A Machine Learning approach for DVCS identification without proton detection

Juan Sebastian Alvarado

IJCLab - Orsay

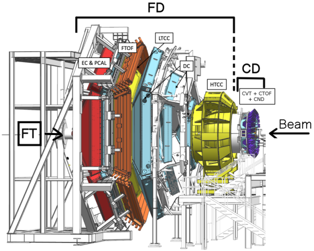
HUGS 2023

15/06/2023

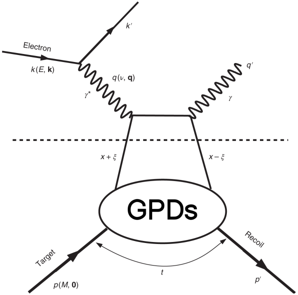
INTRODUCTION

Introduction

Jefferson Lab
Exploring the Nature of Matter



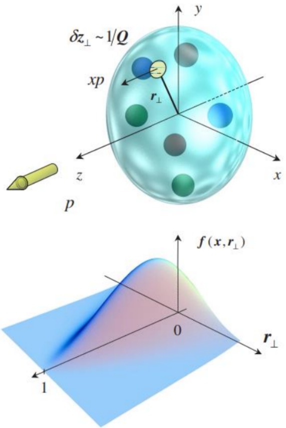
(a) CLAS12 detector



(b) Handbag diagram for DVCS.

- CLAS12 @ Hall B provides measurements of particles in the range $2^\circ < \theta < 135^\circ$.
- We aim for Deeply Virtual Compton Scattering event reconstruction as it is a golden channel for GPD measurements

Introduction



But what are GPDs?

- At low energy QCD we need structure functions to describe the nucleon structure.
- They correlate the transverse position and longitudinal momentum of partons in the nucleon & the spin structure.
- Experimental observables are directly related to Compton Form Factors $\mathcal{H}$

$$\mathcal{H} = \sum_q e_q^2 \left\{ i\,\pi\,[H^q(\xi,\xi,t) - H^q(-\xi,\xi,t)] \;+\; \mathcal{P}\int_{-1}^{1} dx H^q(x,\xi,t)\left[\frac{1}{\xi-x} - \frac{1}{\xi+x}\right] \right\}$$

Models need to map the x dependence

Introduction

In principle, the measurement of only an electron and a photon is enough to reconstruct a DVCS event. We aim for DVCS event reconstruction without requiring final proton information.

Advantages (with respect to $ep\gamma$ detection):

- Higher statistics:
  - Gives access to a wider phase space for GPD studies.
  - Improves GPD studies at low $-t$.
  - More precise BSA measurements.
- Helpful for experiments that do not consider proton detection.

Difficulties:

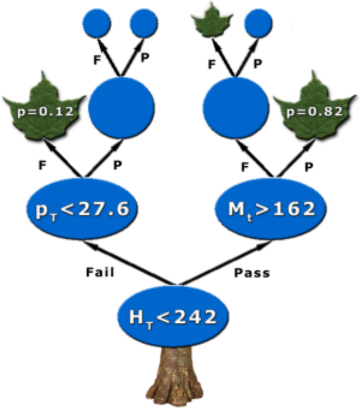
- The $e\gamma$ final state includes background contributions from the whole DIS spectra.
- Only one exclusivity variable available for cuts: Missing mass of $ep \to e\gamma$.

Therefore, we need a method that ensures DVCS identification: Machine Learning

We test the ML approach on experimental data:

1. Validation of the method when we include the proton information.
2. Application to the case without proton information.

ML approach: Boosted Decision Trees (BDT)



Taken from Coadou, Yann. EPJ Web of conferences. Vol. 55. EDP Sciences, 2013.

A decision tree:

- Scans the given variables looking for the point with maximum separation between classes.
- Splits the data recursively until each event lies on a terminal node (leaf) and assigns it a score.

Additionally, boosting is:

- Train iteratively a decision tree.
- At each step, focus the training on misclassified events.
- Final classification is based on the majority of votes.

UNPOLARIZED LH_2

Analysis of $ep \to ep\gamma$

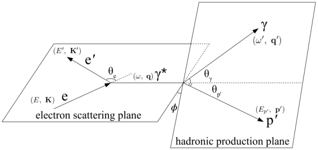
$ep \to e\gamma p$: Data selection

Analyzed data set

- Unpolarized liquid hydrogen target.

Kinematic window:

- $W > 2$ GeV,
- $Q^2 > 1$ GeV^2 ,
- $\mathbf{q}' > 2$ GeV (photon),
- $\mathbf{k}' > 1$ GeV (electron),
- $\mathbf{p}' > 0.3$ GeV (nucleon).



Exclusivity cuts:

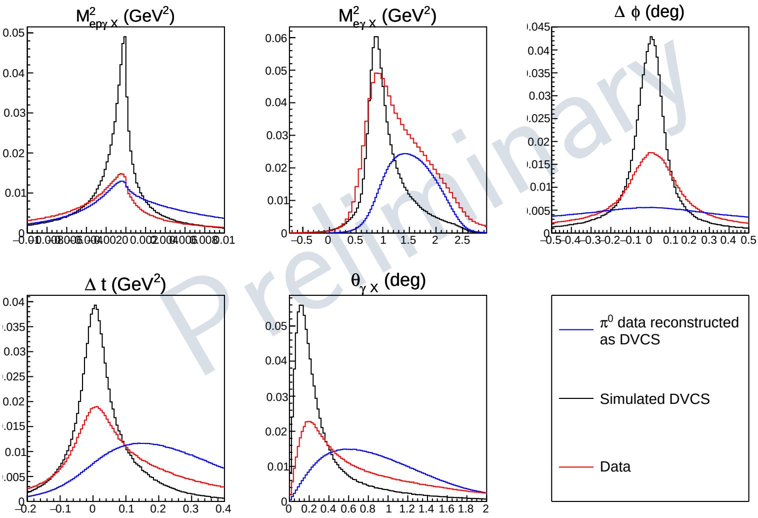
We reconstruct ϕ and t in two ways:

1. Using $\gamma*$ and the outgoing photon γ: $\Rightarrow \phi(\gamma)$
2. Using $\gamma*$ and the recoil proton p: $\Rightarrow \phi(p')$

- $\Delta\phi = |\phi(p') - \phi(\gamma)| < 2°$,
- $\Delta t = |t(p') - t(\gamma)| < 2$ GeV^2,
- $\mathbf{P}_{miss} < 1$ GeV.

$ep \to e\gamma p$: Model training

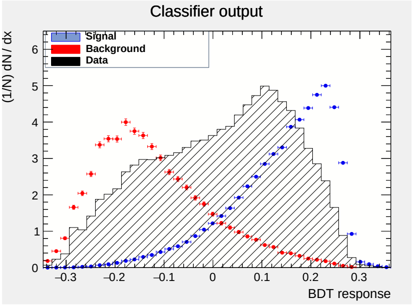
The main contamination channel is $ep \to ep\pi^0 \to ep\gamma(\gamma)$.



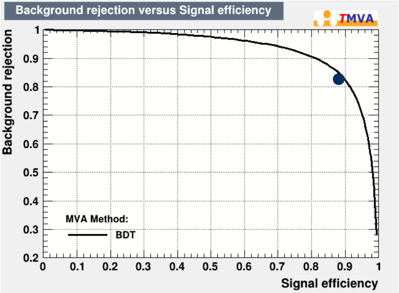
$ep \to e\gamma p$: Background subtraction

To optimize the DVCS event selection, a Boosted Decision Tree (BDT) is trained to classify the events.

- Discriminating variables: $\{M^2_{ep\gamma}, M^2_{e\gamma}, \Delta\phi, \Delta t, \theta_{\gamma X}\}$.
- Simulated DVCS and π^0 production events for training.



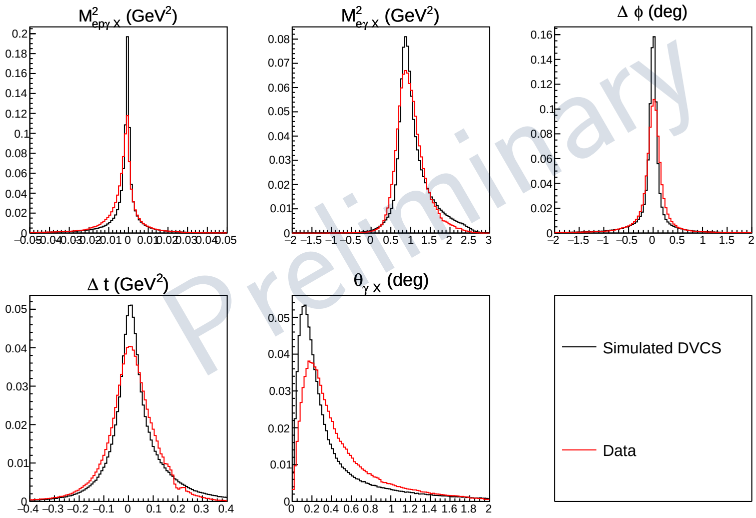
(a) BDT output distributions for different datasets.



(b) ROC curve of the model and applied cut.

$ep \to e\gamma p$: Background substraction

We extract a dataset with $> 89.67\%$ DVCS and $< 10.32\%$ DVMP.



UNPOLARIZED LH_2

Analysis of $ep \to e(p)\gamma$

$ep \to e\gamma(p)$: Data selection

Kinematic window:

We apply the same kinematic restrictions:

- $W > 2$ GeV,
- $Q^2 > 1$ GeV^2 ,
- $\mathbf{q}' > 2$ GeV (photon),
- $\mathbf{k}' > 1$ GeV (electron).
- $-\frac{t}{Q^2} < 1$,

Exclusivity cuts:

However, our exclusivity cuts are no longer useful.

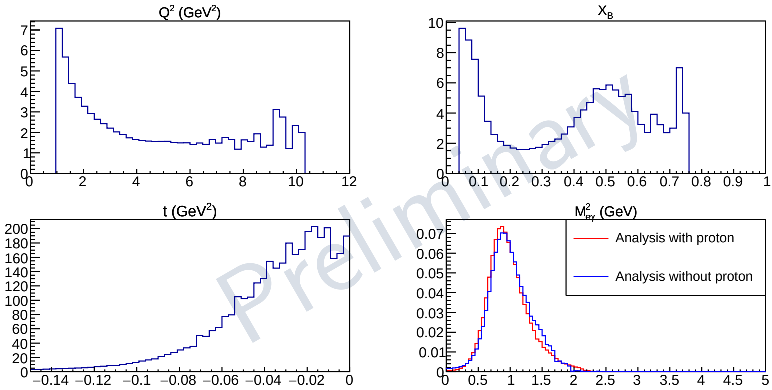
- $\cancel{\Delta\phi = |\phi(p) - \phi(\gamma)| mod(180) < 2^\circ}$,
- $\cancel{\Delta t = |t(p) - t(\gamma)| < 2 \text{ GeV}^2}$,
- $\cancel{\mathbf{P}_{miss} < 1 \text{ GeV}}$.

BDT training:

- Training for π^0 and DIS.
- Discriminating variables: $\{M^2_{e\gamma X}, M^2_{eX}, M^2_{\gamma X}\}$.
- Dataset is splitted in two:
  - Events where the photon is in the FT ($\theta_\gamma < 5^\circ$)
  - Events where the photon is in the FD ($\theta_\gamma > 5^\circ$)

$ep \to e\gamma(p)$: Comparison with $e\gamma p$ detection

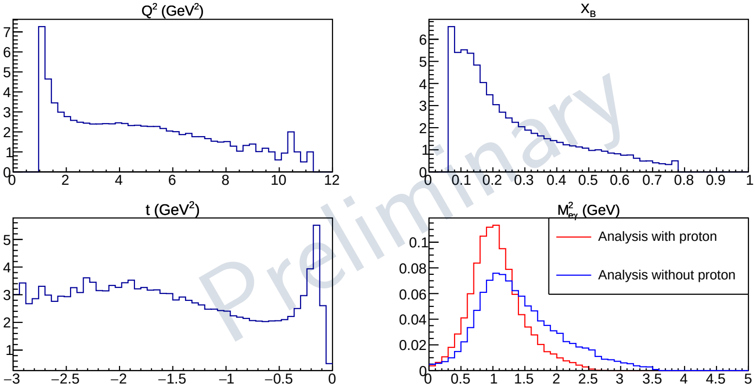
There are more events in general, mostly in the small t region.



$\frac{\text{No proton}}{\text{proton}}$ ratios and $M^2_{e\gamma X}$ for $\theta_\gamma < 5^\circ$

$ep \to e\gamma(p)$: Comparison with $e\gamma p$ detection

There are more events in general, mostly in the small t region.

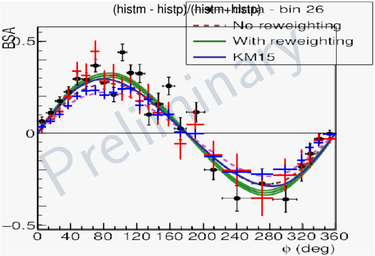


No proton:proton ratios and $M^2_{e\gamma X}$ for $\theta_\gamma > 5^\circ$

Raw Beam Spin Asymmetry

$$A_{LU} \equiv \frac{1}{P}\frac{h^+ - h^-}{h^+ + h^-} \sim \frac{p_0 \; \sin(\phi)}{1 + p_1 \; \cos(\phi)} \sim \frac{\sin(\phi)}{\sigma_{UU}} \Im\left[F_1\mathcal{H} + \xi(F_1 + F_2)\tilde{\mathcal{H}} - kF_2\mathcal{E} + ...\right]$$

$$1.8 \text{ GeV}^2 < Q^2 < 2.4 \text{ GeV}^2 \qquad 0.16 < x_B < 0.26 \qquad t > -0.2 \text{ GeV}^2$$



- BSA measurement from CLAS12 **analysis note (black)***.
- Raw** BSA from the current analysis with proton detection (red)
- Raw** BSA from the current analysis without proton detection (blue)

*G. Christiaens et al. "Deeply Virtual Compton Scattering on proton: Beam Spin Asymmetry extraction". In: CLAS12 Analysis Note (2021).
** Before subtraction of the residual π^0 contamination.

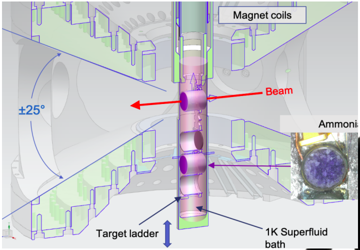
TRANSVERSELY POLARIZED NH_3

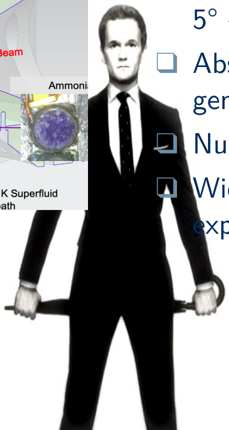
Analysis of $ep \to e(p)\gamma$

NH_3 target

CHALLENGE ACCEPTED.

This experiment presents additional complications:





- Limited acceptance: $5^\circ < \theta < 25^\circ$.
- Absence of a DIS event generator with NH_3 target.
- Nuclear target effects.
- Wider distributions are expected.

NH_3 target

The plan for the moment is:

- Make the best with the LH_2 target data.
- Analyze data from longitudinally polarized NH_3 target.
- Try other methods (sweights, NNs etc.)
- Find the minimum amount of information needed to develop the experiment.

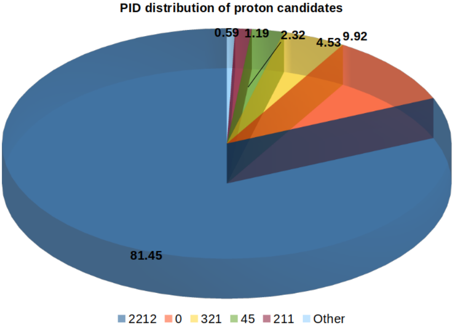


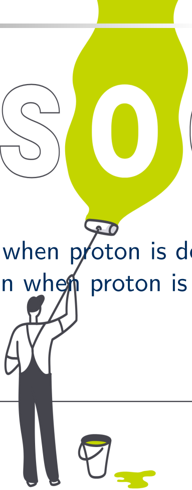
Figure: Preliminary results when a particle tagger is included.

Summary

- Boosted decision trees presents an alternative for channel selection on an event-by-event basis.
- When the final proton is included:
  - DVCS exclusivity variables have a sufficient separation power to allow DVCS and Deep Exclusive π^0 Production identification in an efficient way.
  - A dataset with at least $\sim 90\%$ of DVCS events can be extracted.
- When the final proton information is ignored:
  - There is a strong contribution of DIS processes to the background.
  - We can recover more events, in comparison to the proton-detected case, and directly benefits the small t region.
  - We have smaller statistical error bars on the BSA measurements.

Outlook

COMING SOON



- LH$_2$ target:
  - Subtraction of leftover contamination when proton is detected.
  - Background estimation and subtraction when proton is not detected
  - Extract final BSA measurements.
- Deuterium target:
  - Extraction of nDVCS events.
- Transversely polarized NH$_3$ target
  - Analyze longitudinally polarized NH$_3$ target data.
  - Try several ML methods

Thanks

$ep \to e\gamma p$: Background subtraction

A look to the kinematics:

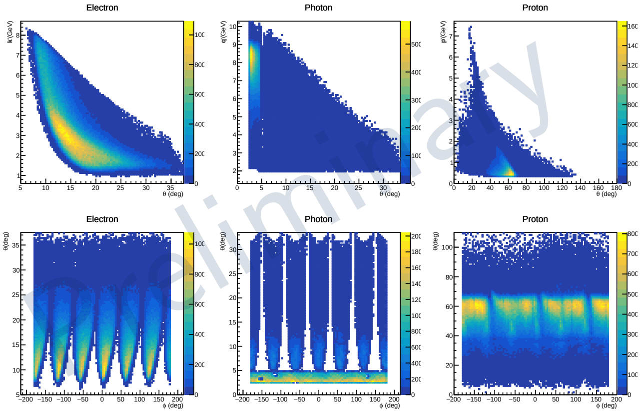


Figure: Momentum of the final particles as a function of the polar angle (first row) and detection polar vs azimuthal angle for each final state particle (second row).

$ep \to e\gamma(p)$: Background subtraction

Let's finally look at the kinematics of the extracted dataset

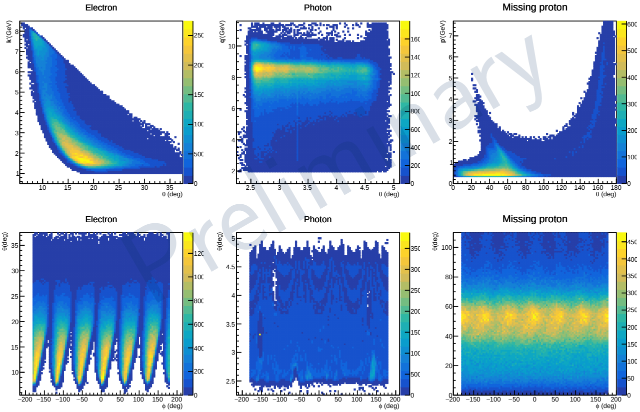


Figure: Momentum of the final particles as a function of the polar angle (first row) and detection polar vs azimuthal angle for each final state for $\theta_\gamma < 5°$

$ep \to e\gamma(p)$: Background subtraction

Let's finally look at the kinematics of the extracted dataset

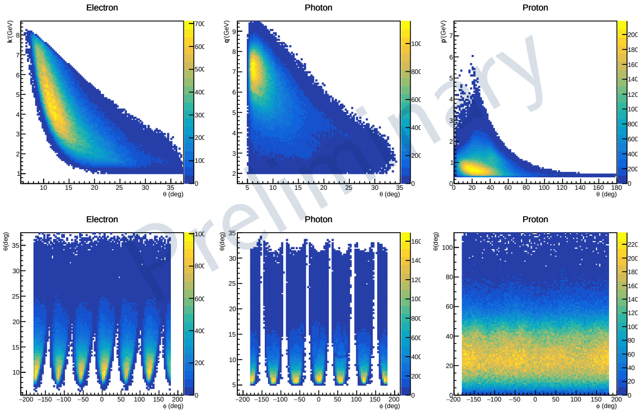
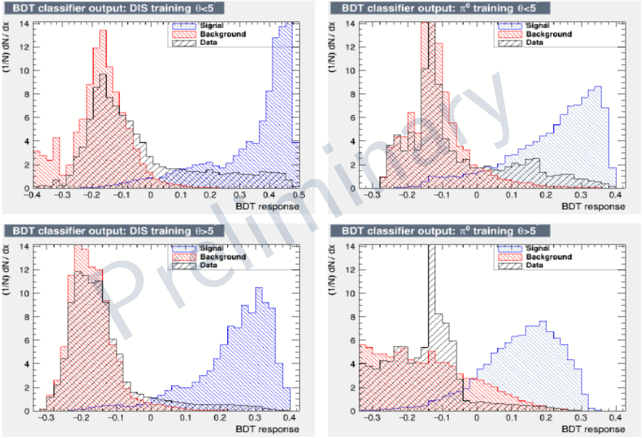


Figure: Momentum of the final particles as a function of the polar angle (first row) and detection polar vs azimuthal angle for each final state for $\theta_\gamma > 5^\circ$

$ep \to e\gamma(p)$: Background subtraction

The training of the BDT results in:



.

Figure: BDT output distributions for the DIS and π^0 background trainings. in the two θ_γ regions.

ROC reconstruction

Using a different sample, we can reconstruct the ROC curve

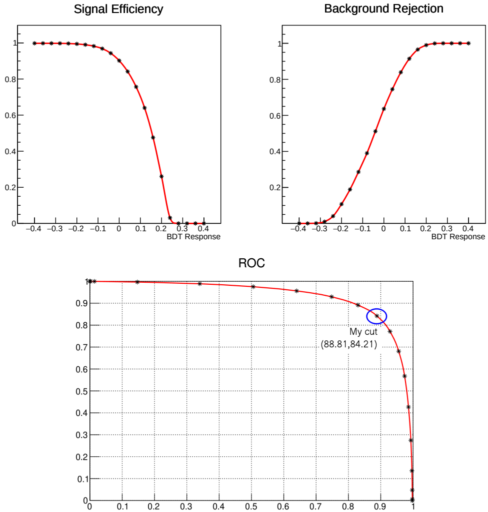


Figure: Reconstructed efficiencies as a function of the BDT response.

ROC reconstruction

From the position on the ROC curve and the number of events, we estimate that the original dataset was 54.58% DVCS and 44.16% π^0 production. So we can create an artificial dataset

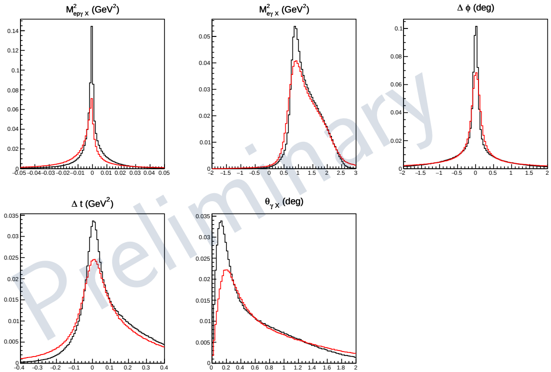


Figure: Normalized DVCS exclusivity variables from data(red) and simulated data (black) before BDT cut.

ROC reconstruction

We observe consistency in the results.

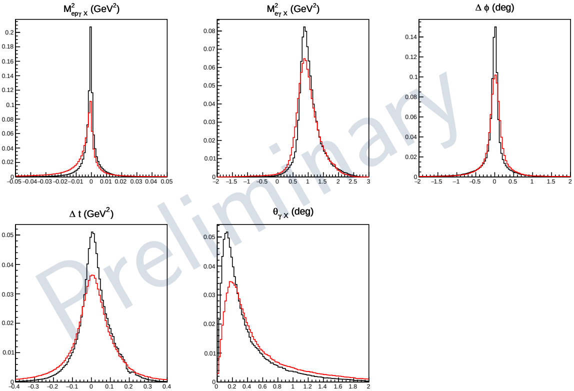
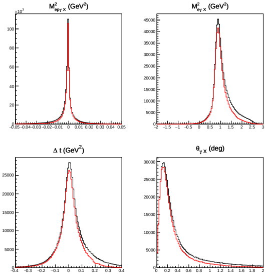


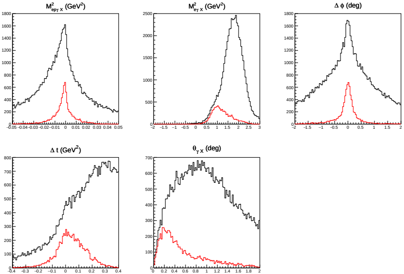
Figure: Normalized DVCS exclusivity variables from data(red) and simulated data (black) after BDT cut.

ROC reconstruction

Looking at each component:



(a) Simulated DVCS before (black) and after (red) BDT cut. signal efficiency: 89.03%, 88.81% expected.



(b) Simulated π^0 data, reconstructed as DVCS, before (black) and after (red) BDT cut.Background rejection: 84.78%, 84.21% expected.

$ep \to e\gamma(p)$: Data selection

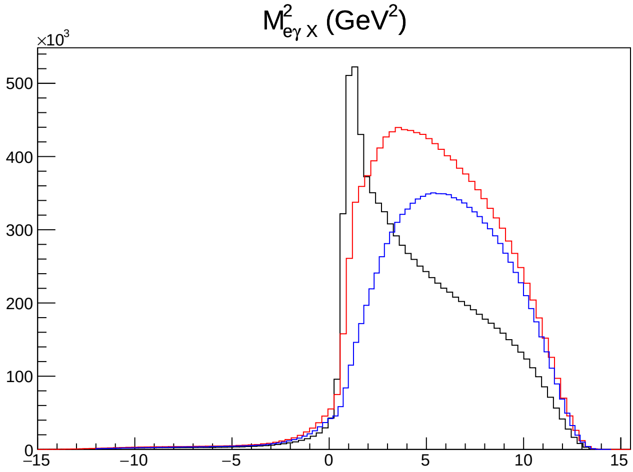


Figure: Missing mass from the process $ep \to e\gamma X$ for events with 1 (black), 2 (red) and 3 (blue) photon detections.

$ep \to e\gamma(p)$: Background subtraction

The training of the BDT results in:

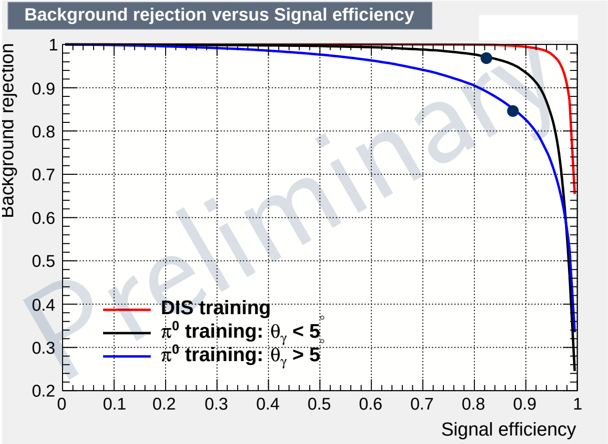
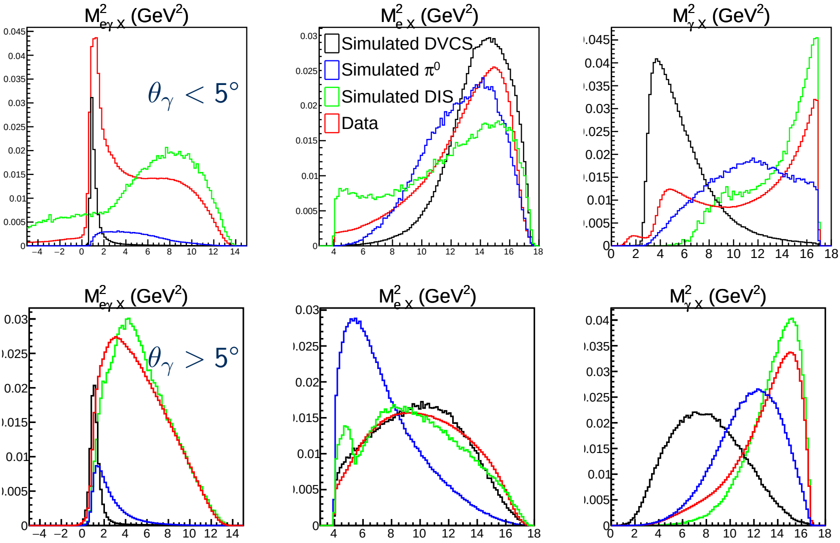


Figure: ROC curve of the model for the DIS and π^0 background trainings indicating the applied cut. DIS training gives similar results in both cases.

$ep \to e\gamma(p)$: Model training



$ep \to e\gamma(p)$: Background subtraction

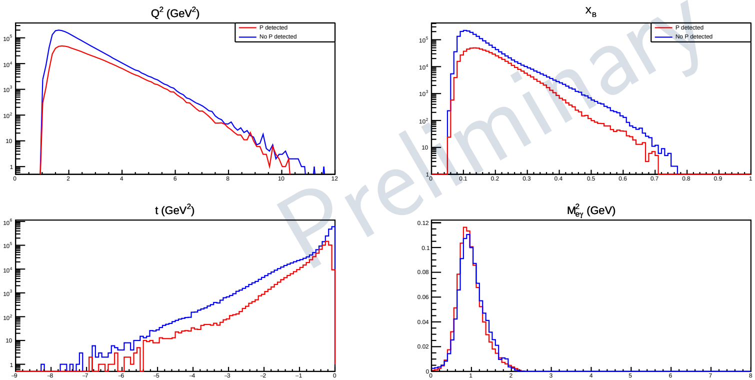


Figure: Q^2, x_B, t and $M^2_{e\gamma}$ distributions, after BDT cut, when the recoil proton is required (red) or not (blue) for $\theta_\gamma < 5°$.

$ep \to e\gamma(p)$: Background subtraction

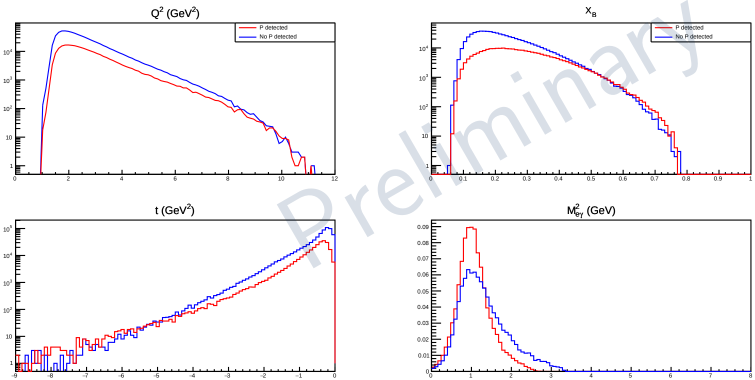


Figure: Q^2, x_B, t and $M^2_{e\gamma}$ distributions, after BDT cut, when the recoil proton is required (red) or not (blue) for $\theta_\gamma > 5°$.

The applied BDT cut leads to the following efficiencies on simulated data:

	$\theta_\gamma < 5°$	Remaining on data	$\theta_\gamma > 5°$	Remaining on data
DVCS	83.5%		86.93%	
π^0	3.64%	<10.3%	16.3%	<80%
DIS	0.044%	<1.2%	0.77%	<9.16%

- DVCS data can be extracted when photons are detected in the Forward Tagger
  - Remaining DIS contamination can be taken into the systematics.
- Remaining contamination is the upper bound coming from the hypothesis that 70% (80%) of the data in the FT (FD) is background data.
- From simulations, we estimate that events at $\theta_\gamma > 5$ contain 3 times more DIS events, 5 times more π^0 events and 3 times **less** DVCS events. (e.g. if 20% of the background is π^0 production, after BDT cut there will be 40% of π^0 contamination.)

Beam Spin Asymmetry

We construct bins of equal number of events before background substraction

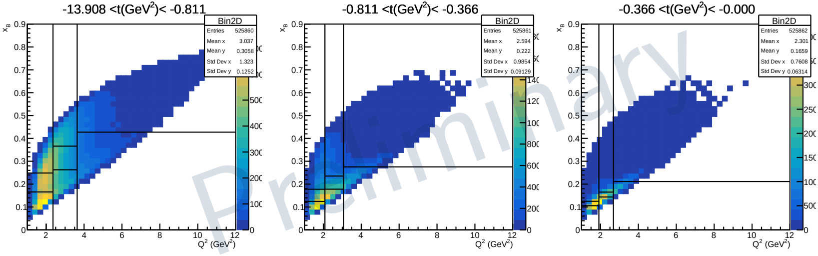


Figure: Binning scheme for BSA measurements in regions with $60K$ events.

Raw Beam Spin Asymmetry

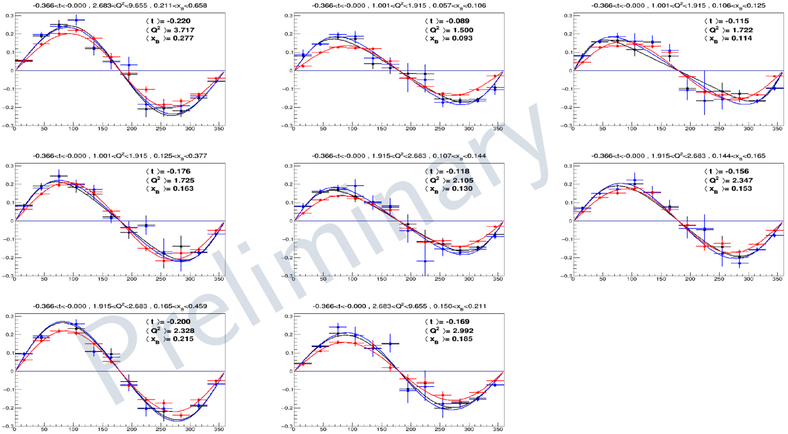
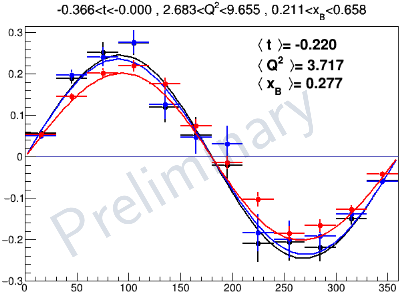


Figure: raw BSA measurements for global training with proton (blue), without proton (red) and **binned training with proton (black)**.

Raw Beam Spin Asymmetry

$$A_{LU} \equiv \frac{1}{P}\frac{h^+ - h^-}{h^+ + h^-} \sim \frac{p_0 \ \sin(\phi)}{1 + p_1 \ \cos(\phi)} \sim \frac{\sin(\phi)}{\sigma_{UU}} \Im\left[F_1\mathcal{H} + \xi(F_1 + F_2)\tilde{\mathcal{H}} - kF_2\mathcal{E} + ...\right]$$



	GP	BP	GNP
$\langle t \rangle$ (GeV)^2	-0.251	-0.247	-0.220
$\langle Q^2 \rangle$ (GeV)^2	3.753	3.759	3.717
$\langle x_B \rangle$	0.260	0.264	0.277

- Global training with proton (blue): GP
- **Binned training with proton (black)**: BP
- Global training without proton (red): GNP

- Training on bins gives very similar results to the global training.
- Training without proton information has an additional systematic shift.