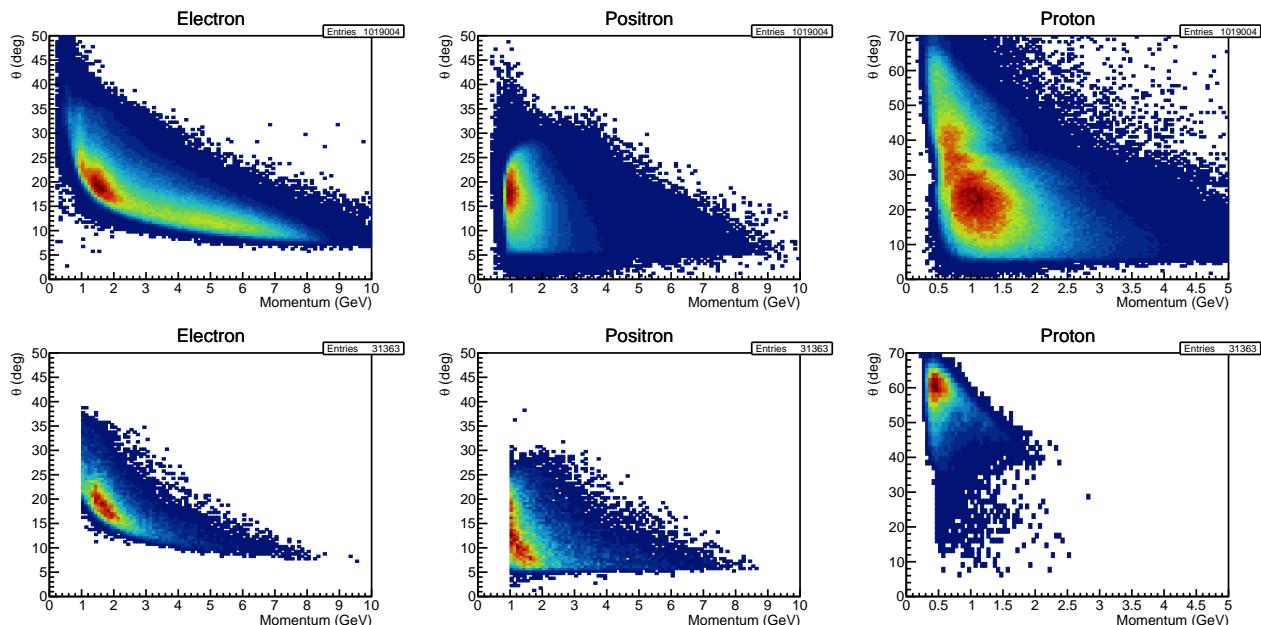


Figure C.1: Top: Kinematics of the TCS final state particles (polar angle in the laboratory frame vs momentum) for events with one proton and two opposite-sign leptons. Bottom: Same plots after the exclusivity cuts to select the photoproduction events are applied.

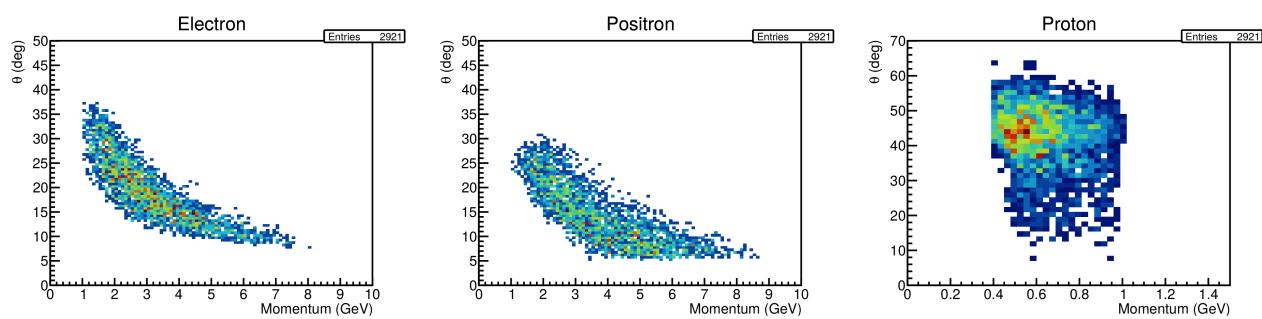


Figure C.2: Kinematics of the final state particles for TCS events with invariant mass between 1.5 and 3 GeV.

# <sup>1477</sup> Bibliography

- <sup>1478</sup> [1] E.R. Berger, M. Diehl, and B. Pire. Timelike compton scattering: exclusive photoproduction of  
<sup>1479</sup> lepton pairs. *The European Physical Journal C - Particles and Fields*, 23(4):675–689, Apr 2002.
- <sup>1480</sup> [2] Boér, M., Guidal, M., and Vanderhaeghen, M. Timelike compton scattering off the proton and  
<sup>1481</sup> generalized parton distributions. *Eur. Phys. J. A*, 51(8):103, 2015.
- <sup>1482</sup> [3] R. Paremuzyan. *Timelike Compton Scattering*. PhD thesis, Yerevan Physics Institute, 2010.
- <sup>1483</sup> [4] Markus Diehl, Thierry Gousset, Bernard Pire, and John P. Ralston. Testing the handbag contribu-  
<sup>1484</sup> tion to exclusive virtual compton scattering. *Physics Letters B*, 411(1):193 – 202, 1997.
- <sup>1485</sup> [5] Oleksii Gryniuk and Marc Vanderhaeghen. Accessing the real part of the forward  $j/\psi-p$  scattering  
<sup>1486</sup> amplitude from  $j/\psi$  photoproduction on protons around threshold. *Phys. Rev. D*, 94:074001, Oct  
<sup>1487</sup> 2016.
- <sup>1488</sup> [6] Y.G. Sharabian et al. The clas12 high threshold cherenkov counter. *Nuclear Instruments and*  
<sup>1489</sup> *Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated*  
<sup>1490</sup> *Equipment*, page 163824, 2020.
- <sup>1491</sup> [7] Andreas Hoecker et al. Tmva: Toolkit for multivariate data analysis. *PoS*, ACAT:040, 2007.
- <sup>1492</sup> [8] CLAS collaboration. Fiducial studies. [https://clasweb.jlab.org/wiki/index.php/Run\\_Group\\_A#tab=Fiducial\\_Studies](https://clasweb.jlab.org/wiki/index.php/Run_Group_A#tab=Fiducial_Studies), 2020.
- <sup>1493</sup> [9] Tetsuo Abe. Grape-dilepton (version 1.1): A generator for dilepton production in ep collisions.  
<sup>1494</sup> *Computer Physics Communications*, 136(1):126 – 147, 2001.
- <sup>1495</sup> [10] Paul Kessler. Photon fluxes and the EPA. In *Two-Photon Physics at LEP and HERA - Status of*  
<sup>1496</sup> *Data and Models*, pages 183–190, 5 1994.
- <sup>1497</sup> [11] Eero Byckling and K. Kajantie. *Particle Kinematics*. University of Jyvaskyla, Jyvaskyla, Finland,  
<sup>1498</sup> 1971.
- <sup>1499</sup> [12] CLAS collaboration. Genev webpage. [https://clasweb.jlab.org/wiki/index.php/Event\\_generator](https://clasweb.jlab.org/wiki/index.php/Event_generator), 2014.
- <sup>1500</sup> [13] M. Guidal, M. V. Polyakov, A. V. Radyushkin, and M. Vanderhaeghen. Nucleon form factors  
<sup>1501</sup> from generalized parton distributions. *Phys. Rev. D*, 72:054013, Sep 2005.
- <sup>1502</sup> [14] Michel Guidal, Hervé Moutarde, and Marc Vanderhaeghen. Generalized parton distributions  
<sup>1503</sup> in the valence region from deeply virtual compton scattering. *Reports on Progress in Physics*,  
<sup>1504</sup> 76(6):066202, 2013.
- <sup>1505</sup> [15] Paweł Sznajder. private communication.
- <sup>1506</sup> [16] B. Berthou, D. Binosi, N. Chouika, L. Colaneri, M. Guidal, C. Mezrag, H. Moutarde, J. Rodríguez-  
<sup>1507</sup> Quintero, F. Sabatié, P. Sznajder, and J. Wagner. Partons: Partonic tomography of nucleon  
<sup>1508</sup> software. *The European Physical Journal C*, 78(6):478, Jun 2018.
- <sup>1509</sup>
- <sup>1510</sup>

- 1511 [17] K. Goeke, M.V. Polyakov, and M. Vanderhaeghen. Hard exclusive reactions and the structure of  
1512 hadrons. *Progress in Particle and Nuclear Physics*, 47(2):401 – 515, 2001.
- 1513 [18] Raphael Dupre, Michel Guidal, and Marc Vanderhaeghen. Tomographic image of the proton.  
1514 *Physical Review D*, 95, 06 2016.
- 1515 [19] Haakon Olsen and L. C. Maximon. Photon and electron polarization in high-energy  
1516 bremsstrahlung and pair production with screening. *Phys. Rev.*, 114:887–904, May 1959.
- 1517 [20] Jonathan Dumas. *Feasibility studies of a polarized positron source based on the bremsstrahlung of*  
1518 *polarized electrons*. PhD thesis, Université Grenoble Alpes, 2011.
- 1519 [21] M. Vanderhaeghen, P. A. M. Guichon, and M. Guidal. Hard electroproduction of photons and  
1520 mesons on the nucleon. *Phys. Rev. Lett.*, 80:5064–5067, Jun 1998.
- 1521 [22] M. Vanderhaeghen, P. A. M. Guichon, and M. Guidal. Deeply virtual electroproduction of photons  
1522 and mesons on the nucleon: Leading order amplitudes and power corrections. *Phys. Rev. D*,  
1523 60:094017, Oct 1999.
- 1524 [23] S. V. Goloskokov and P. Kroll. An attempt to understand exclusive  $\pi^+$  electroproduction. *The*  
1525 *European Physical Journal C*, 65(1):137, Nov 2009.
- 1526 [24] R. Dupré, M. Guidal, S. Niccolai, and M. Vanderhaeghen. Analysis of deeply virtual compton scat-  
1527 tering data at jefferson lab and proton tomography. *The European Physical Journal A*, 53(8):171,  
1528 Aug 2017.
- 1529 [25] B. Pasquini, M.V. Polyakov, and M. Vanderhaeghen. Dispersive evaluation of the d-term form  
1530 factor in deeply virtual compton scattering. *Physics Letters B*, 739:133 – 138, 2014.