Updates on the analysis of J/ψ photoproduction on the proton

Pierre Chatagnon

8th of February 2023

Jefferson Lab clas

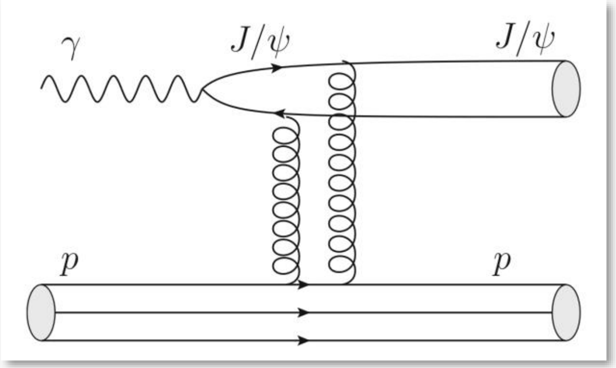
Outline

I – Motivations, general considerations and timeline

Motivations for dilepton final state measurement

J/ψ photoproduction at threshold

$$\gamma p \to J/\psi\, p \to e^+ e^- p'$$



- The t-dependence of the cross-section allow to access gluon Gravitational Form Factors (GFFs), mass radius of the nucleon and gluon GPDs (under 2-gluon exchange assumption and no open-charm contributions)
- Model-dependent limit on the branching ration of the Pc pentaquark.

Publications at JLab

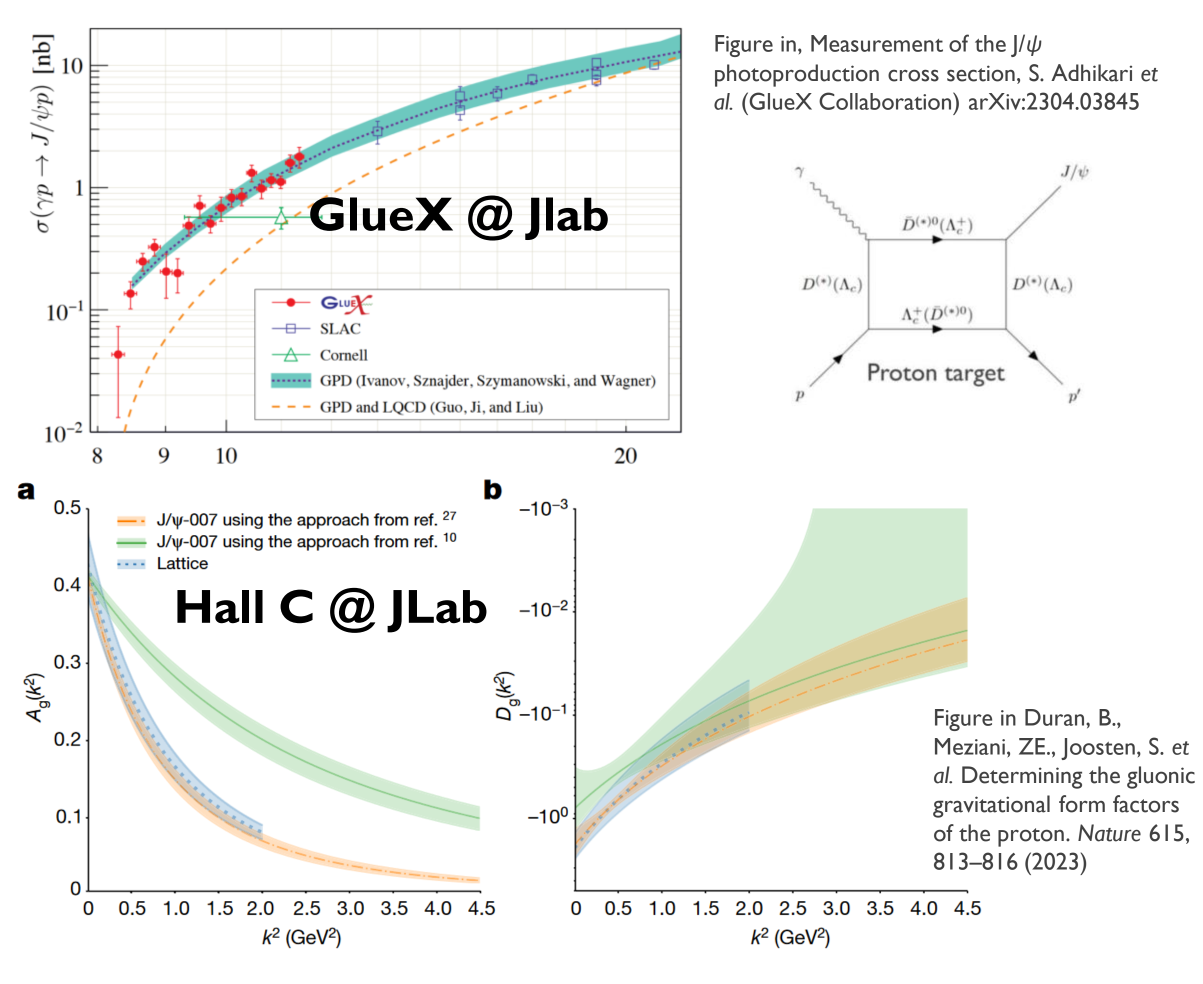


Figure in, Measurement of the J/ψ photoproduction cross section, S. Adhikari et al. (GlueX Collaboration) arXiv:2304.03845

Figure in Duran, B., Meziani, ZE., Joosten, S. et al. Determining the gluonic gravitational form factors of the proton. Nature 615, 813–816 (2023)

General analysis strategy

1) CLAS12 PID + Positron NN PID

$$ep \rightarrow (e')\gamma p \rightarrow (e')J/\psi\ p' \rightarrow (X)e^+e^-p'$$

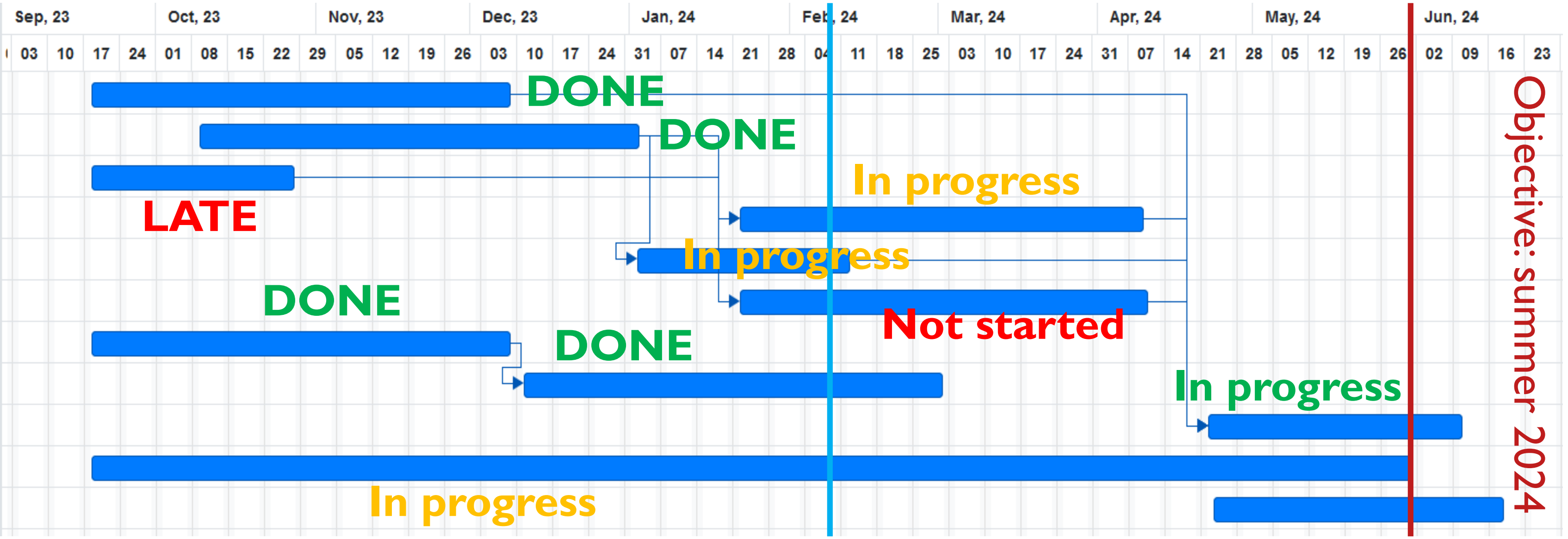
$p_X = p_{beam} + p_p - p_{e^+} - p_{e^+} - p_{p'}$ → 2) $|M_X^2| < 0.4 GeV^2$ → 3) $Q^2 < 0.5$ GeV^2

Event selection

- Event topology:
  - exactly one electron in FD
  - exactly one positron in FD
  - exactly one proton
  - anything else
- HTCC and ECAL hits in the same sector
- HTCC lepton time within 2ns
- Lepton momenta > 1.7 GeV
- Proton in the FD
- Sampling Fraction > 0.15
- Lepton AI PID score > 0.05 (trained on pass 2 simulation)
- Exclusivity cuts:
  - $|MM^2|<0.4$ GeV^2
  - $|Q^2|<0.5$ GeV^2

Timeline for the tools and task for a dilepton publication

ID	Name	Start Date	End Date
1	Lepton AI PID	Sep 18, 2023	Dec 08, 2023
2	Data processing	Oct 09, 2023	Jan 02, 2024
4	Energy loss corrections	Sep 18, 2023	Oct 27, 2023
5	Momemtum correction	Jan 22, 2024	Apr 09, 2024
9	Fiducial cuts	Jan 02, 2024	Feb 12, 2024
6	Momentum smearing	Jan 22, 2024	Apr 10, 2024
7	Radiative correction validations	Sep 18, 2023	Dec 08, 2023
8	Radiative corrections	Dec 11, 2023	Mar 01, 2024
10	Systematics	Apr 22, 2024	Jun 10, 2024
11	Analysis note writting	Sep 18, 2023	May 31, 2024
12	Article writting	Apr 23, 2024	Jun 18, 2024



- On time for PID, Data processing and radiative corrections
- Still some tools required/preferred for the analysis (Momentum corrections/smearing)
- Still on track for analysis note submission by the summer

II – Data, Monte Carlo samples and analysis tools

Data samples

- Analysis on Pass 2 data only
- *jpsitcs* train is used
- All *main* Fall 18 (Inbending and outbending) and Spring 19 runs are processed.
- The supplemental runs of Fall 2018 (run number <5000) are not used.

Integrated luminosity calculation

- The QADB tool is used (https://github.com/JeffersonLab/clas12-qadb) to retrieve the accumulated charge per DST files
- The tag *OkForAsymmetry* is used, with additional post-checks (remove low current runs, etc)
- The RCDB interface of clas12root is used (https://github.com/JeffersonLab/clas12root/blob/a7ba949fc92ed355e47f993f8342b0acb4b9303d/RunRoot/Ex8_RcdbReader.C) to query the RCDB for each run and retrieve *Beam current* and *Requested beam current*
- Accumulated charge is computed per beam current for each configuration (In/Out Fall 18, Spring 19)

Config.	Beam currents			Total
Fall 18 Inbending	45 nA 26.312 mC	50 nA 4.000 mC	55 nA 5.355 mC	35.667 mC
Fall 18 Outbending	40 nA 11.831 mC	50 nA 20.620 mC		32.451 mC
Spring 19	50 nA 45.994 mC			45.994 mC

MC samples

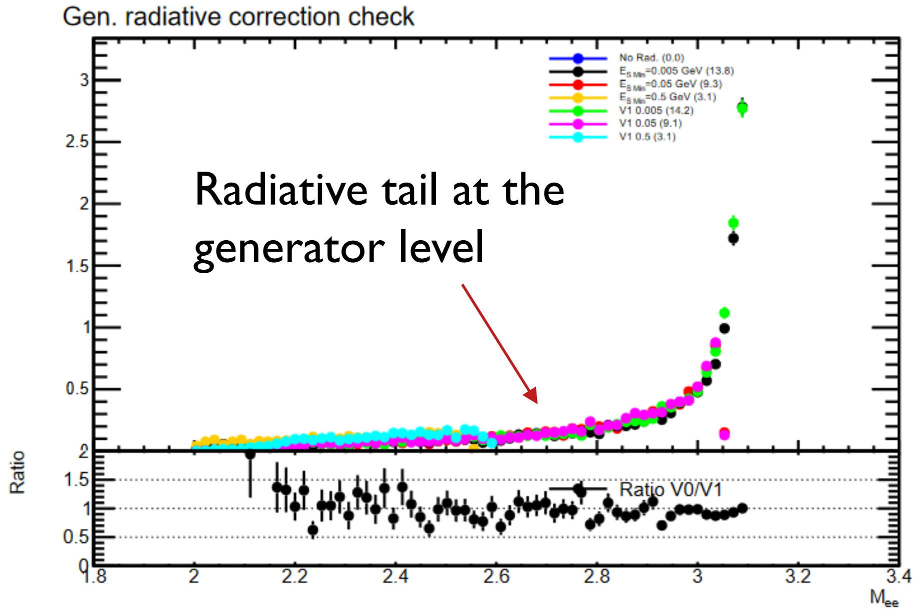
- 3 Generators are used: TCSGen (real photon BH), JPsiGen (Jpsi signal), Grape (Virtual photon BH)
- Simulations are processed through OSG with pass 2 configuration

Generator	Config / Beam currents					
	Fall 18 In.			Fall 18 Out.		Sp. 19
	45 nA	50 nA	55 nA	40 nA	50 nA	50 nA
Grape	8.2M each					6.7 M
TCSGen	2M each					1.5 M
JPsiGen	2M each					
JPsiGen (No rad.)	3M each					
Total of 24 MC samples						

Radiative corrections

- Inclusion of radiative effect is done in all generators according to formulas in: *Matthias Heller, Oleksandr Tomalak, and Marc Vanderhaeghen. Soft-photon corrections to the bethe-heitler process in the γp → l+l−p reaction. Phys. Rev. D, 97:076012, Apr 2018.*
- The JpsiGen and TCSGen generator with radiative effect are on Github
  - https://github.com/JeffersonLab/JPsiGen/tree/Rad_Corr
  - https://github.com/JeffersonLab/TCSGen/tree/Rad_Corr
  - https://github.com/PChatagnon/Grape_Rad_Corr

  …not yet on OSG
- A full note on the algorithm is ready and will be included in the analysis note.
- The work was presented at the CLAS collaboration meeting in July 23: https://indico.jlab.org/event/724/contributions/13105/attachments/10013/14848/RadiativeCorrections_for_BH.pdf



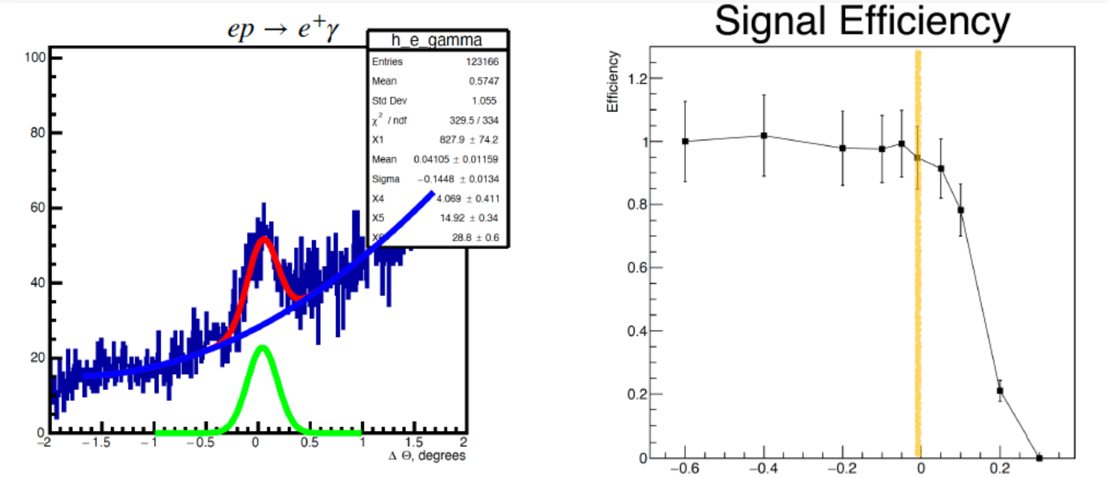
Lepton PID using AI

- Multiple evidences for large contamination from pions in the positron sample at high momenta (P>4.5 GeV)
- We developed a PID algorithm to use on top of the EB PID for leptons (electron, positron, muon(soon))
- Multivariate classifier using calorimeter responses only
  - Extension to Pass2 to the work that was done for the Pass1 TCS analysis

- One classifier per configuration and lepton flavor (6 in total)
- Soon available through Iguana
- Trained on simulation and validated on data

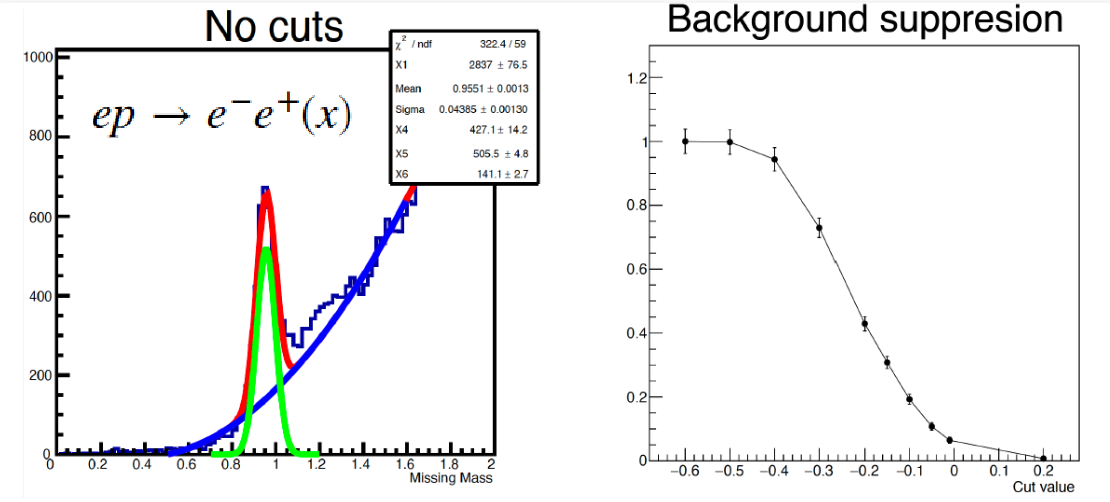
Work by Mariana Tenorio

Signal efficiency validation on data



All plots from Mariana Tenorio

Background rejection validation on data



Other tools

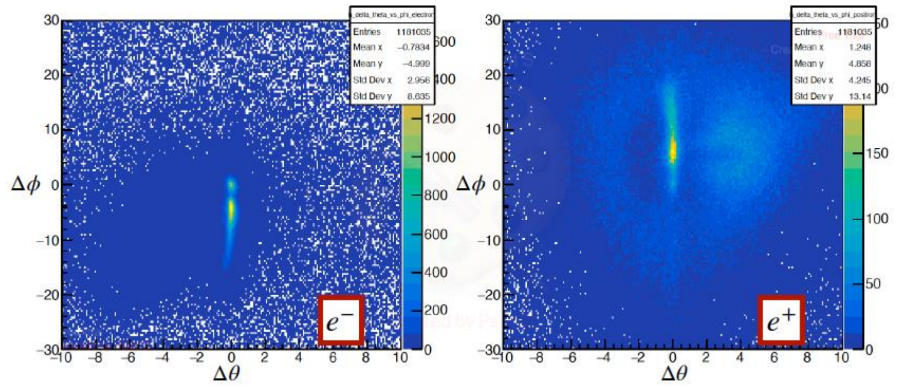
Fiducial cuts/dead paddle cuts

- Pass1 fiducial cuts on the PCAL (~ 8-9cm on V and W)
- Additional dead paddle cut, cross-check with Valerii Klimenko

Radiated photon correction

- Loop over photons in the event
- Add 4-vectors to the lepton if $\Delta\theta<1.5$ deg.

Plots from Mariana Tenorio



Github repository

- Github repository getting populated, some scripts still missing

https://github.com/PChatagnon/TCS_Analysis

Missing tools

- Not included in the following:
  - Energy loss / Energy corrections
  - Momentum smearing
  - *Edge*-based fiducial cuts

IV – Background modelisation and normalization

Jefferson Lab

Data/MC normalization

- Each event is weighted by:

$$\omega = \frac{\mathcal{L}\cdot\sigma_{tot}}{nb_{GEN}}$$ for generator providing integrated CS, $$\omega = \frac{\mathcal{L}\cdot w_{GEN}}{nb_{GEN}}$$ for weighted generator.

- Where the luminosity is obtained from target specification:

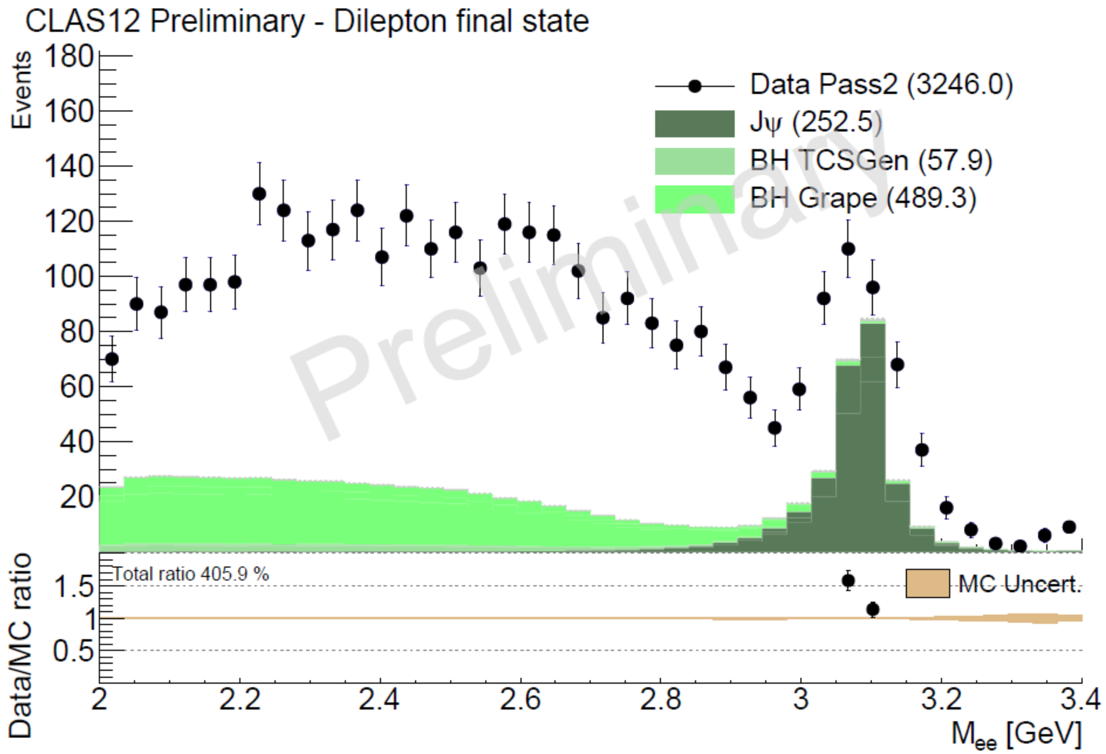
$$\mathcal{L} = \frac{l\cdot\rho\cdot N_A\cdot C\cdot Q}{e} = 1316.875 \cdot Q(\text{in mC})$$

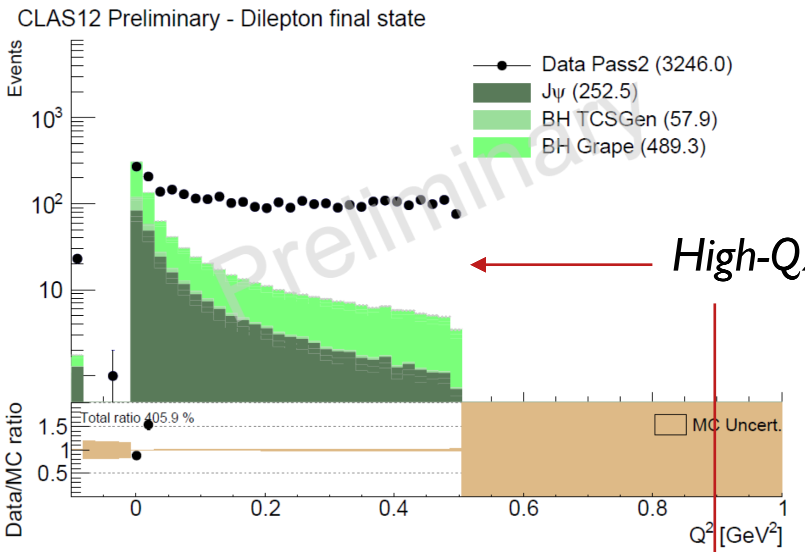
Length of the target $l = 5$ cm
Density of the target $\rho = 0.07$ g/cm^3
Avogadro constant $N_A = 6.02\text{x}10^{23}$ mol^{-1}
Unit charge $e = 1.6\text{x}10^{-19}$ C
Conversion to pb $C = 10^{-36}$

https://clasweb.jlab.org/rungroups/tlc/wiki/images/e/e7/Normalization_MC_Data-5.pdf

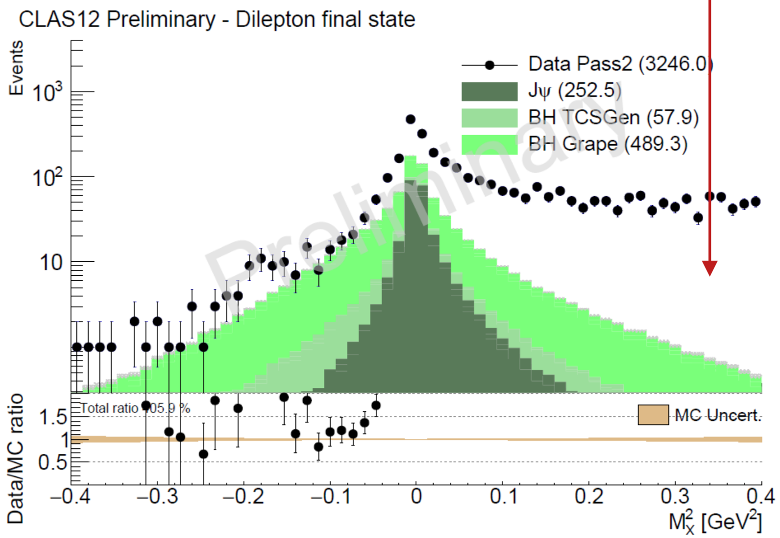
Comparison data/MC – Fall 2018 inbending

- Plotting conventions
  - Color-filled histograms are *stacked*, ie they show the total number of events with contributions for different channels "on top of each other"
  - Marker histograms are *not stacked* and simply superimposed

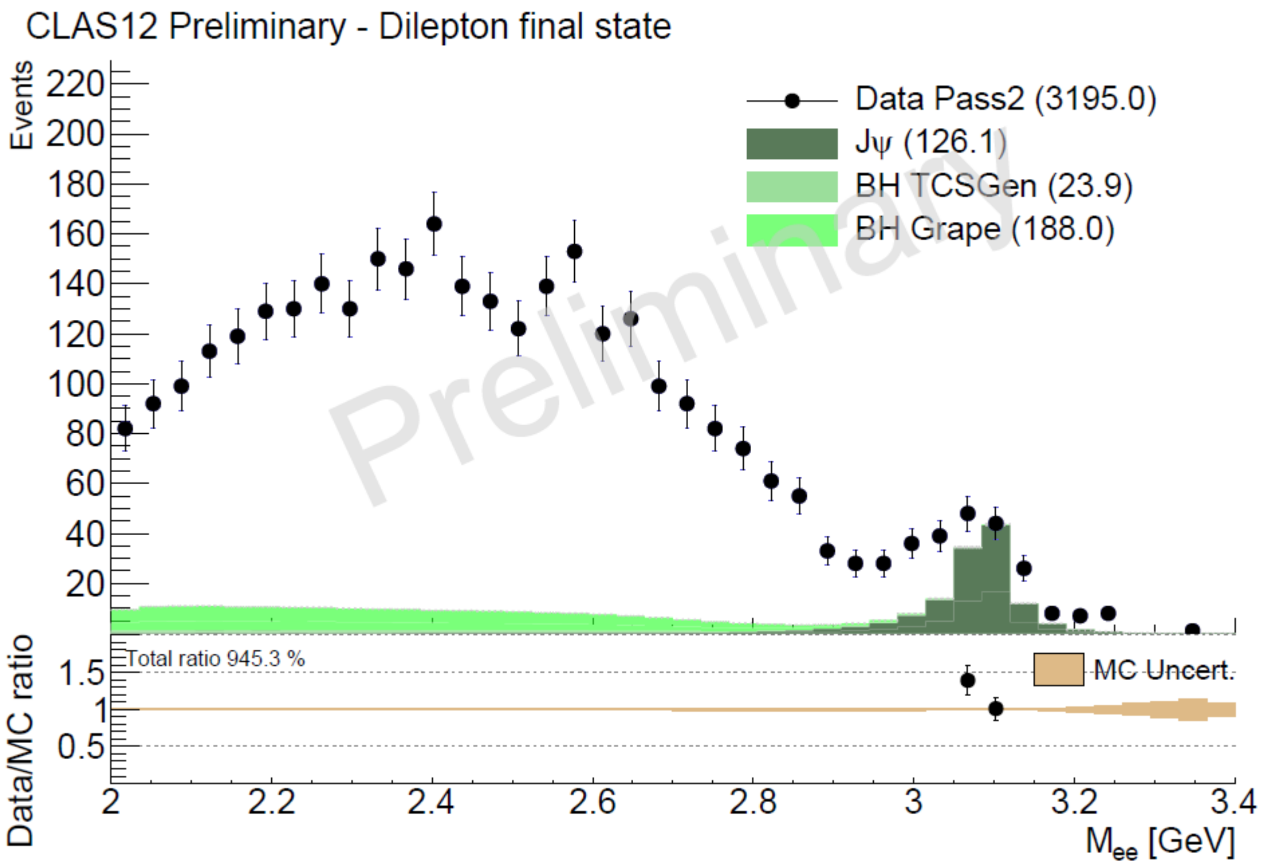


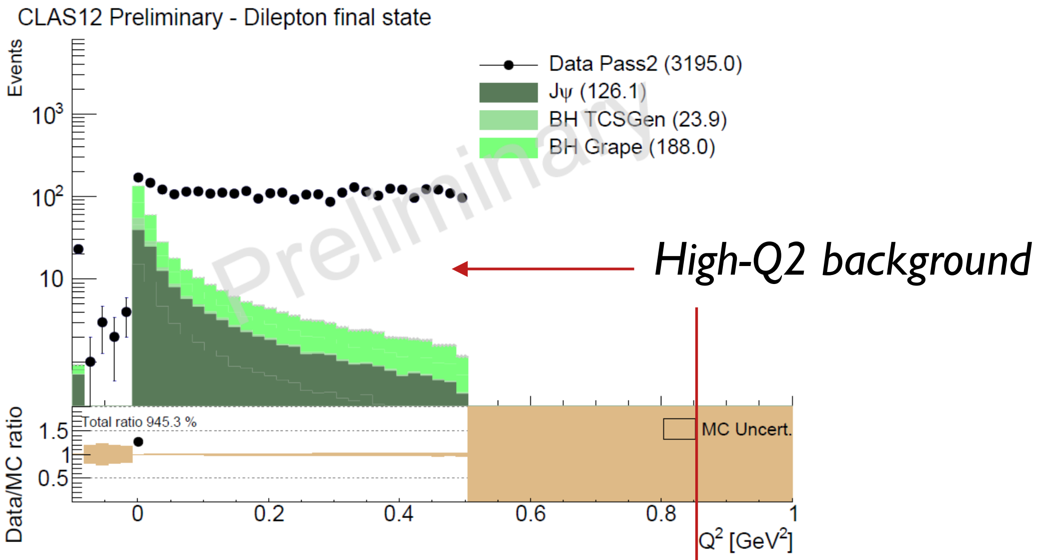


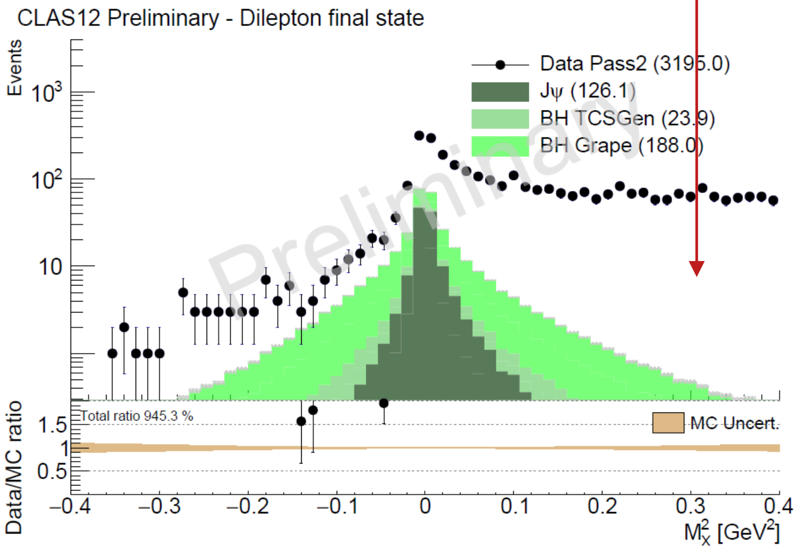
High-Q2 background



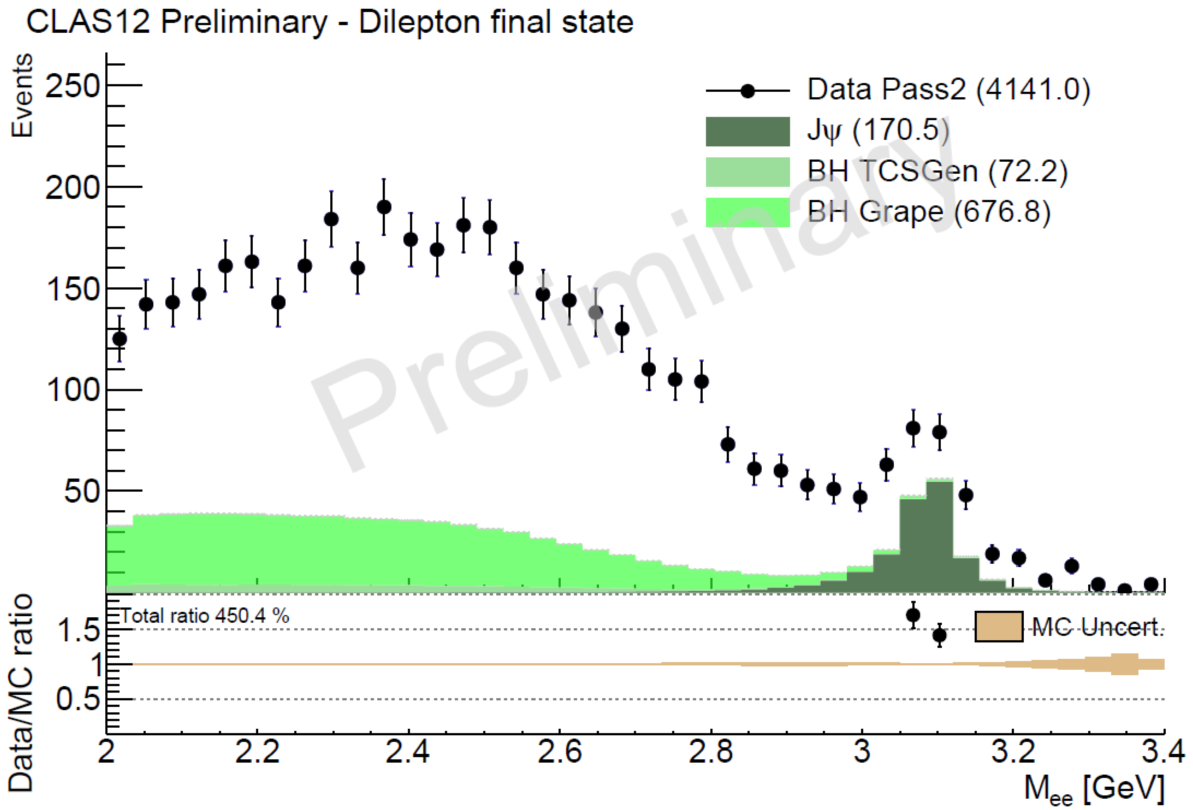
Comparison data/MC – Fall 2018 outbending

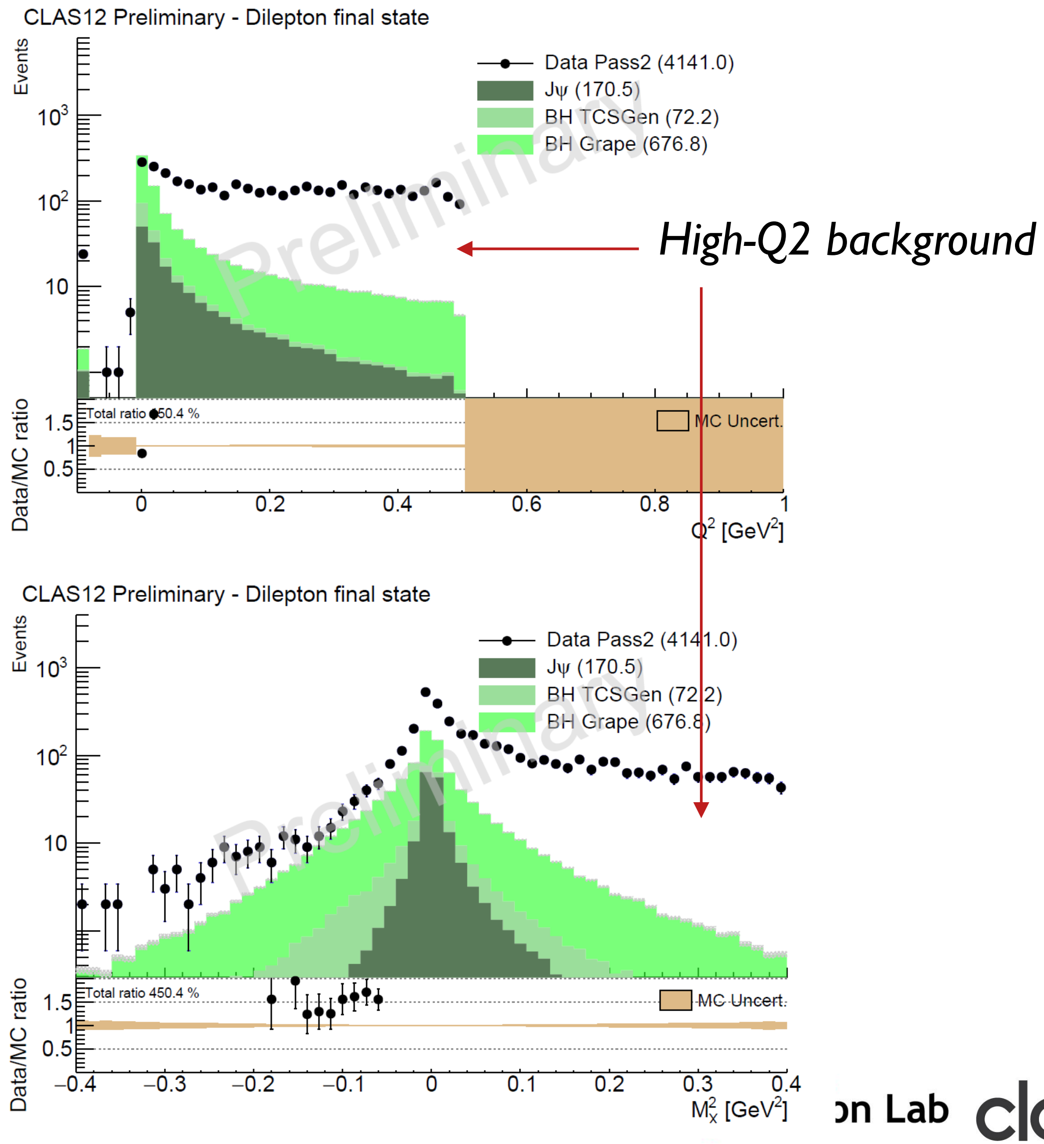






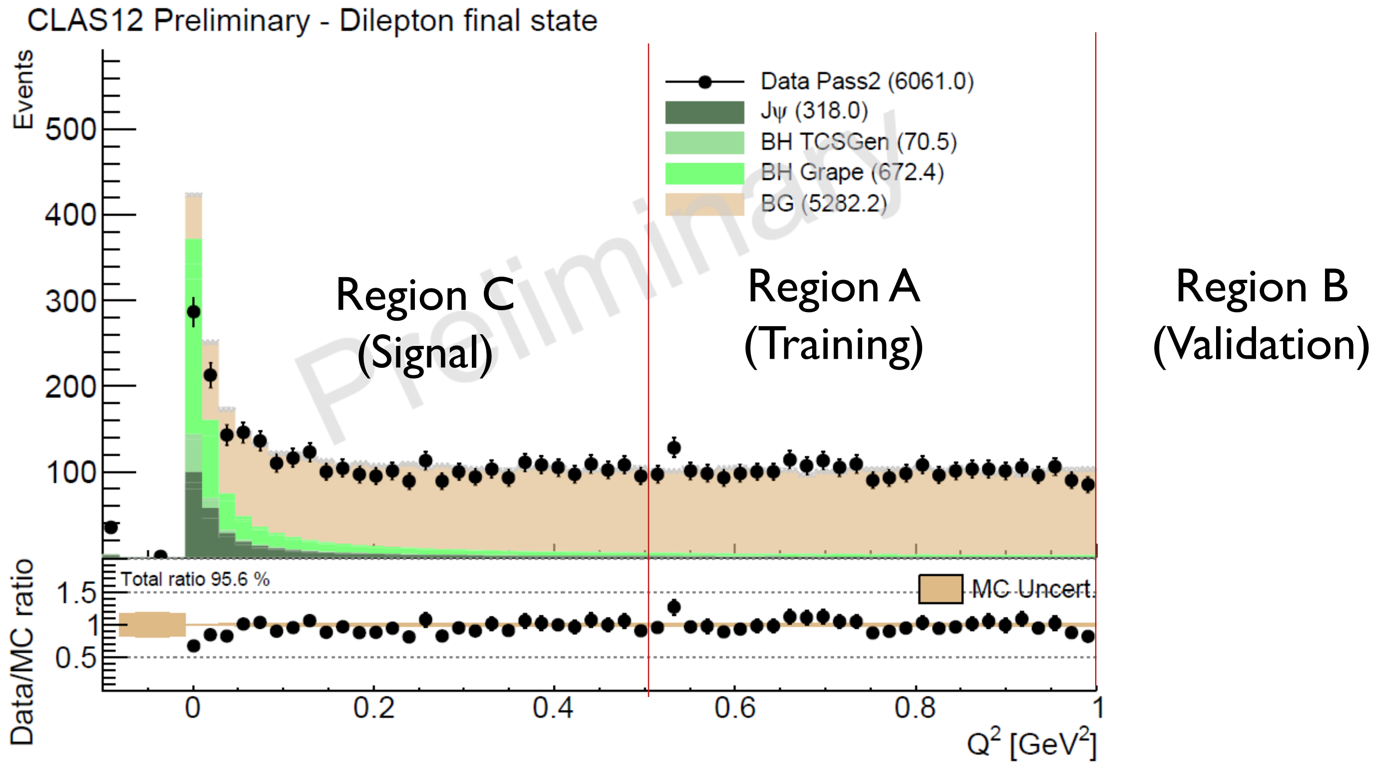
Comparison data/MC – Spring 2019





Overall strategy for background modelization

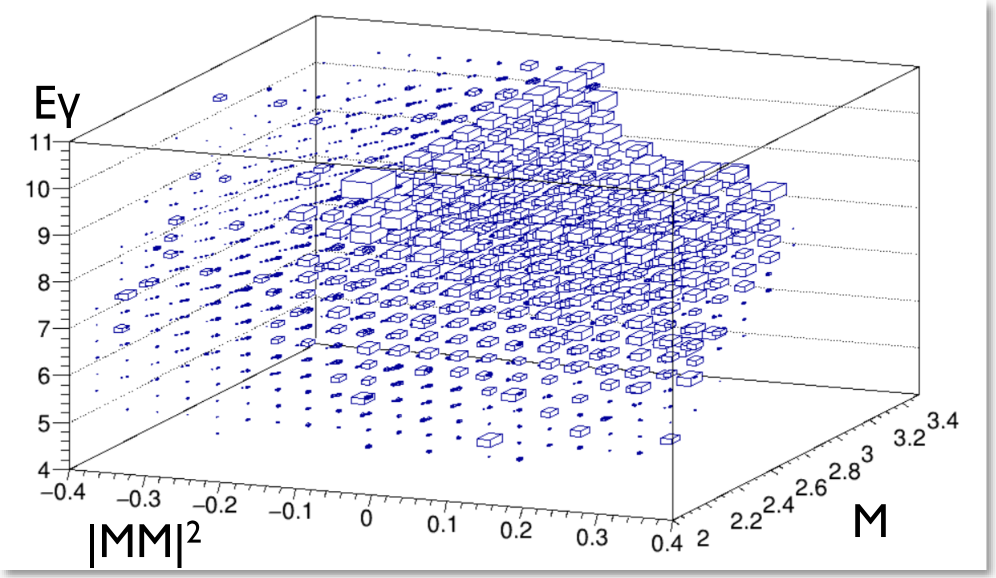
1) Event mixing
   - From data randomly select electron, positron, proton (from different events)
   - Construct kinematics and make sure they are within the region of interest (M_{ee}>2 GeV, $|MM|^2$<0.4 GeV^2, Q^2<2 GeV^2)
2) Reweight events to match data in the training region
3) Validate the weights on region B
4) Apply weights on region C and obtained BG-subtracted yields



Reweighting methods

Binned weights

- Compute ratio (Data-Signal)/(Mixed_BG), and apply to event from the mixed BG sample
  - Inconvenient method
    1) Need to track bin indices
    2) Curse of dimensionality: the more variable, the less events per bins



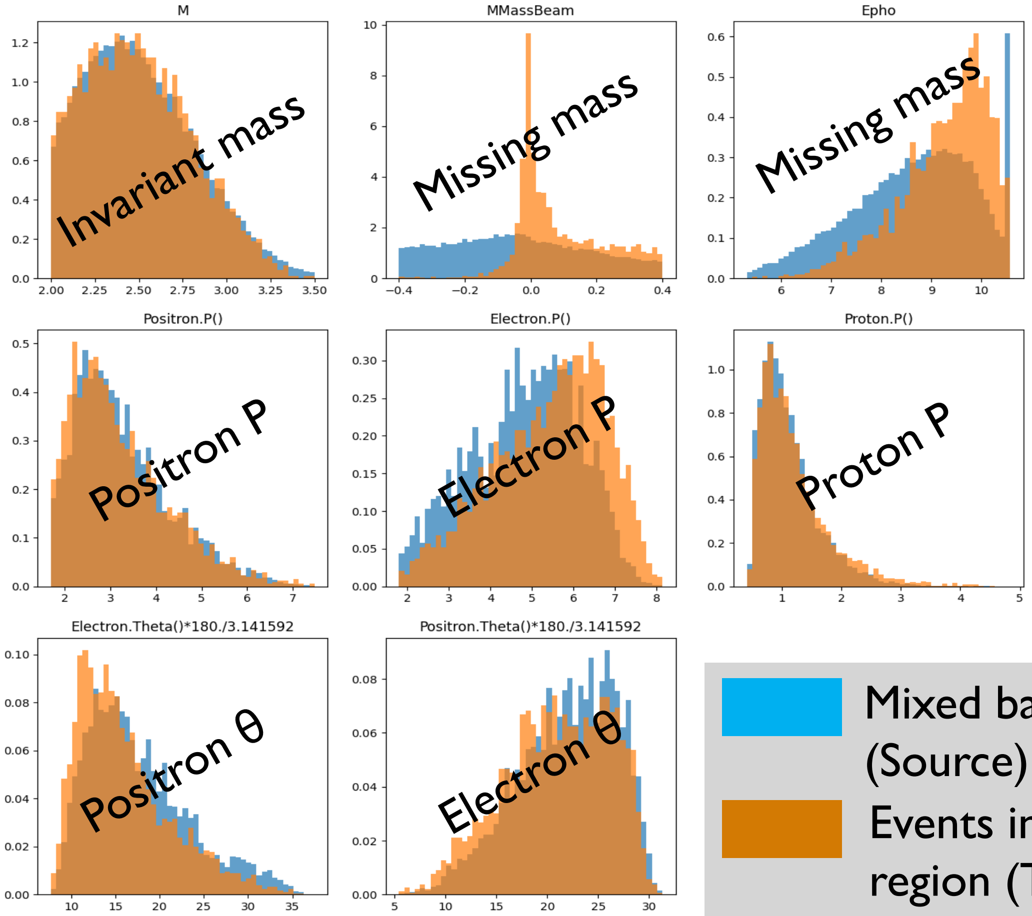
Boosted decision trees

- Use a BDT to compute a weight event-by-event so that source and target distribution match
- Using method from Alex Rogozhnikov 2016 *J. Phys.: Conf. Ser.* **762** 012036
- Code available here: (https://github.com/arogozhnikov/hep_ml)
  - Advantages
    1) As many variables as needed can be matched
    2) No/less of a dimensionality curse
    3) Easy to use, no need to handle complex bin indexing

Reweighting method using BDT

Using BDT reweighting from *Alex Rogozhnikov 2016 J. Phys.: Conf. Ser.* ***762*** *012036* (https://github.com/arogozhnikov/hep_ml)

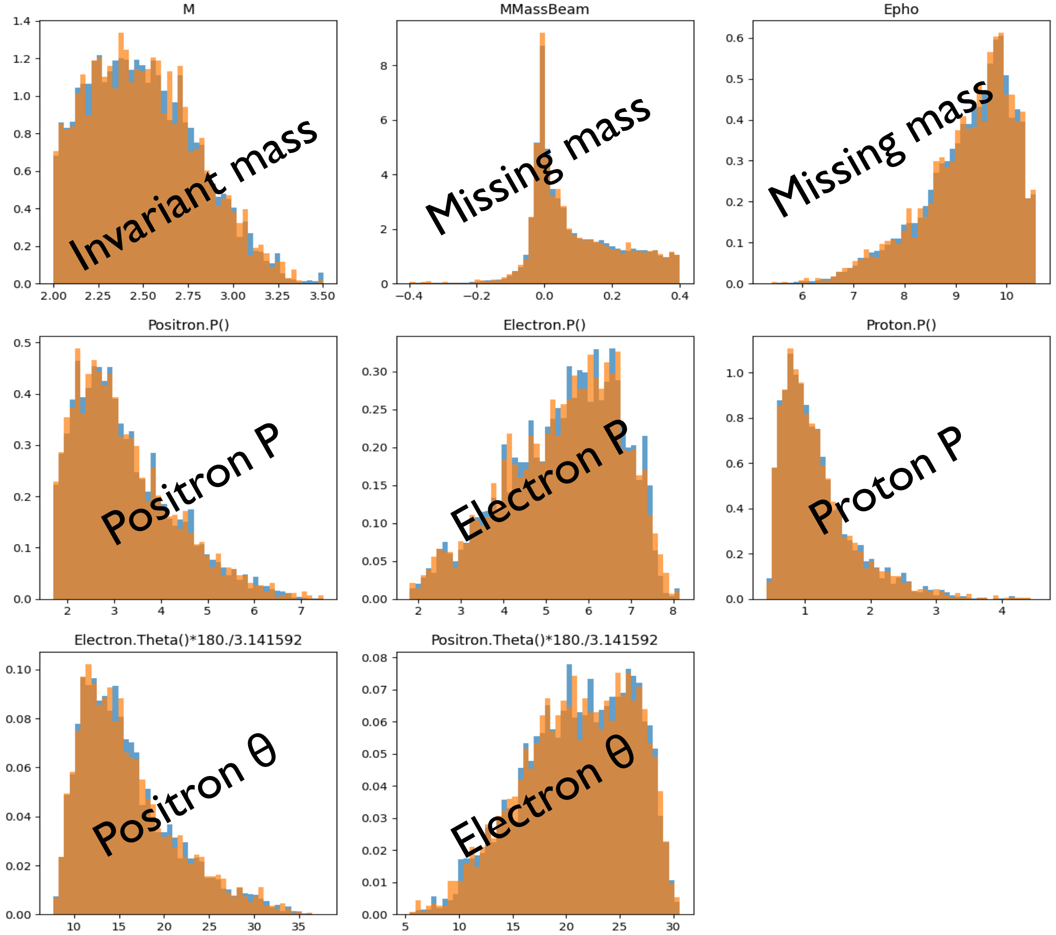
Before reweighting



Mixed background (Source)

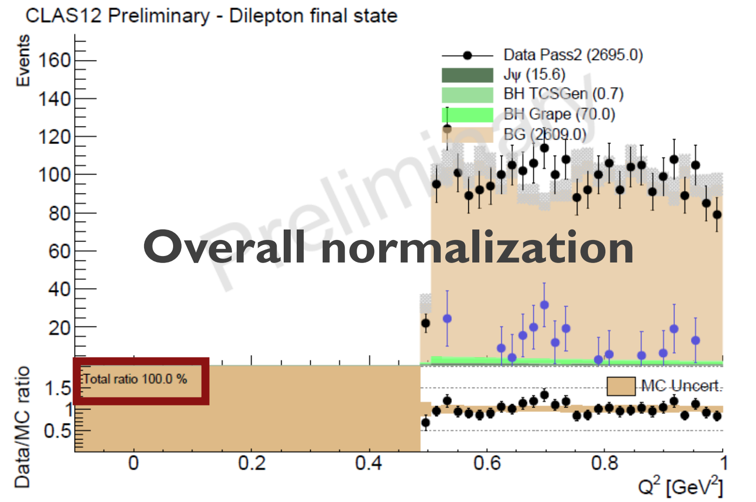
Events in the training region (Target)

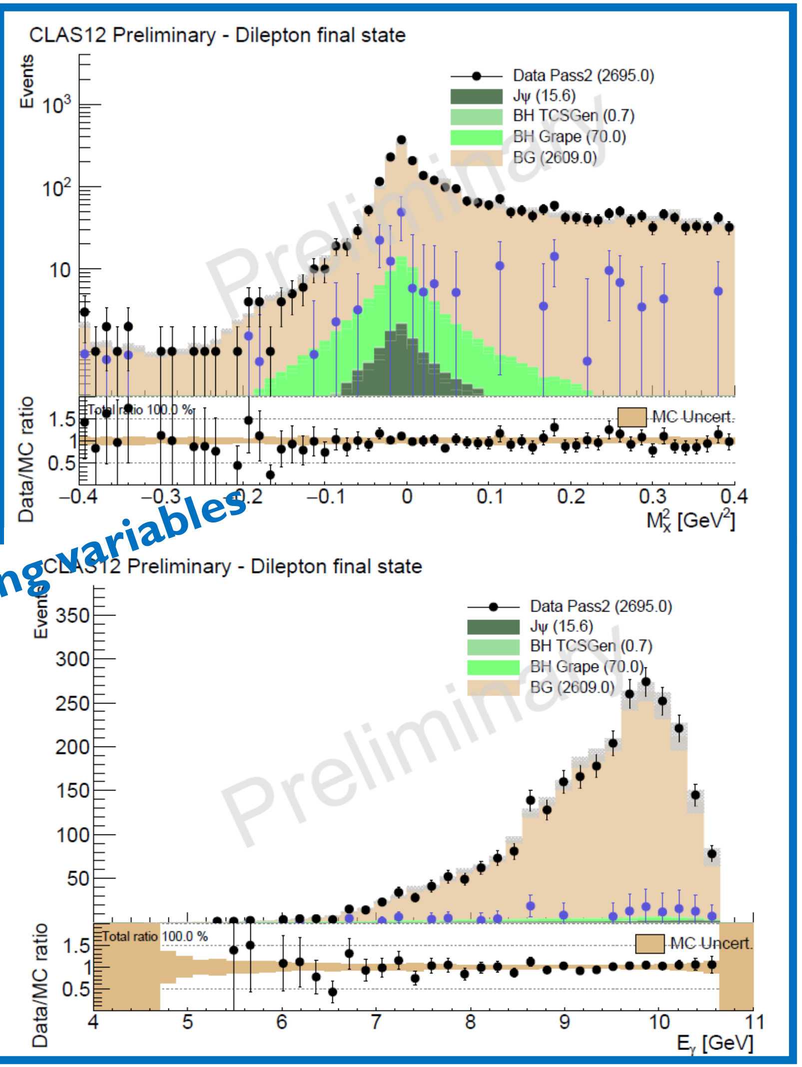
After reweighting

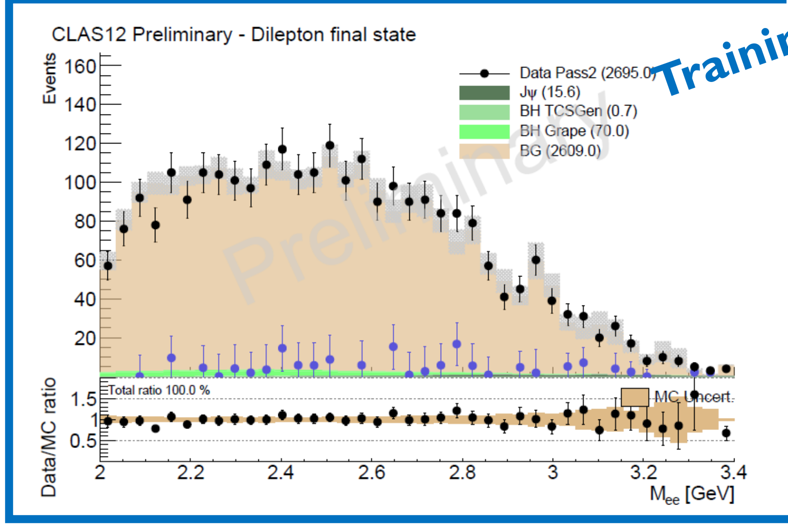


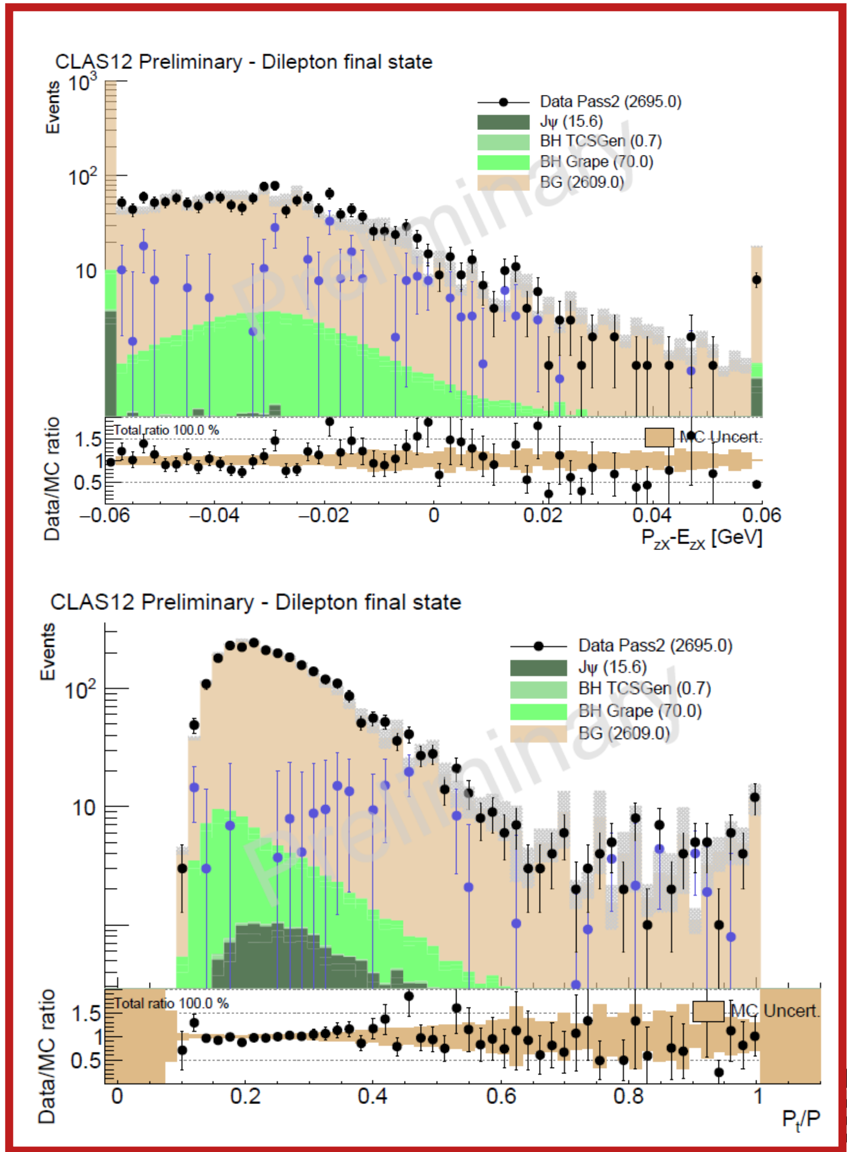
Reweighting validation – Fall 2018 inbending (I)

$Q^2 \in [0.5, 1.0] GeV^2$ Region A (Training)



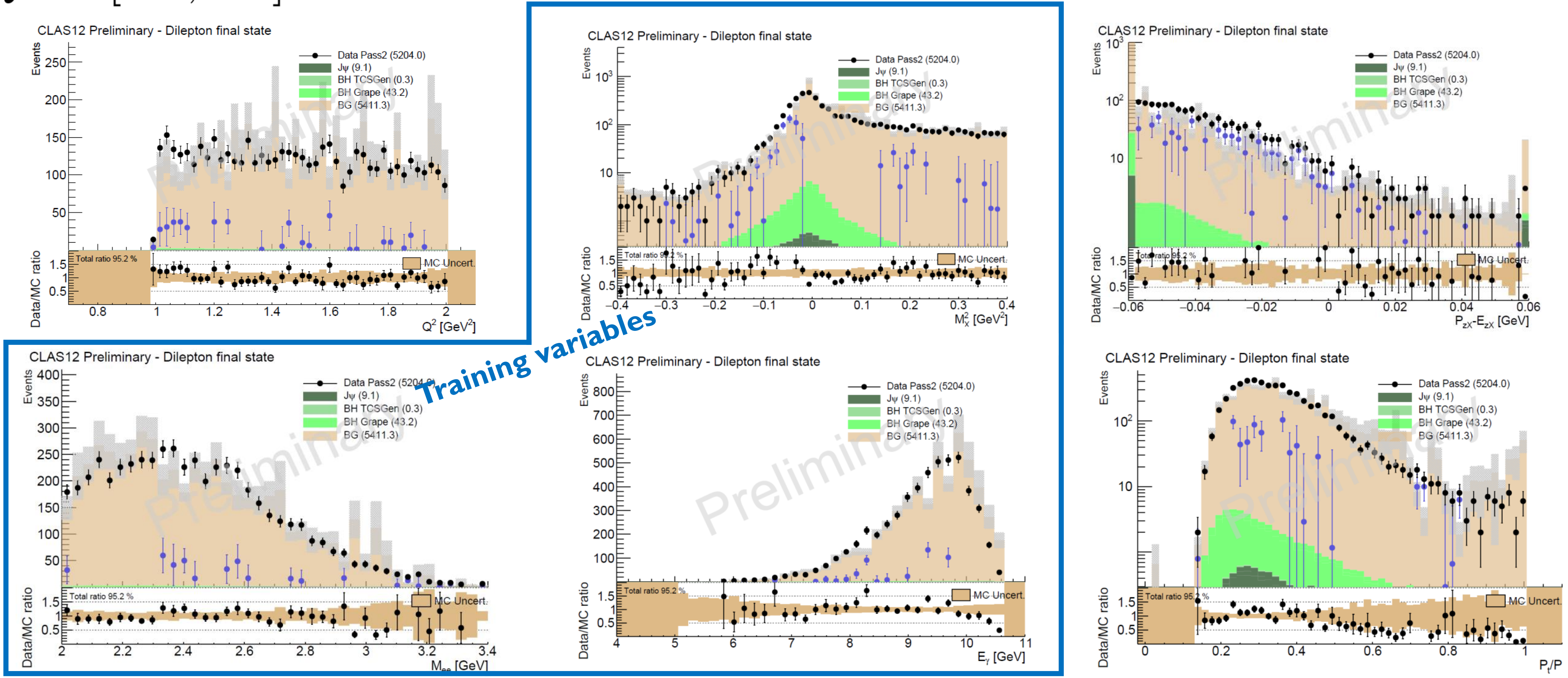






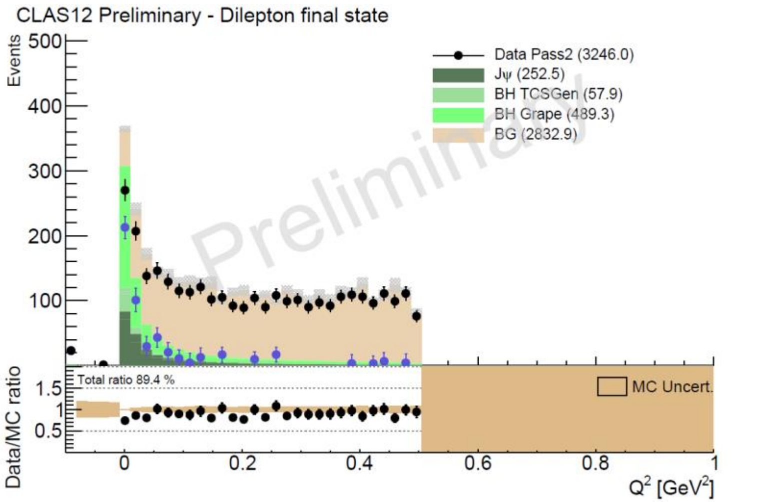
Reweighting validation – Fall 2018 inbending (II)

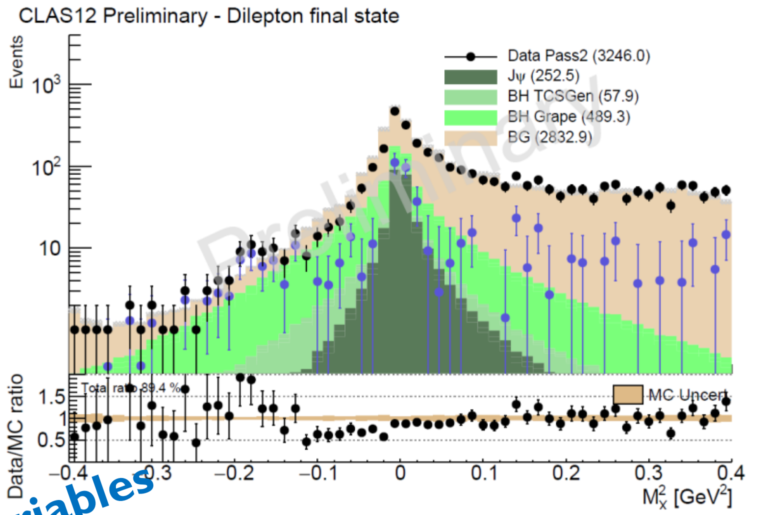
$Q^2 \in [1.0, 2.0] GeV^2$ Region B (Validation)

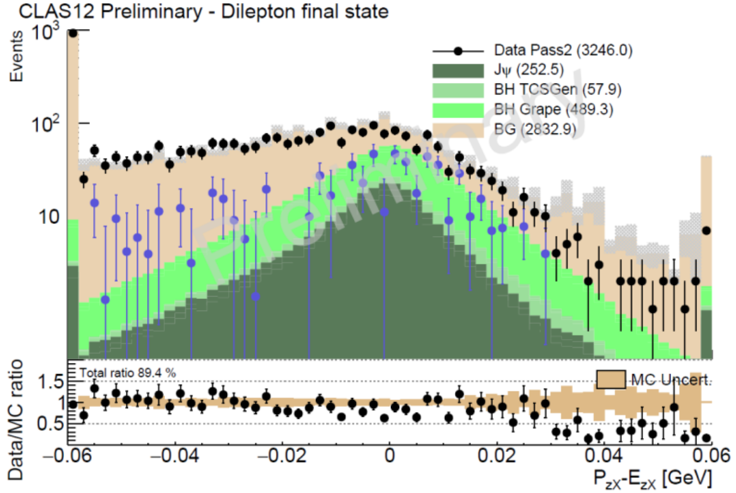


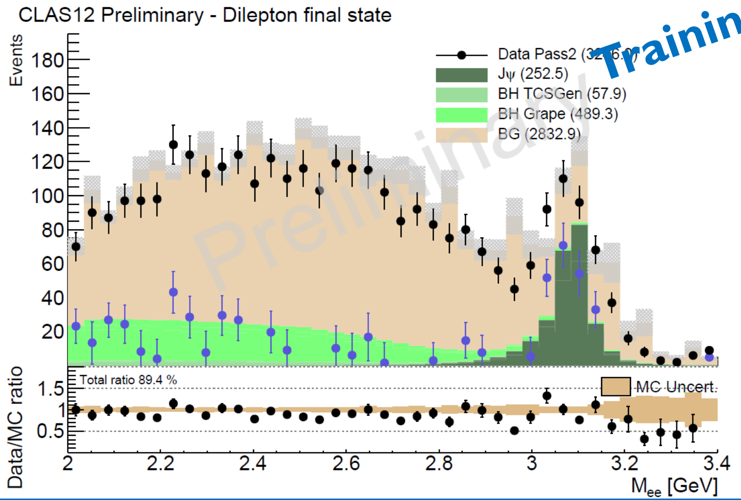
Full comparison data/MC – Fall 2018 inbending (1)

$Q^2 \in [0.0, 0.5] GeV^2$ Region C (Signal)

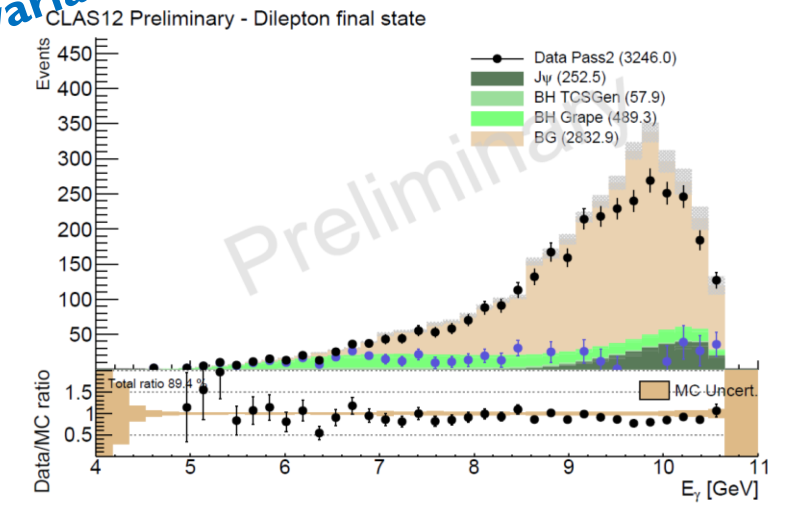


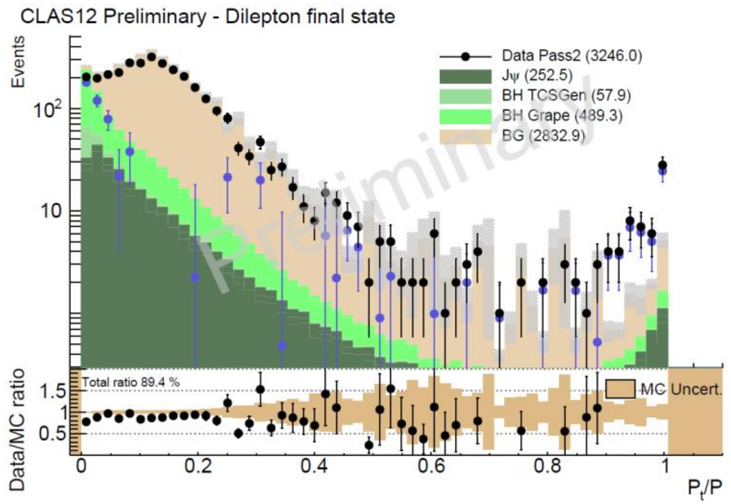






Training variables

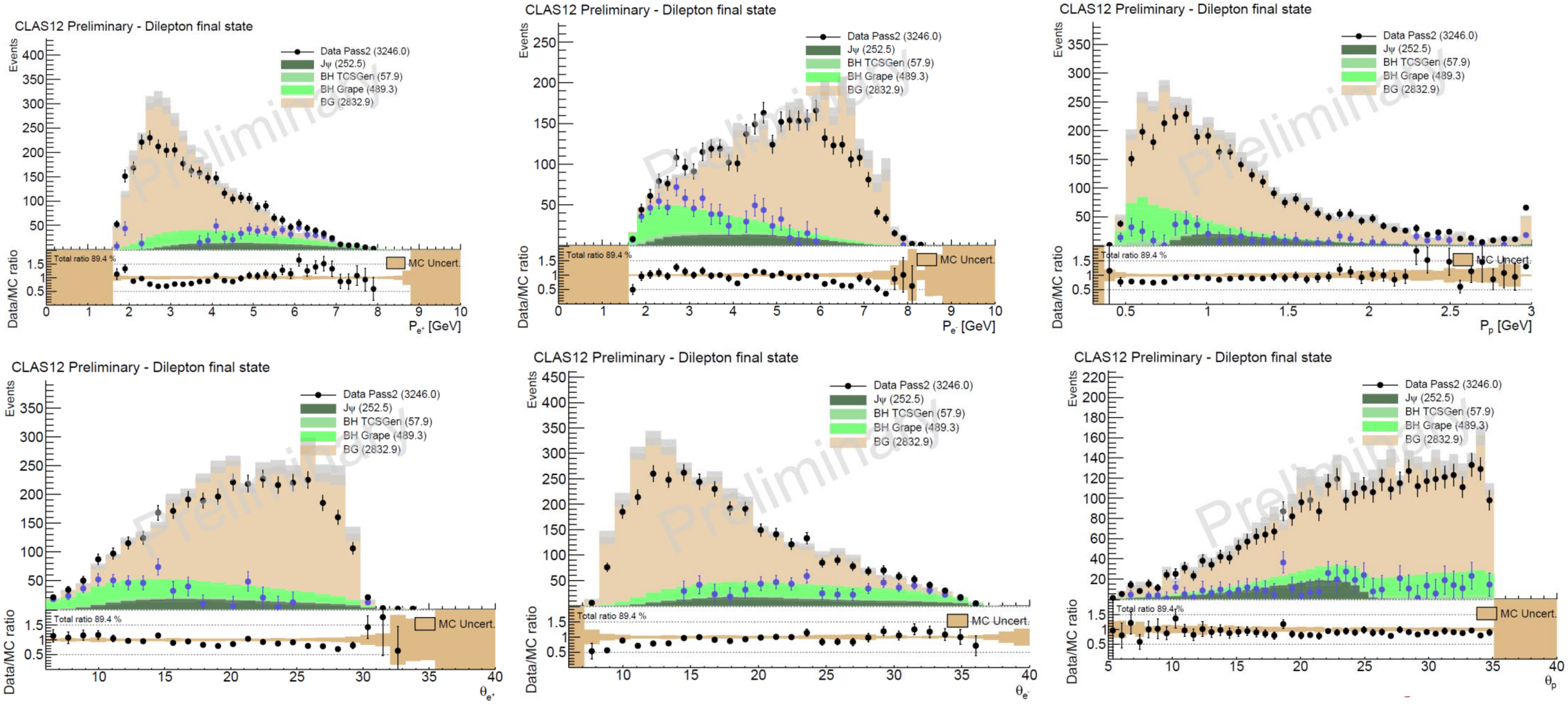




Full comparison data/MC – Fall 2018 inbending (2)

$Q^2 \in [0.0, 0.5] GeV^2$ Region C (Signal)

Final state particle kinematics

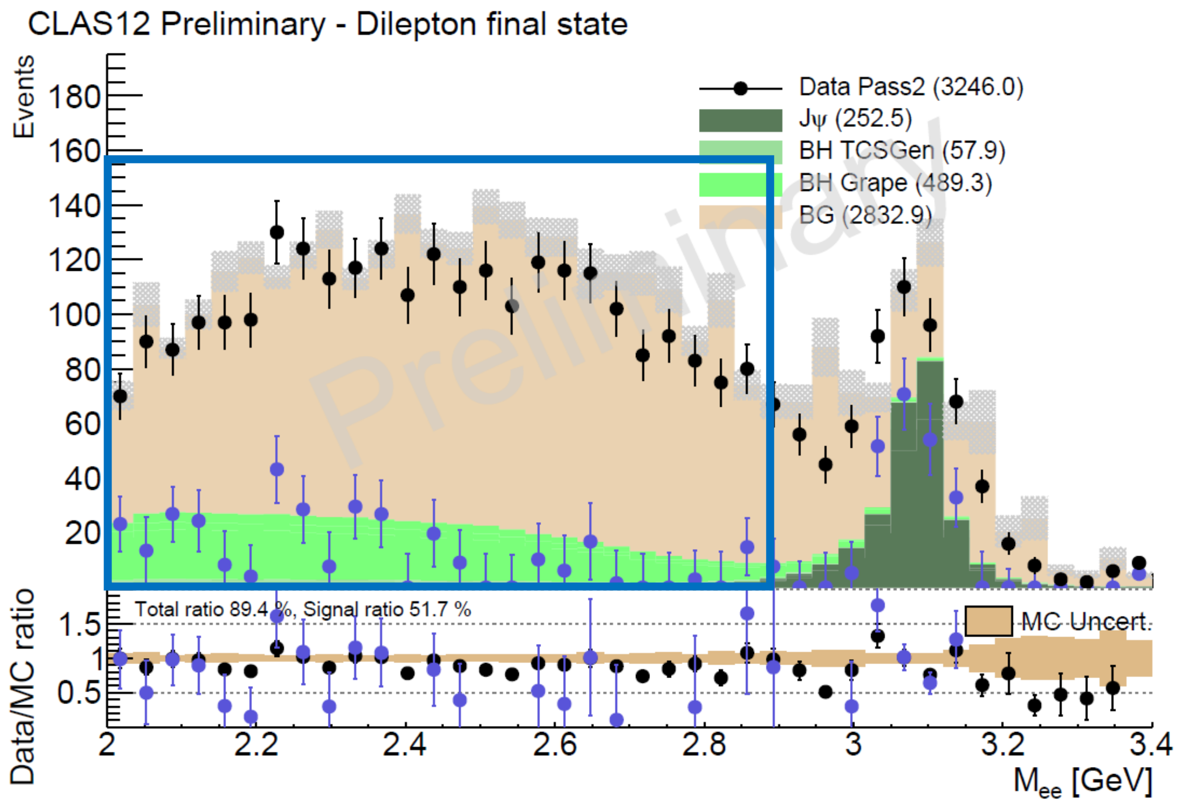


Normalization factor

- Normalization factor can be computed as:

$$\omega_c = \frac{N_{Data} - N_{BG}}{N_{SIM\ BH}}$$

- Here this would be 49.6% for inbending 2018 with large statistical error
- Requires larger BG samples
- To be continued



V - Results

Cross-section computation

$$\sigma_j = \frac{N_{J/\psi j}}{\mathcal{F}_j \cdot \mathcal{L} \cdot \omega_{cj} \cdot B_r \cdot \epsilon_j \cdot \epsilon_{Rad/j}}$$

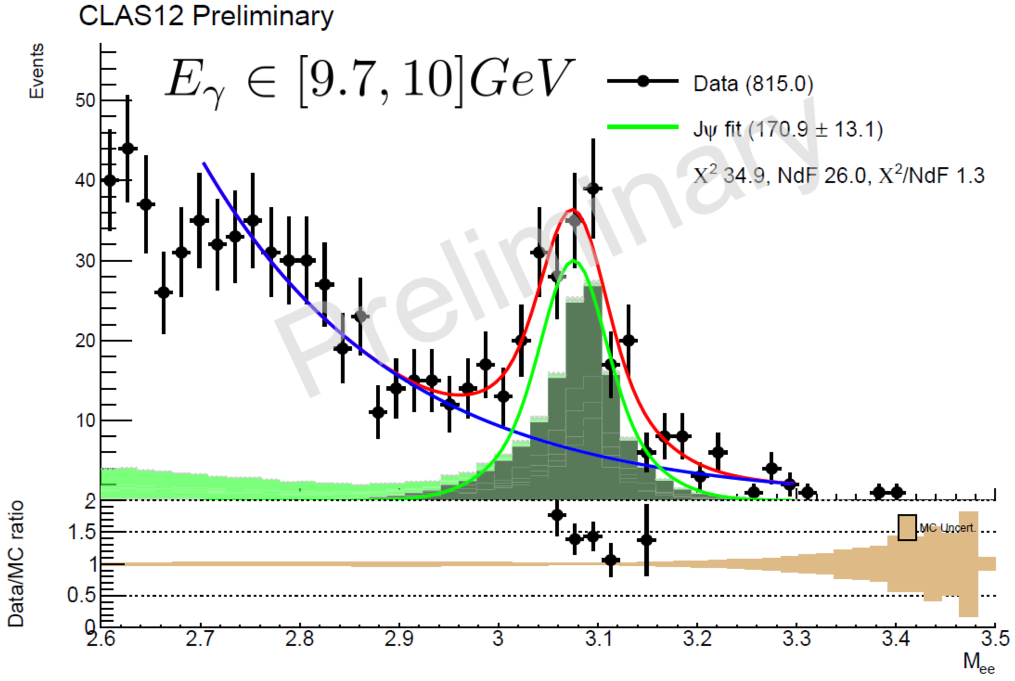
Number of photons and Number of targets

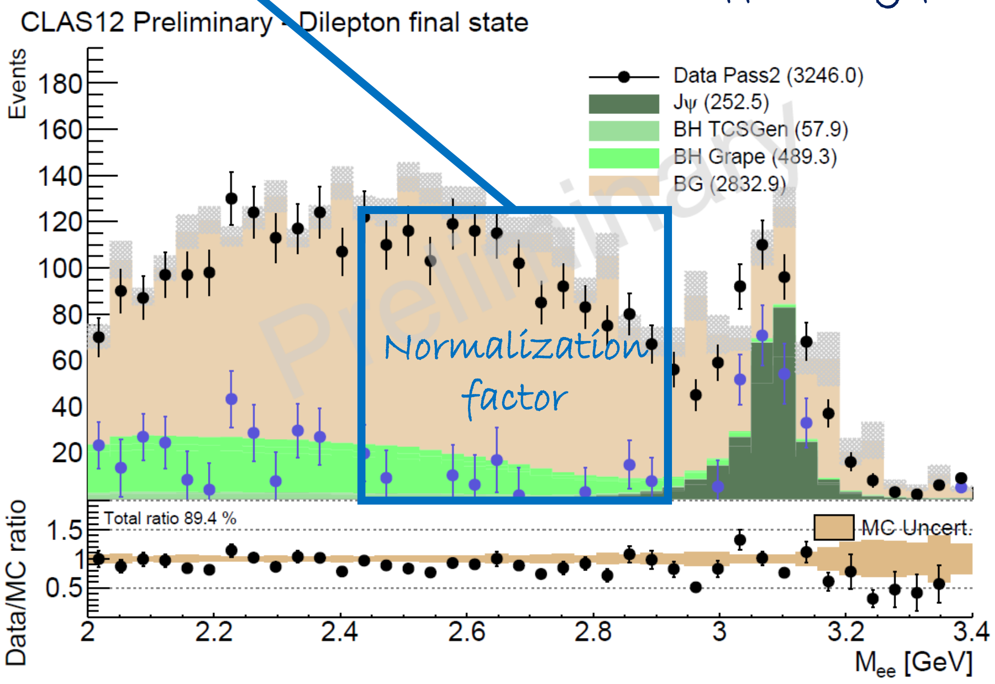
Branching ratio: 6%

Radiative corrections from MC

Reconstruction efficiency from MC

Normalization factor



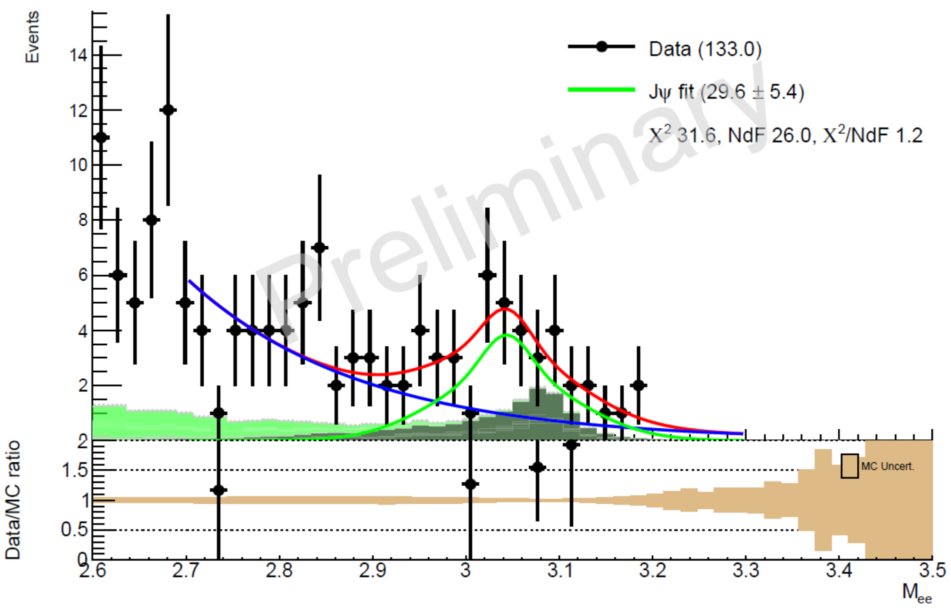


Number of J/Psi

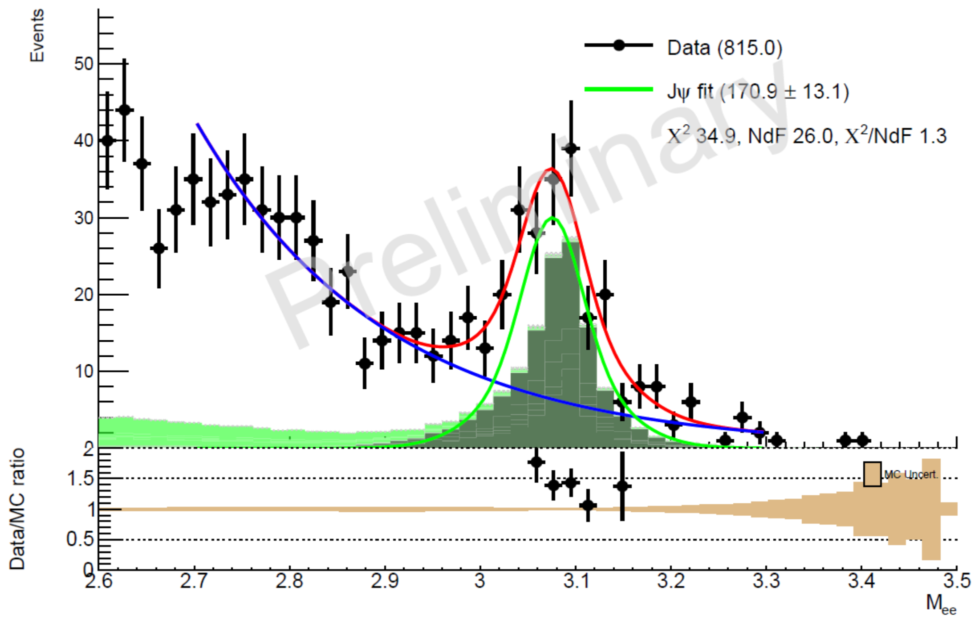
- All data samples are combined and **fitted together**
- Double-gaussian with common mean is used to fit the peak
- Error bar on number of J/Psi is set to sqrt(N)
- Systematic study to be performed on the fit function

$$\sigma_j = \frac{N_{J/\psi j}}{\mathcal{F}_j \cdot \mathcal{L} \cdot \omega_{cj} \cdot B_r \cdot \epsilon_j \cdot \epsilon_{Rad/j}}$$

$E_\gamma \in [8.9, 9.05] GeV$

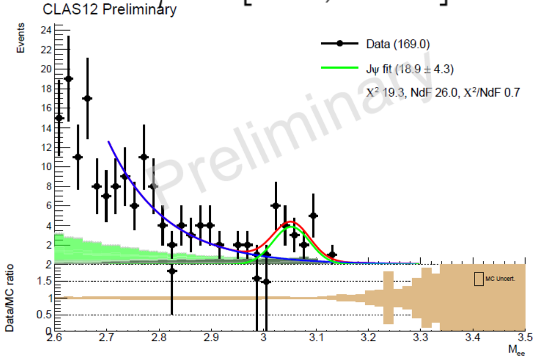


$E_\gamma \in [9.7, 10] GeV$

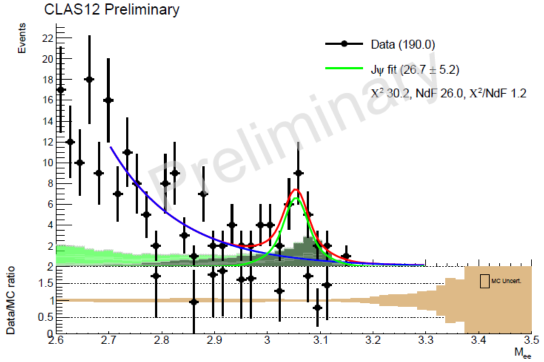


All Data fits

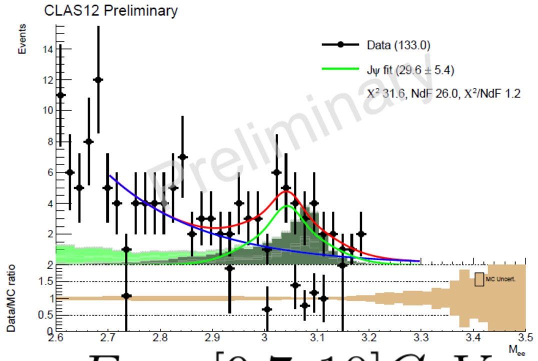
$E_\gamma \in [8.2, 8.65] GeV$



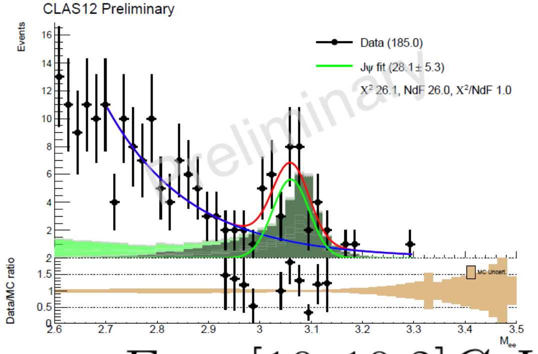
$E_\gamma \in [8.65, 8.9] GeV$



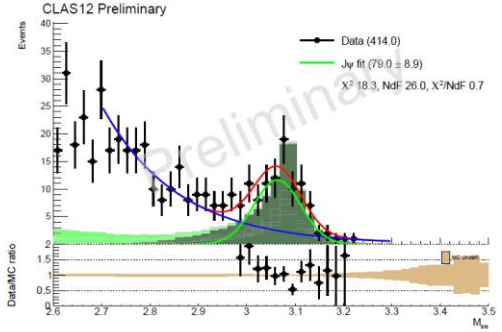
$E_\gamma \in [8.9, 9.05] GeV$



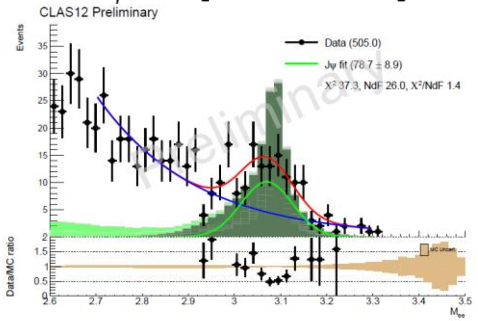
$E_\gamma \in [9.05, 9.2] GeV$



$E_\gamma \in [9.2, 9.46] GeV$

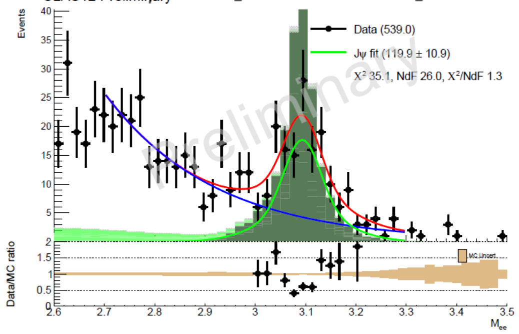


$E_\gamma \in [9.46, 9.7] GeV$

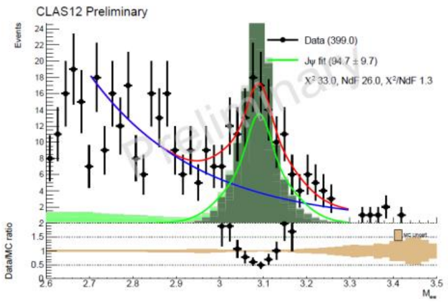


$E_\gamma \in [9.7, 10] GeV$

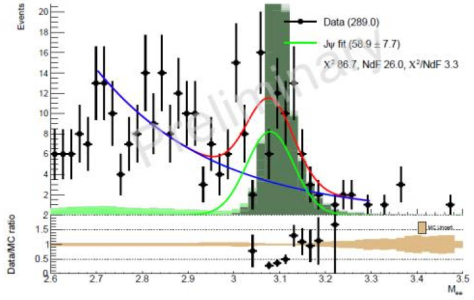
$E_\gamma \in [10, 10.2] GeV$



$E_\gamma \in [10.2, 10.4] GeV$



$E_\gamma \in [10.4, 10.6] GeV$



Total of 707 J/ψ

Jefferson Lab clas

Photon flux

1) Real and virtual flux are provided event by event by the JPsiGen Generator: https://github.com/JeffersonLab/JPsiGen/blob/eb40dd934bb9f022873414a57e0dad9d1ccbcbdf/src/KinFunctions.cc#L38
2) The integral over the range of energy of the bin j is done using the integral/mean theorem:

$$\mathcal{F}_{c/j} = \int_j \mathcal{F}_c dE = \Delta E \frac{1}{N} \frac{\sum_{i=1}^{N} \mathcal{F}_c(E_{GEN/i}) \cdot \omega_i}{\sum_{i=1}^{N} \omega_i}$$

3) Each flux (one per configuration) is multiplied by the corresponding accumulated charge:

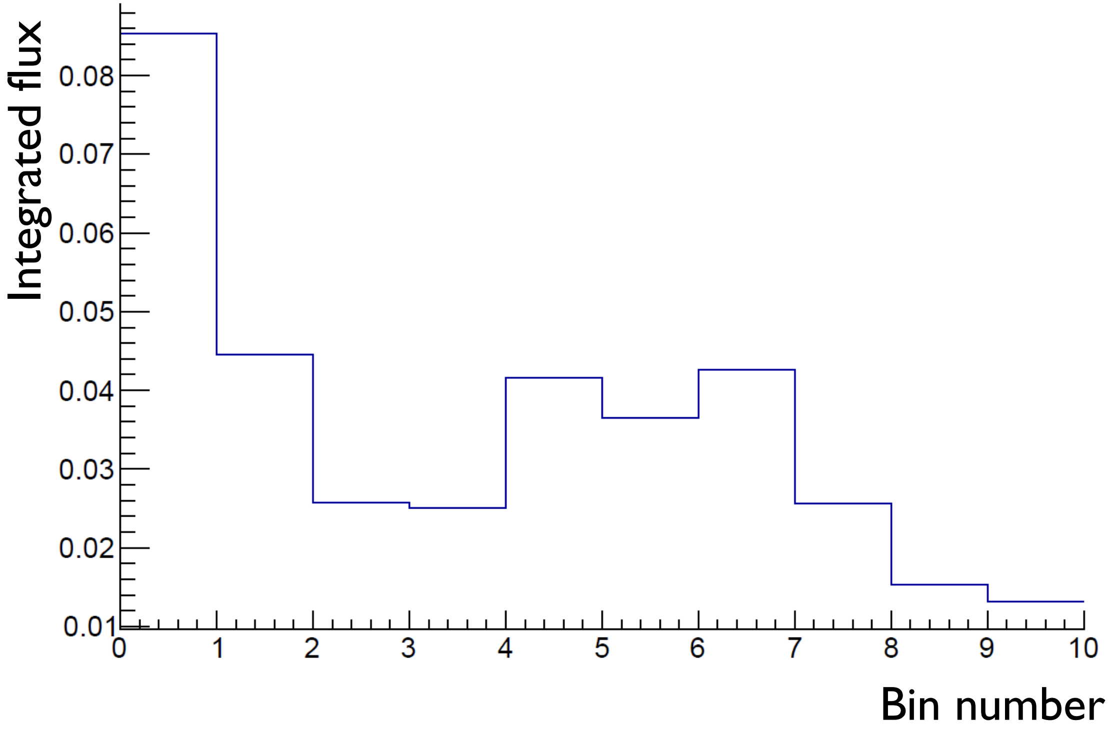
$$\mathcal{F}_j = \sum_c C_c \cdot \mathcal{F}_{c/j}$$

Total number of photon in the bin j in unit of e

4) The results is multiplied by the luminosity factor to recover the correct normalizing factor:

$$\mathcal{L} = \frac{l \cdot \rho \cdot N_A \cdot C}{e}$$

$$\sigma_j = \frac{N_{J/\psi j}}{\mathcal{F}_j \cdot \mathcal{L} \cdot \omega_{cj} \cdot B_r \cdot \epsilon_j \cdot \epsilon_{Rad/j}}$$

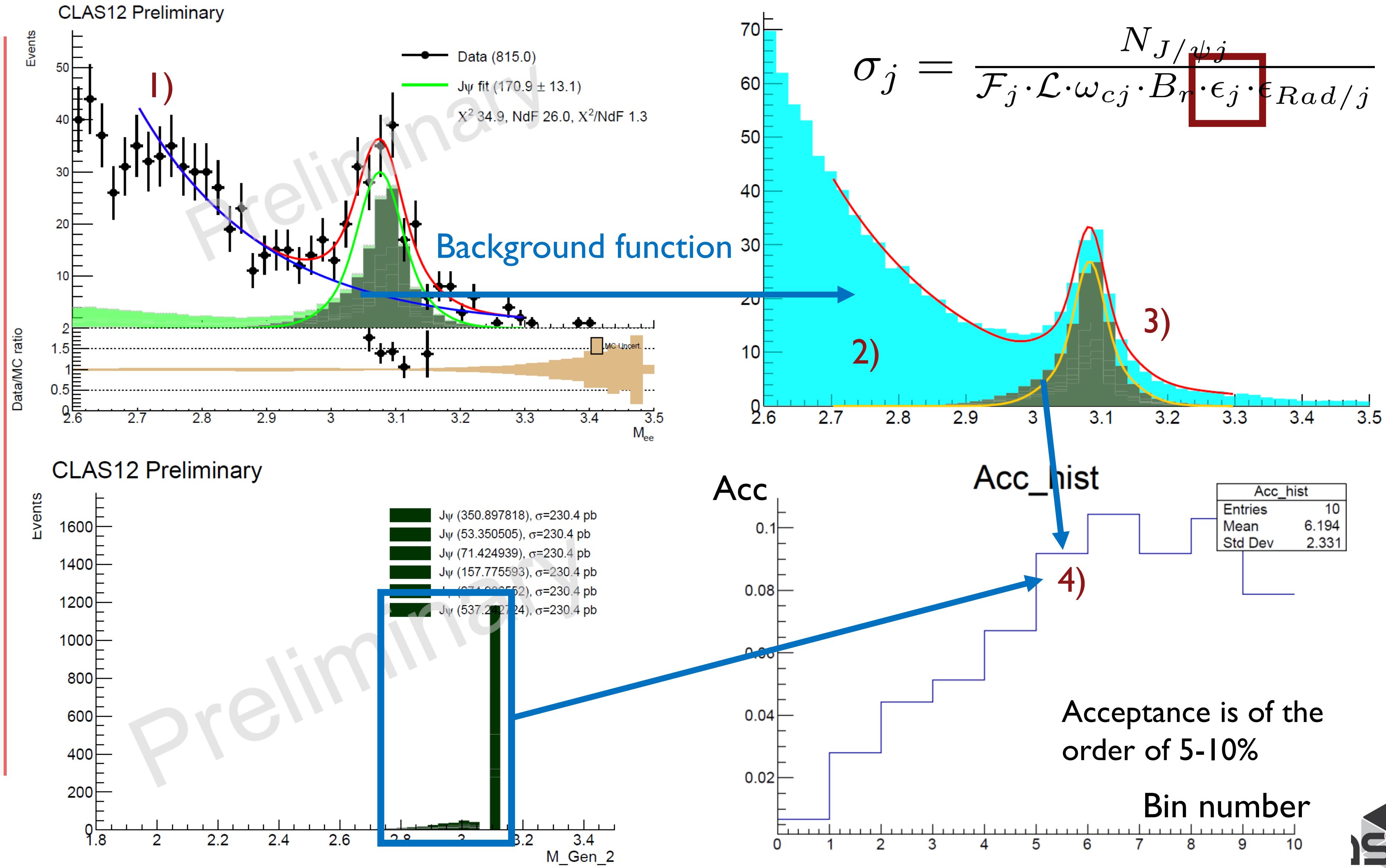


Detection efficiency computation

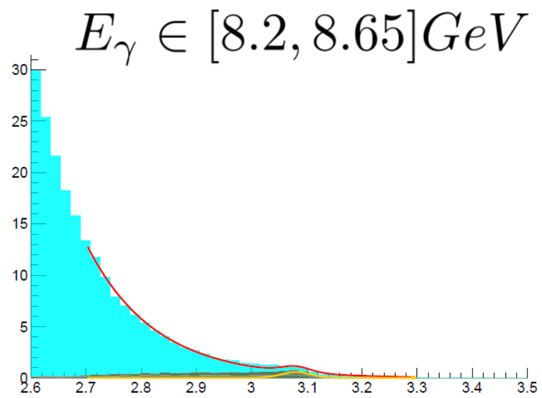
1) From the data fit a second order polynomial background function is extracted
2) Events are generated according to this background function and added to the Jpsi signal MC sample
3) The obtained distribution is fitted with the same function as the data
4) The acceptance correction is then:

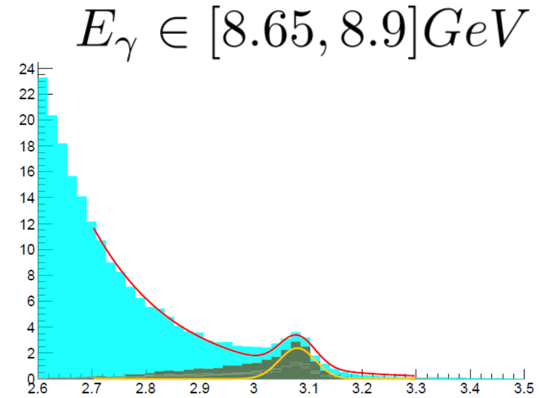
$$\epsilon_j = \frac{N_{J/\psi}\big|_{j/REC}}{N_{J/\psi}\big|_{j/RAD}}$$

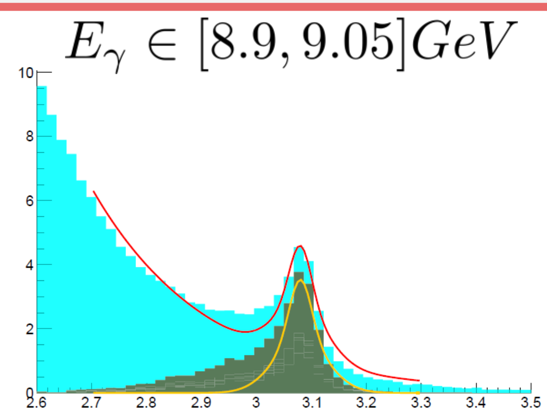
$$\sigma_j = \frac{N_{J/\psi j}}{\mathcal{F}_j \cdot \mathcal{L} \cdot \omega_{cj} \cdot B_r \cdot \epsilon_j \cdot \epsilon_{Rad/j}}$$

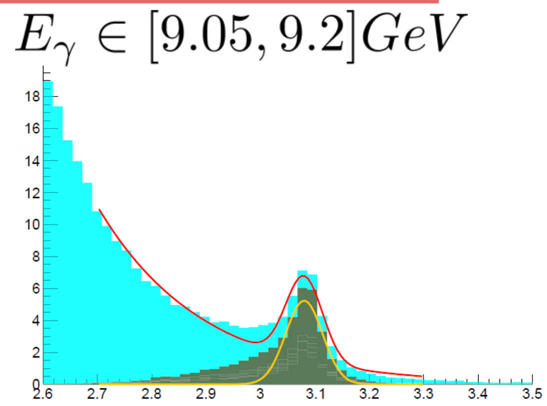


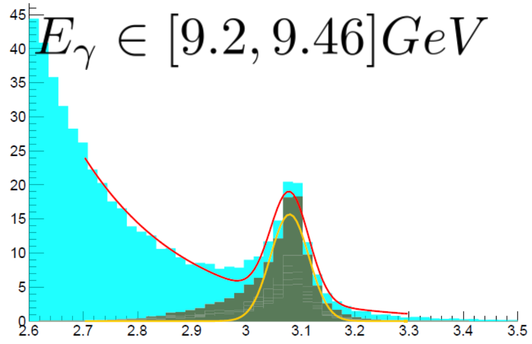
All MC fits

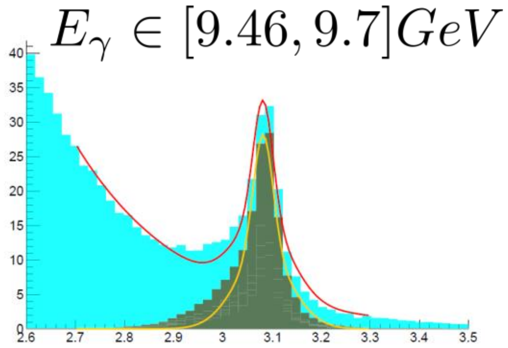


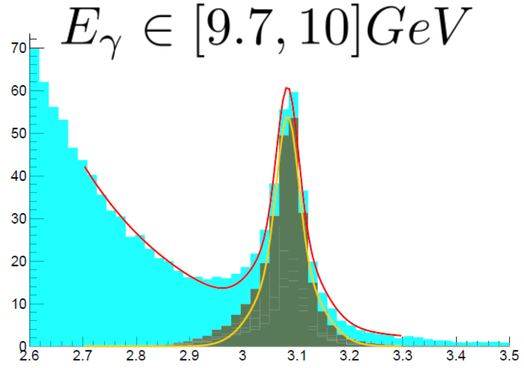


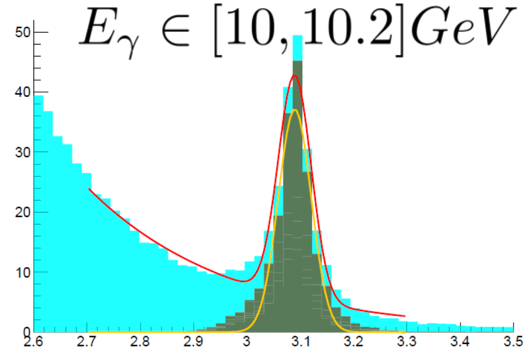


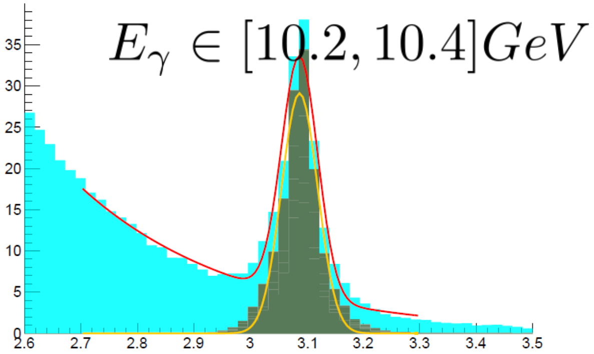


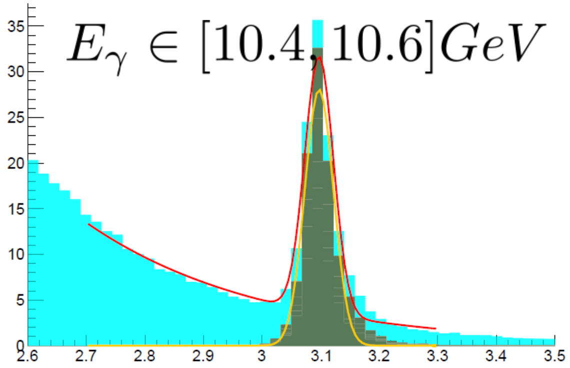












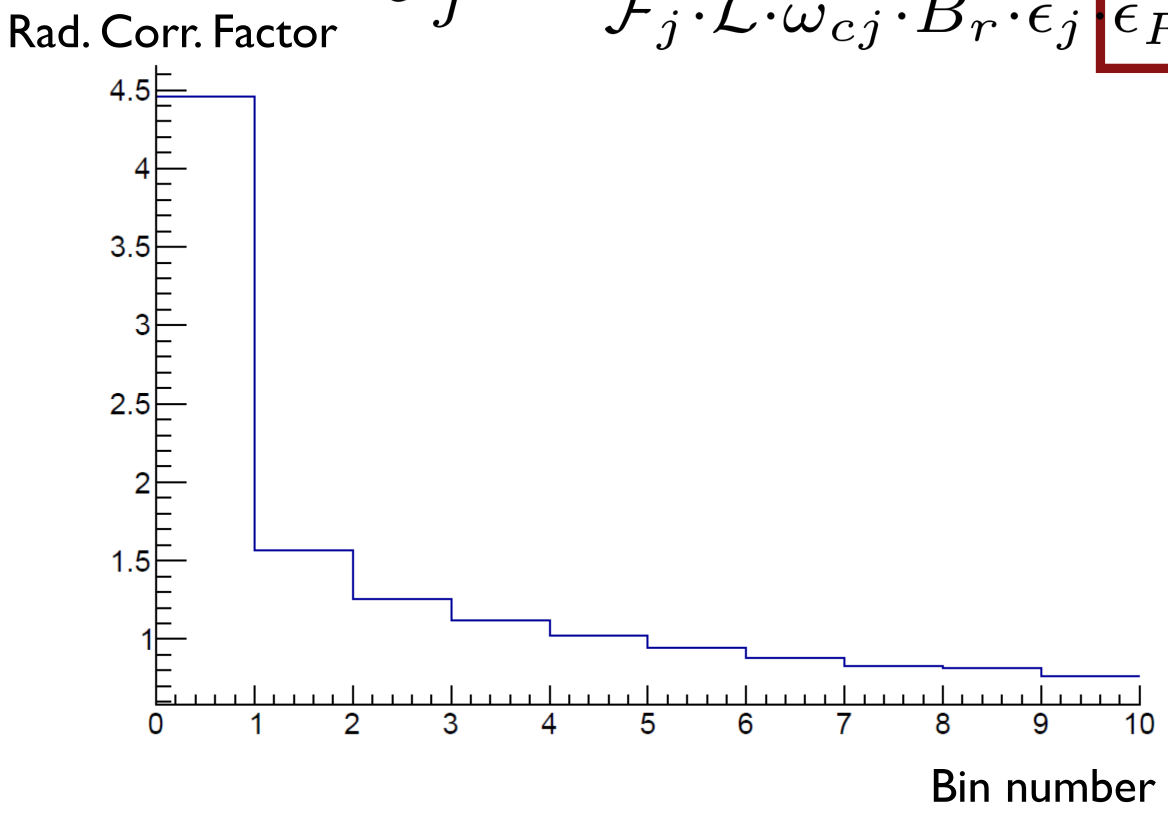
Jefferson Lab clas

Radiative correction

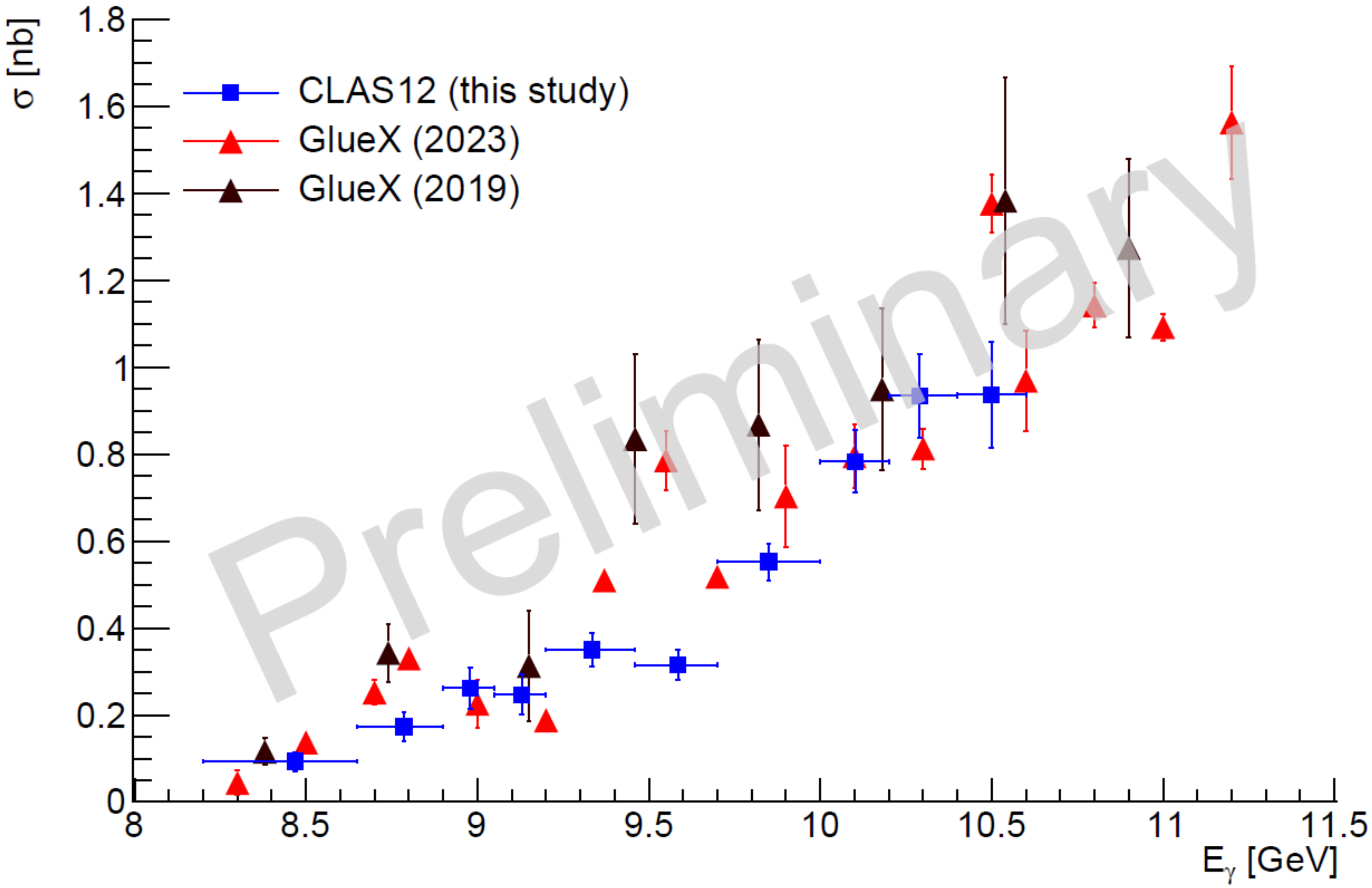
1) Jpsi samples without radiative effects are produced
2) The radiative correction is defined using the GEN kinematics as:

$$\epsilon_{Rad/j} = \frac{N_{J/\psi}\big|_{j/RAD}}{N_{J/\psi}\big|_{j/GEN}}$$

$$\sigma_j = \frac{N_{J/\psi j}}{\mathcal{F}_j \cdot \mathcal{L} \cdot \omega_{cj} \cdot B_r \cdot \epsilon_j \cdot \boxed{\epsilon_{Rad/j}}}$$

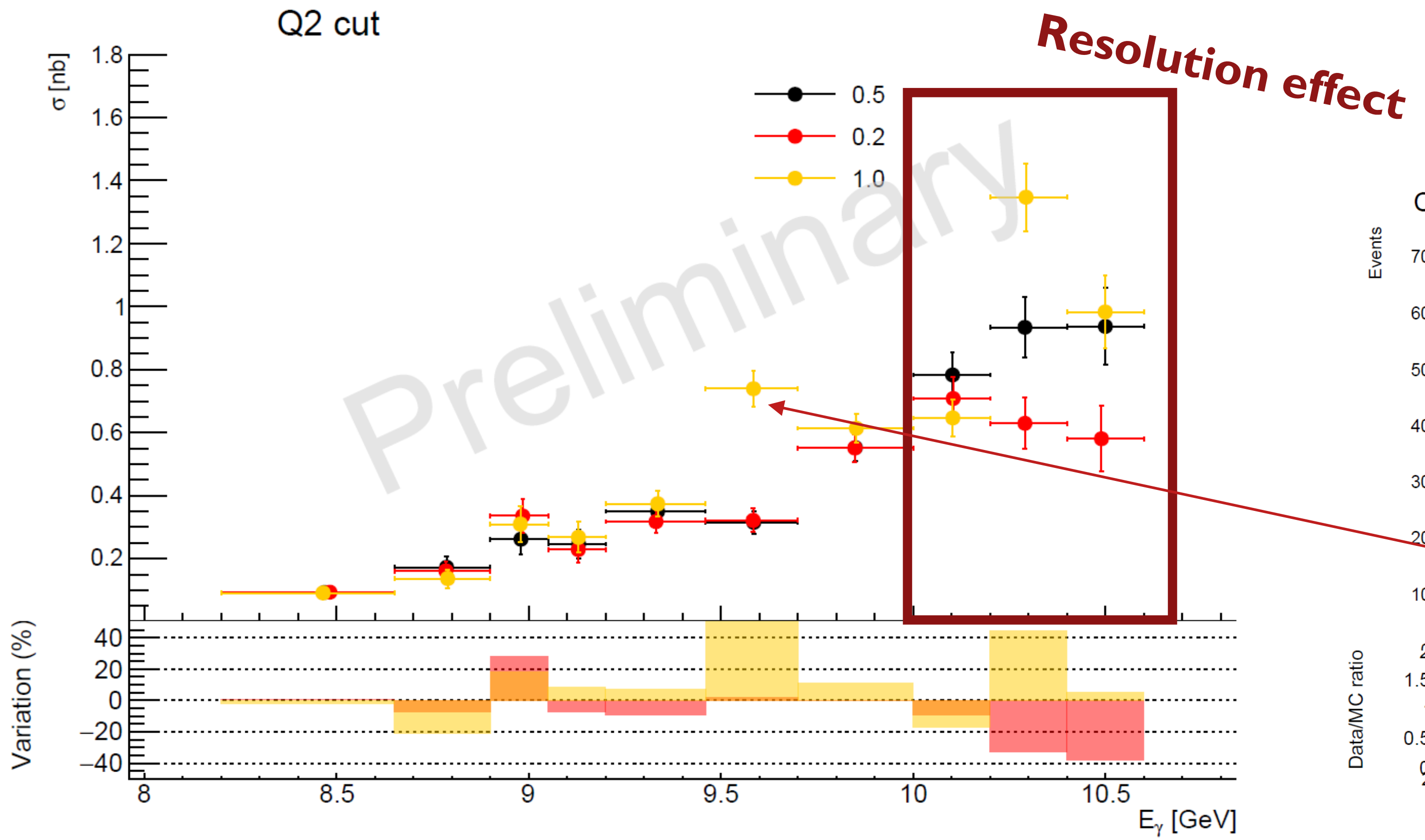


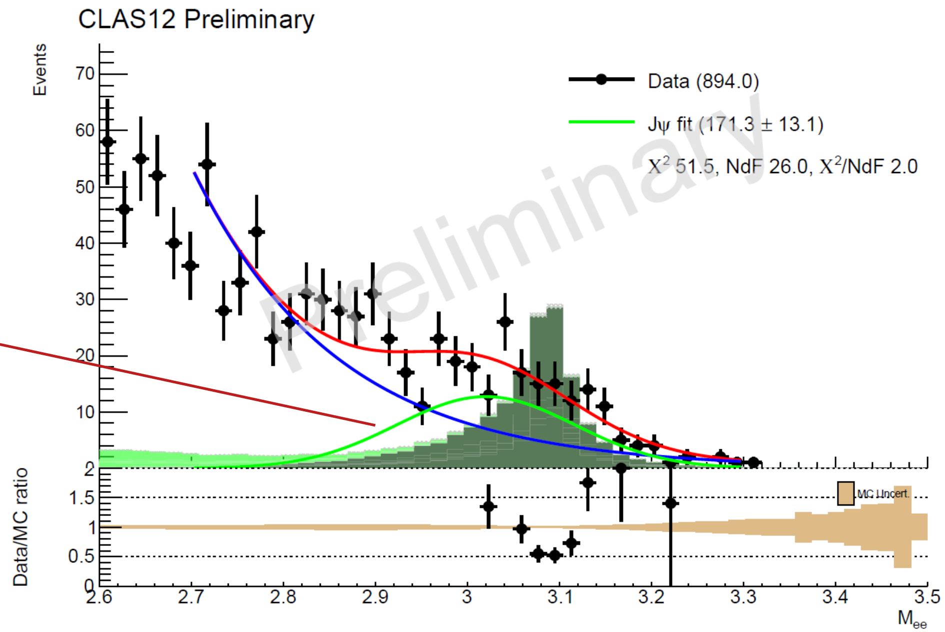
Preliminary cross-section as a function of Eγ



Selection cut systematics

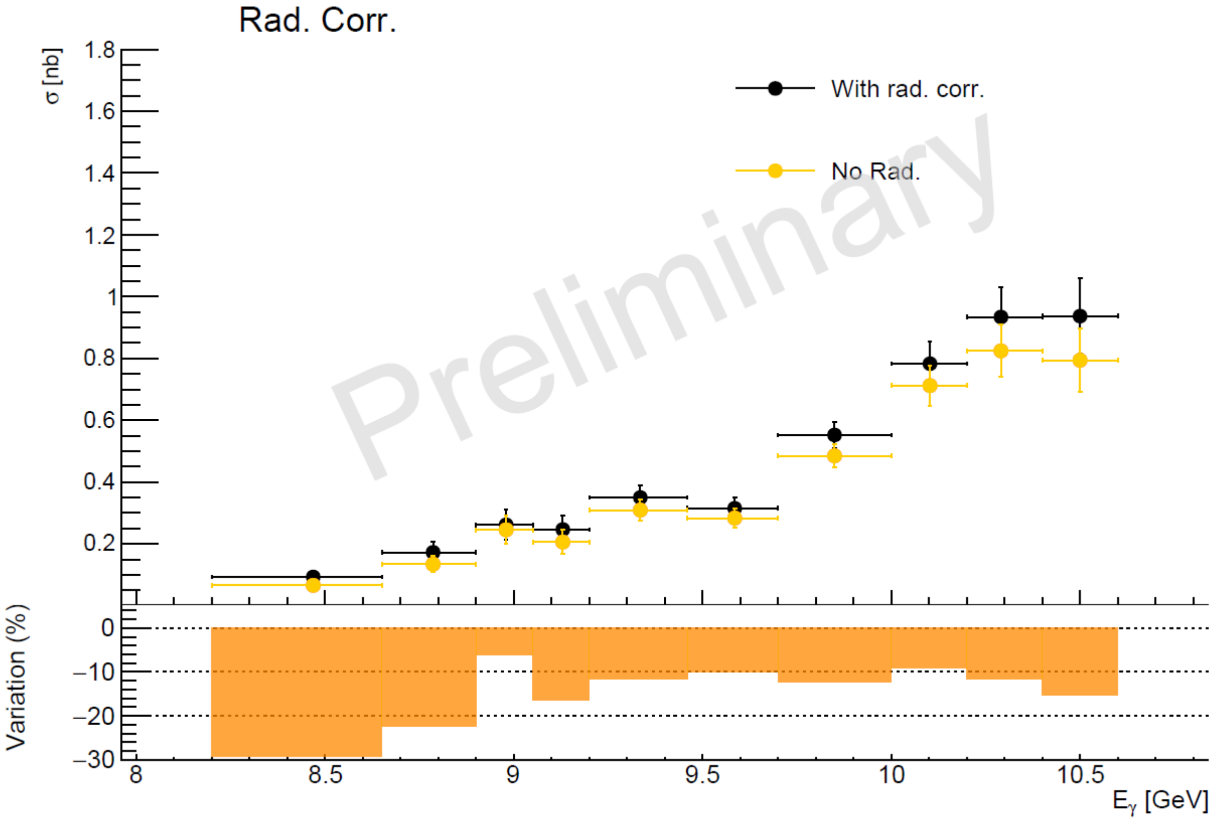
- Every step of the analysis is repeated with different cuts:
  - Q^2 **DONE**
  - $|MM|^2$ **To be done**
  - Proton PID **To be done**
  - Lepton momenta cut **To be done**
  - Lepton ID cut **To be done**





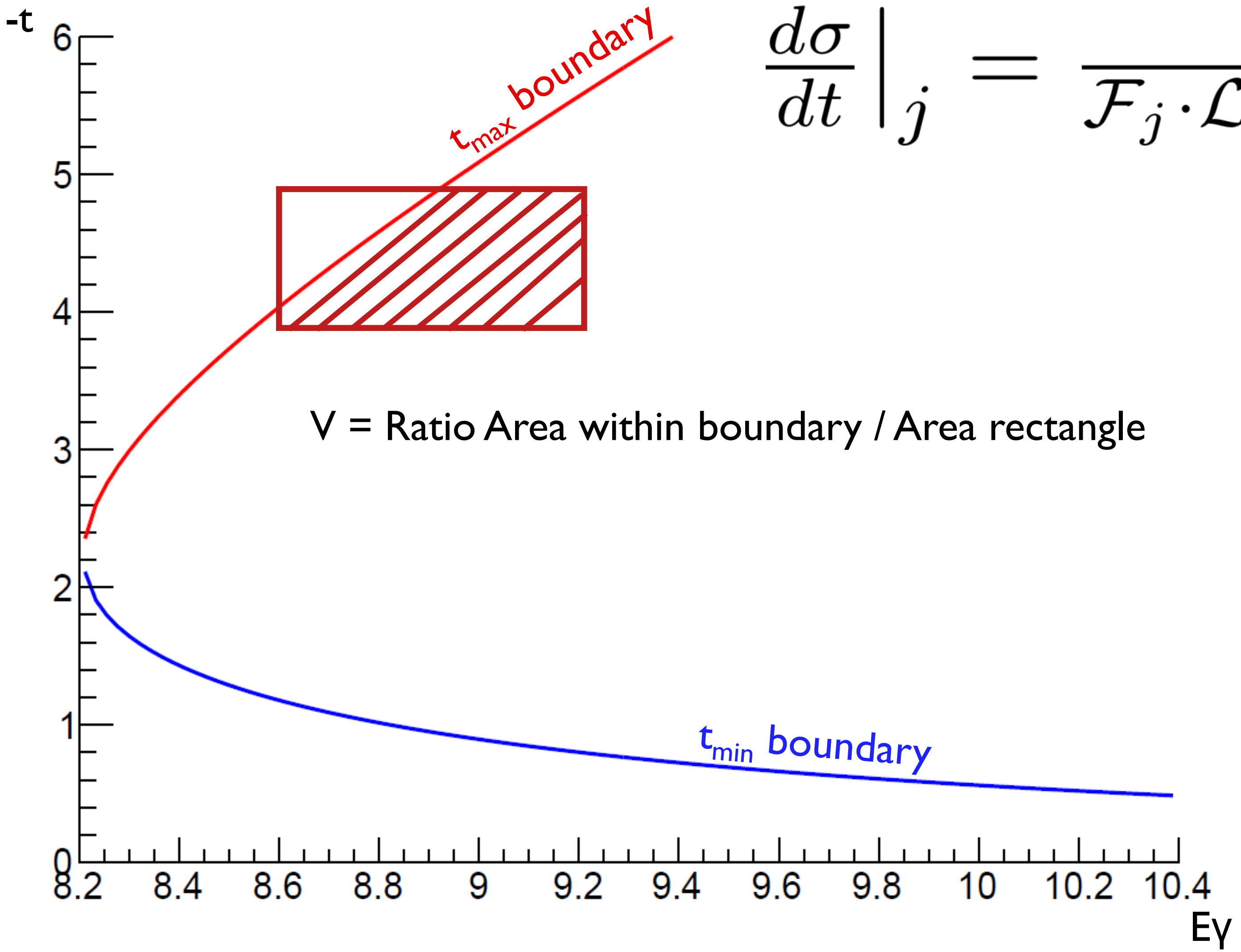
Radiative correction systematic

- The standard CS is extracted using the Radiated Jpsi MC samples and radiative correction
- The alternate is using non-radiated MC samples
- The effect is of the order of 10% (GlueX quoted 8.5%)



+ Closure test (Implemented but not presented here)

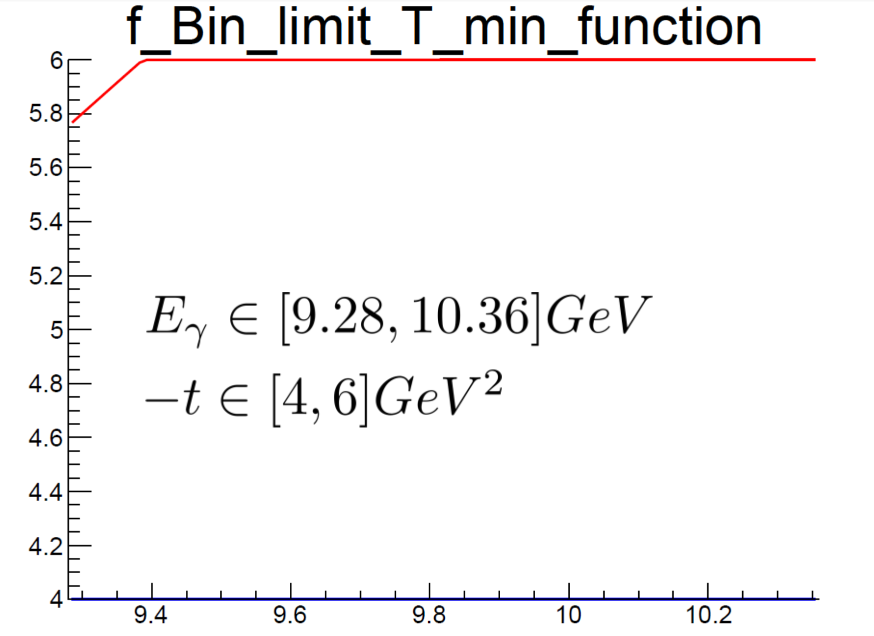
Bin volume correction



$$\left.\frac{d\sigma}{dt}\right|_j = \frac{N_{J/\psi/j}}{\mathcal{F}_j \cdot \mathcal{L} \cdot \omega_{c/j} \cdot B_r \cdot \epsilon_j \cdot \epsilon_{Rad/j} \cdot \mathcal{V}_j \cdot \Delta t_j}$$

V = Ratio Area within boundary / Area rectangle

- In practice is this readily done using integral of functions in root



t-dependence of the cross-section

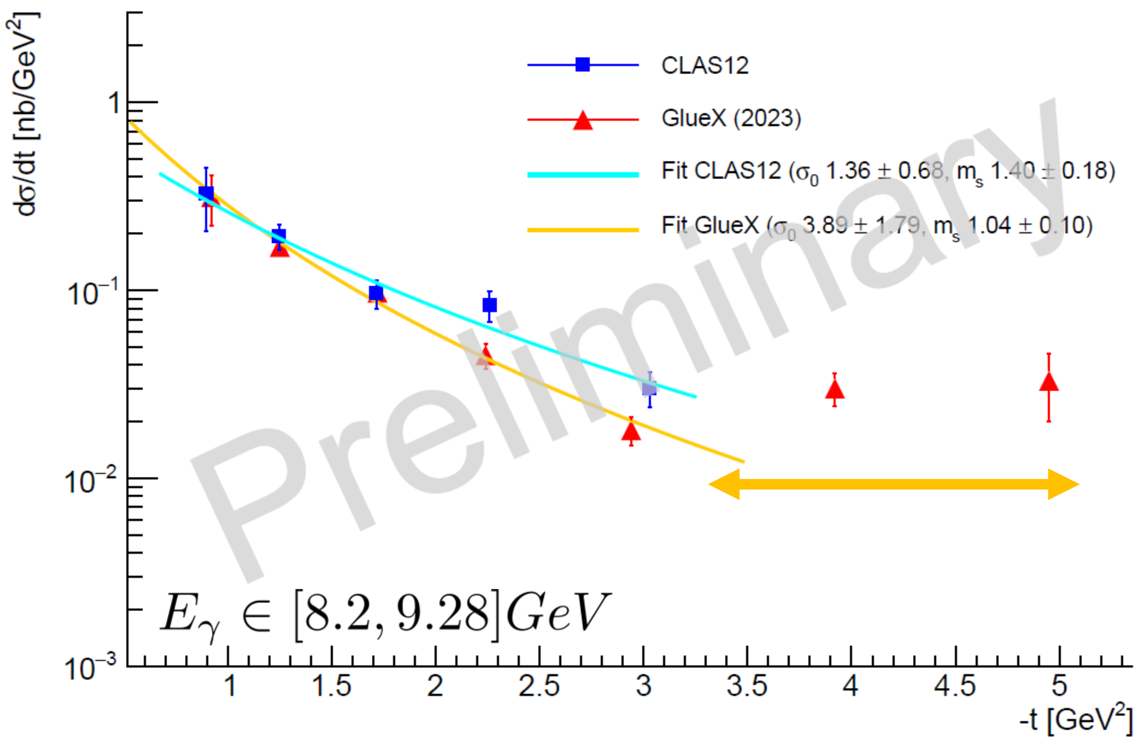
- The t-dependent cross-section can be parametrized as:

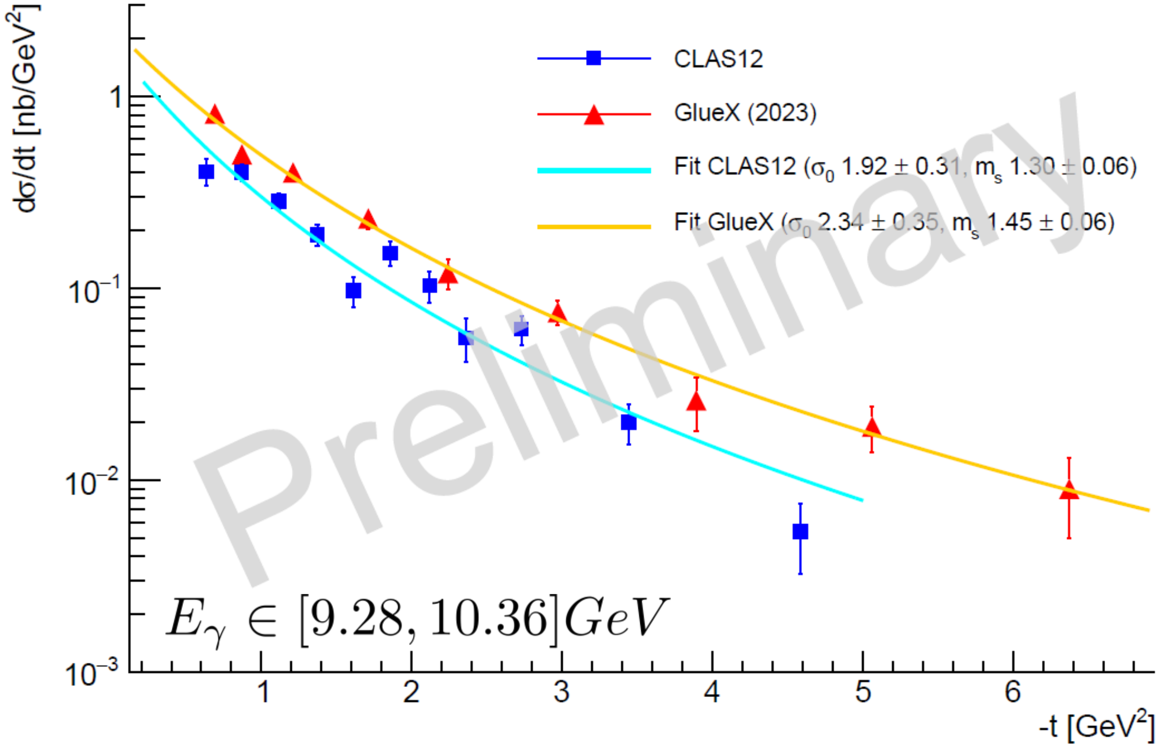
$$\frac{d\sigma}{dt} = \frac{d\sigma}{dt}\Big|_0 \cdot \frac{1}{(1-t/m_S^2)^4}$$

- ms can be interpreted as the mass radius of the proton.
- Our results are consistent with GlueX
- We do not see/are not sensitive to the cusp at small E\gamma and large t

GlueX 2023 publication

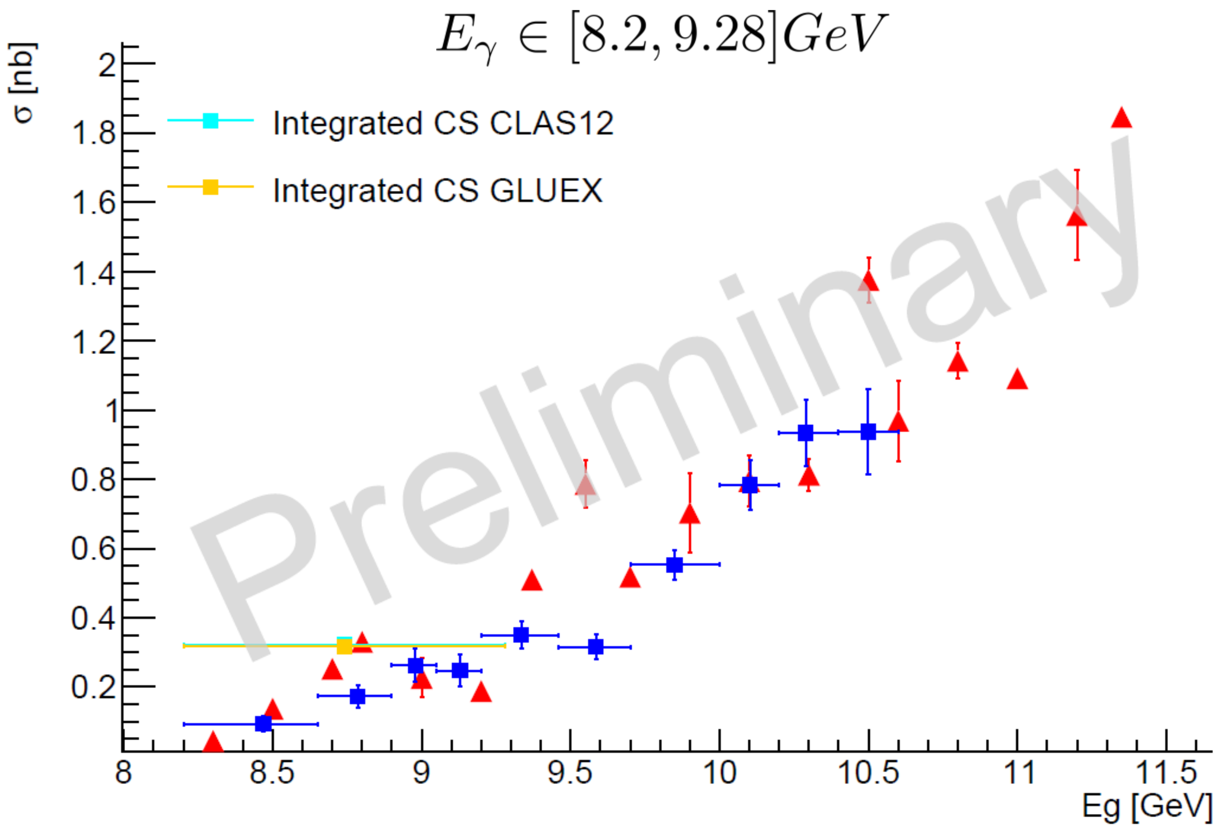
$\langle E_\gamma \rangle$ [GeV]	8.93	9.86	10.82
q [GeV]	0.499	0.767	0.978
$d\sigma/dt(0)$ [nb/GeV^2]	2.863 $\pm$1.95	2.205 $\pm$0.380	4.268 $\pm$0.564
m_s [GeV]	1.105 $\pm$0.168	1.472 $\pm$0.075	1.313 $\pm$0.049

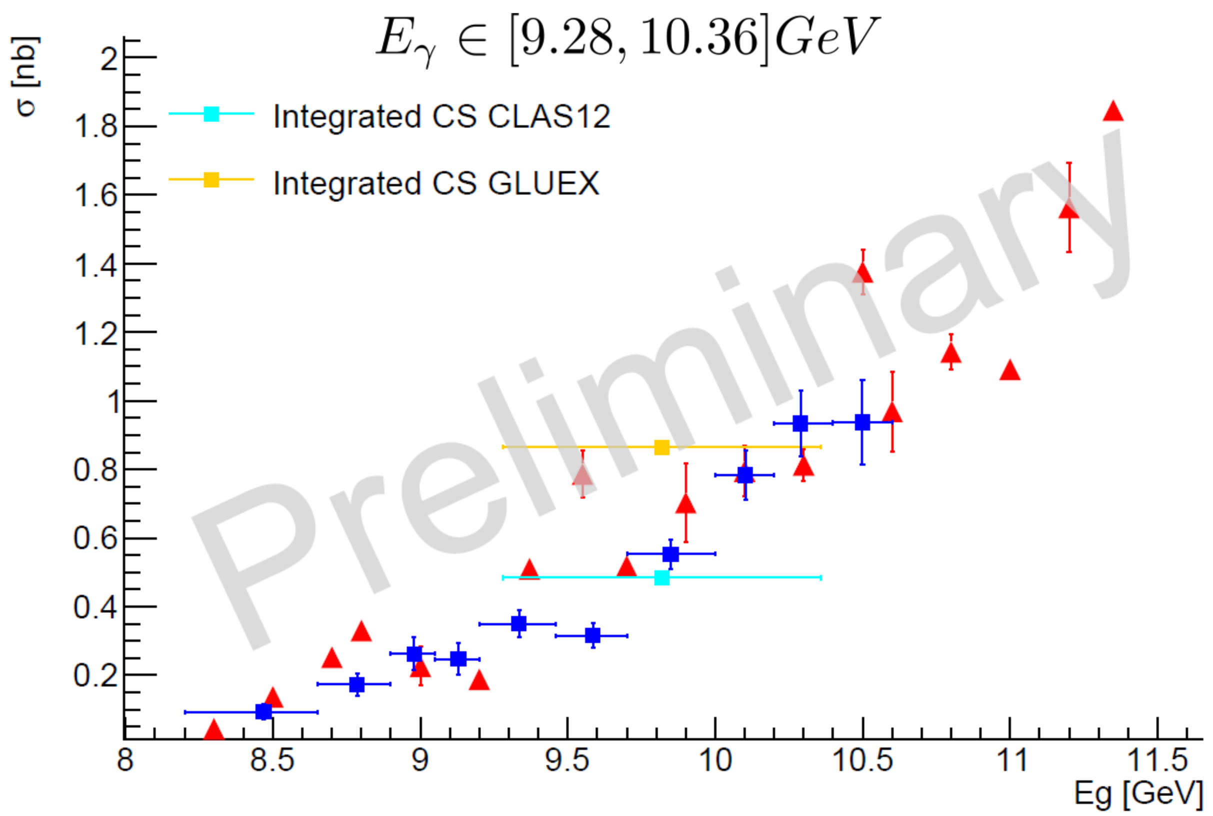




Integrated t-dependent cross-section

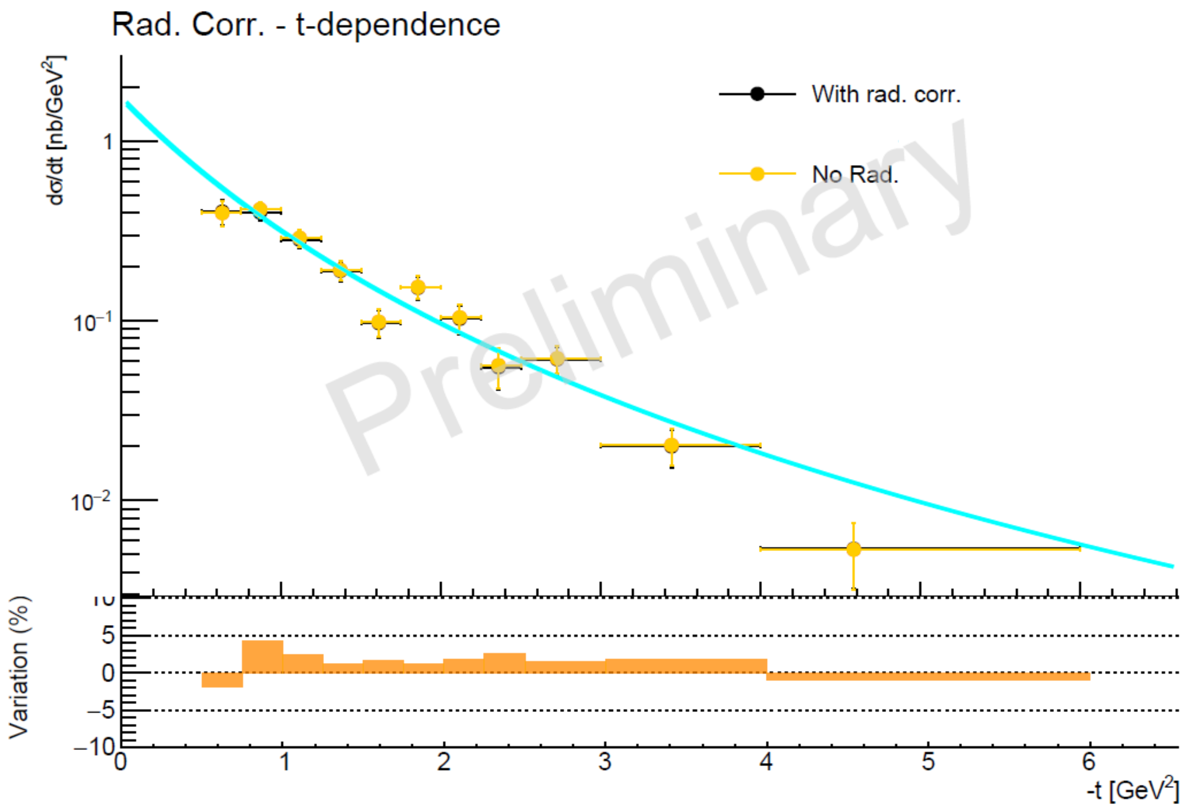
- The integral of the t-dependent cross section is done bin-by-bin: $\sigma = \sum_j \left.\frac{d\sigma}{dt}\right|_j \cdot \Delta t_j$
- And compared to the total CS



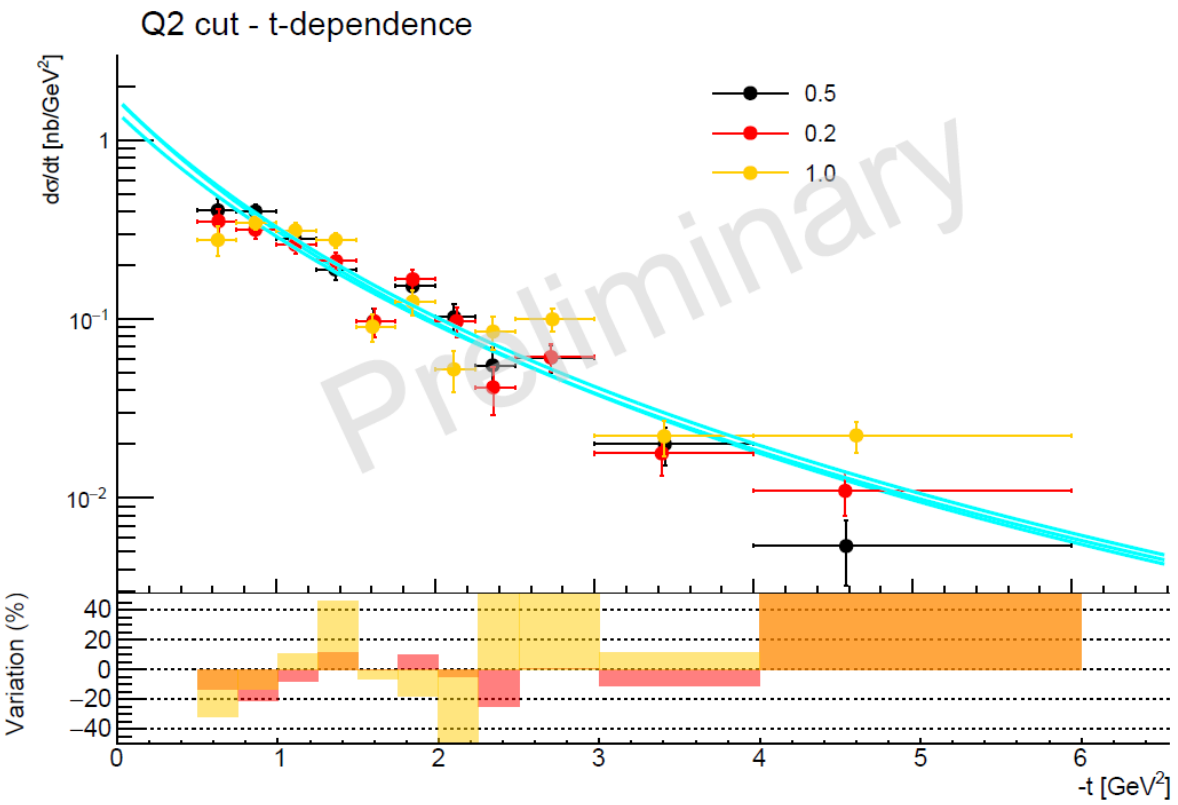


- Qualitative good agreement !

Systematic studies on the t-dependent cross-section



- No dependence on –t (expected)
- ~2-3% variation



- Large variation mostly due to the fitting
  → Systematic way to choose the binning

Take-aways and path going forward

I The JPsi analysis is at an advanced stage.

II Data and MC samples have been produced, the framework to analyze them is final.

III Some common tools remain to be developed and used in the analysis.

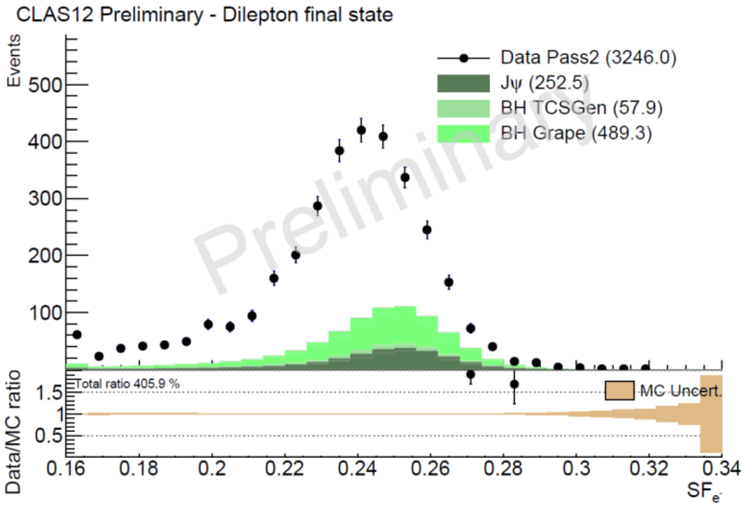
IV Current effort in on finalizing the background modelization and normalization factor (Final results for the collaboration meeting).

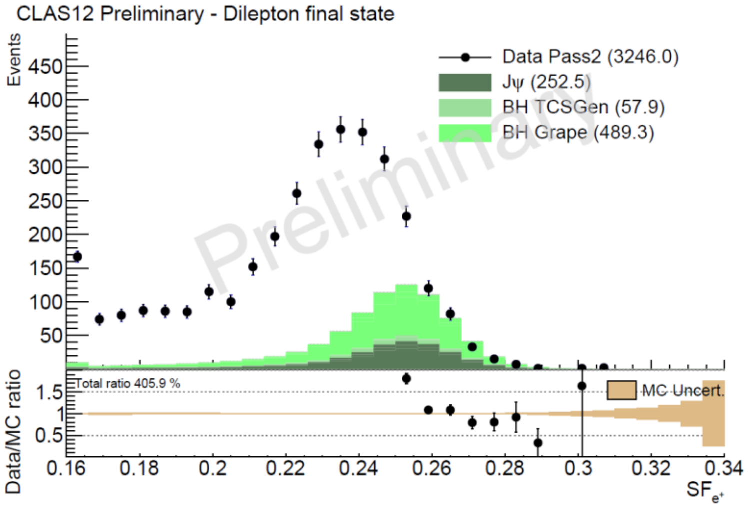
V **A release note will be ready by early March at the latest. An analysis note will be ready for the summer**

Back-up

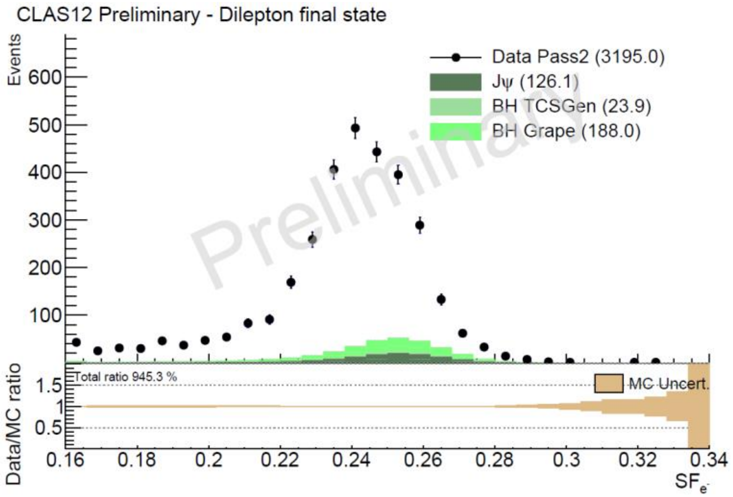
Sampling fraction MC/Data mismatch

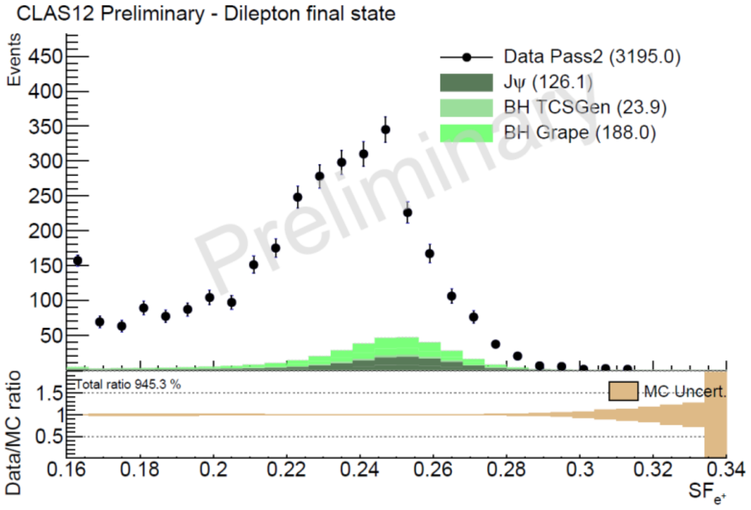
Inbending Fall 2018





Outbending Fall 2018





Jefferson Lab clas

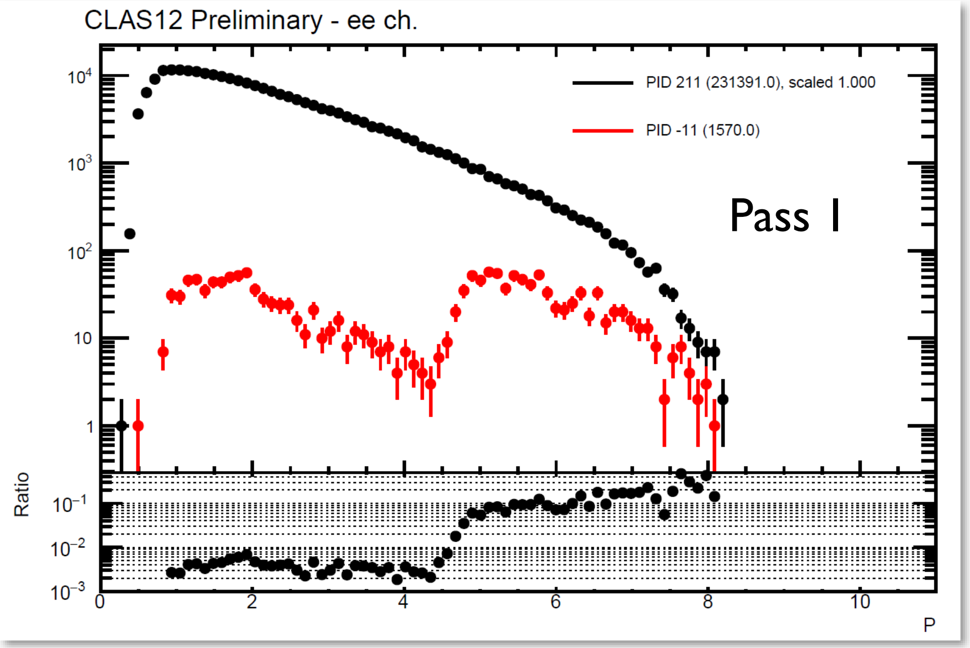
Lepton PID using machine learning (Status as of October 2023)

Motivations and previous work

Motivations

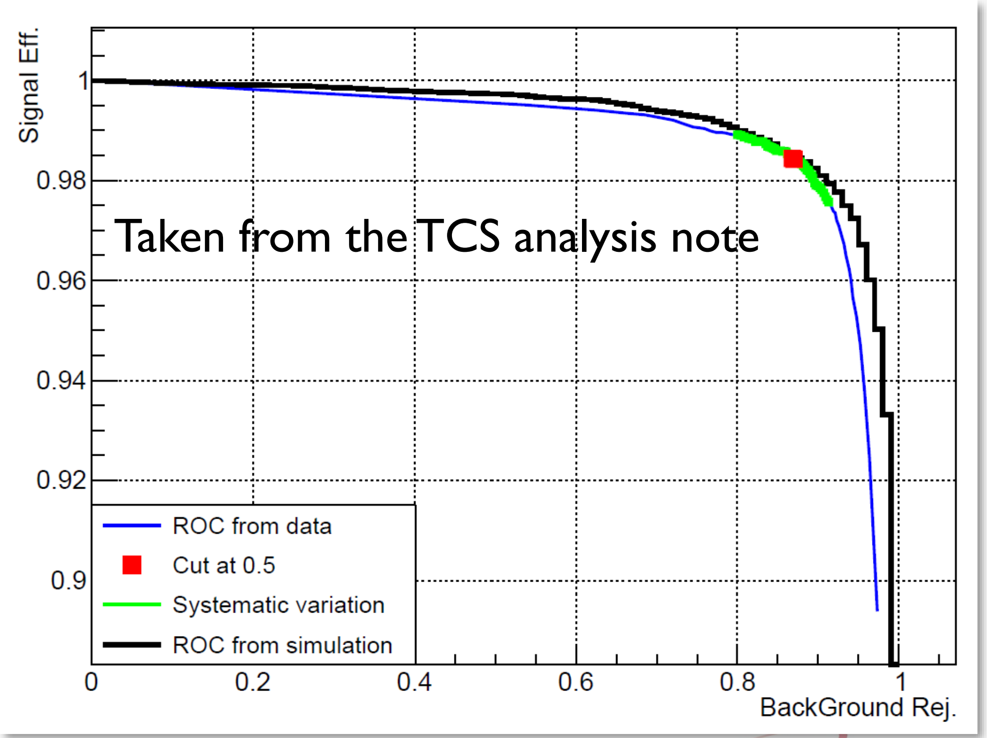
- Above the HTCC, threshold both pions and leptons produce a HTCC signal. In the EB, only ECAL provide a separation between the two.
- ep → epπ$^+$(π$^-$) is a large background at large positron momenta

$$ep \to ep\pi^+\pi^- \text{ VS } ep \to epe^+\pi^-$$



Previous work and motivations

- Long standing feature, already solved for the TCS publication
- Use the layer segmentation of the ECAL to provide separation
  - Variables used: SFs and m2 of PCAL, ECIN, ECOUT
  - Method tested: **NN**, BDT



Current status

All material of this section provided by M. Tenorio Pita

Approach

- For both electrons and positrons, and for each RGA configuration:

2 (e^+/e^-) x 3 (Spring19/Fall18 in/out) = 6 classifiers

- Use the layer segmentation of the ECAL to provide separation

Variables used: P, θ, ϕ, SFs and m2 of PCAL, ECIN, ECOUT

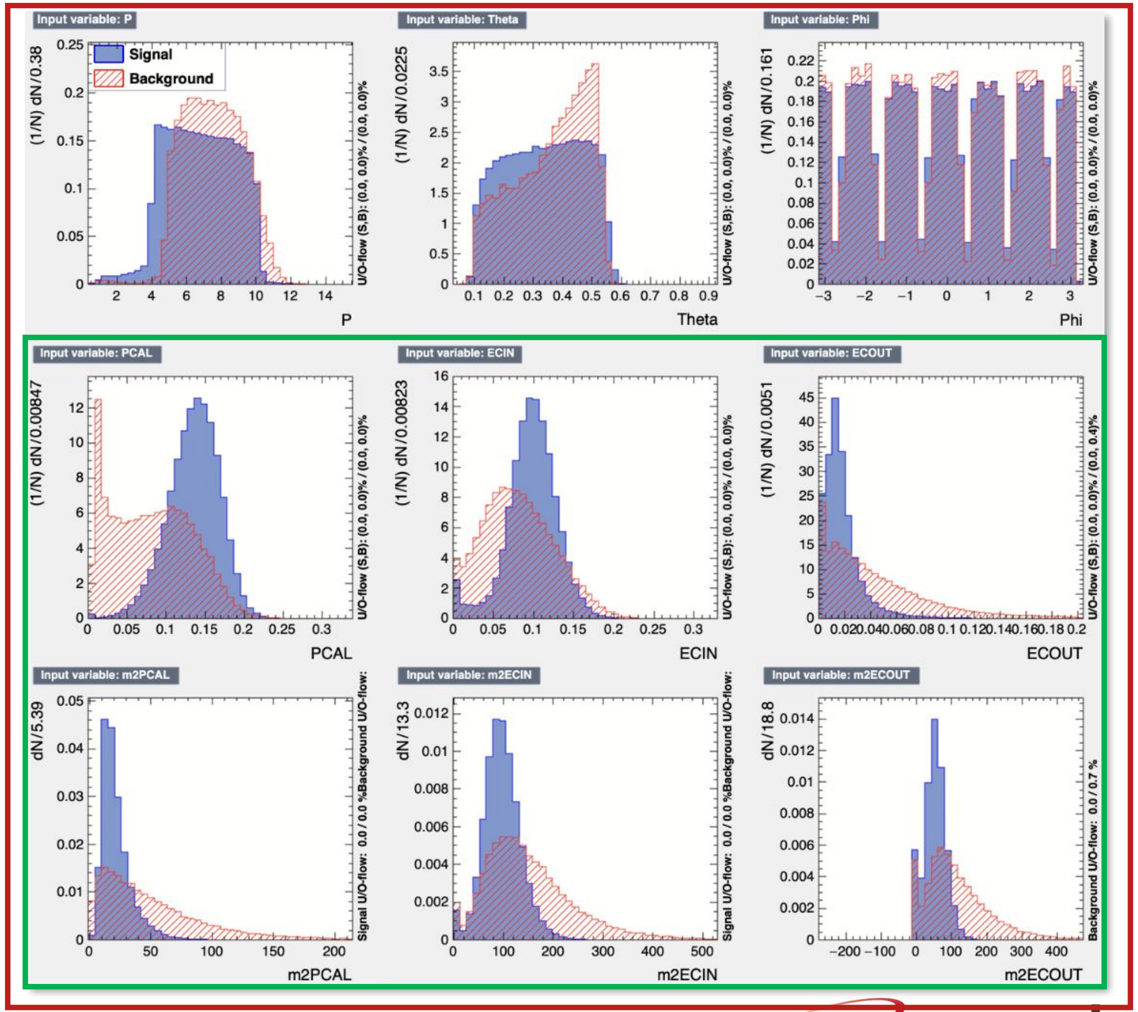
Method tested: NN, BDT

- Trained on simulation:

Signal: flat $e^{+/-}$ distribution, reconstructed as $e^{+/-}$

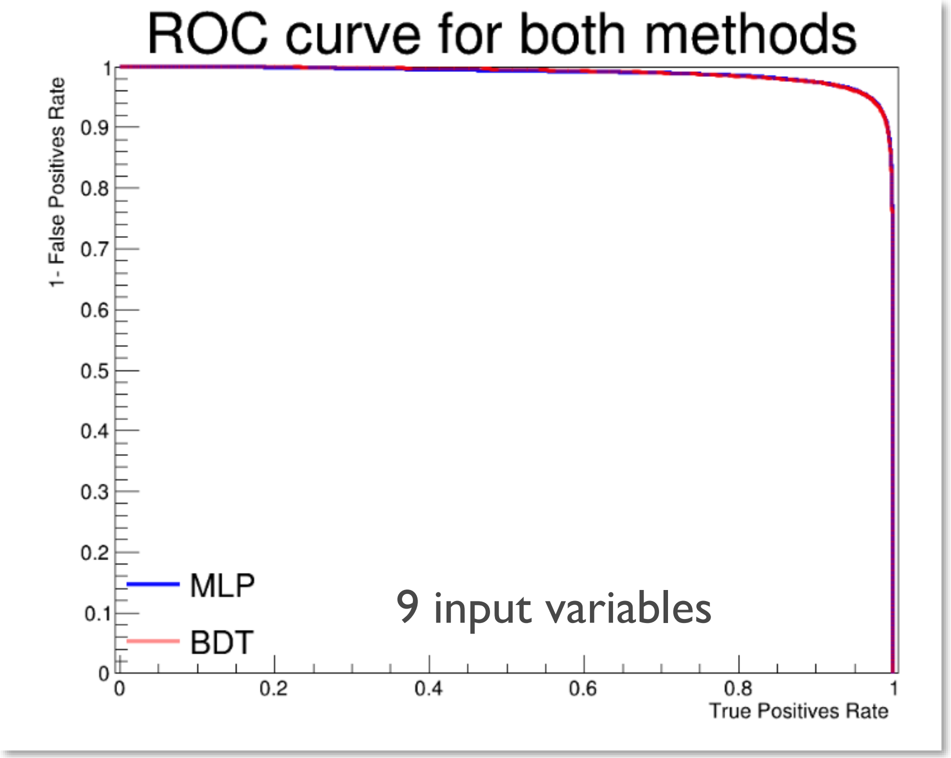
Background: flat $\pi^{+/-}$ distribution, reconstructed as $e^{+/-}$

Input variables for signal (blue) and background (red)



Performances

- We tested both 6 and **9 input variables**, for 2 methods **NN** and BDT.
- Signal efficiency: 99.4 %
- Background rate: 10%



NN 6 var.	Actual e+ (653771)	Actual π+ (101499)
Predicted e+	647688	12805
Predicted π+	6083	88694
	TPR 99.1 %	FPR 12.6 %

NN 9 var.	Actual e+	Actual π+
Predicted e+	649244	10158
Predicted π+	4527	91341
Performances	TPR 99.4%	FPR 10%

Validation on simulation

- Signal efficiency and background reduction as a function of particle kinematics
- Done on separate samples

NN. 9 Variables