



UNIVERSITÀ DEGLI STUDI DI MILANO

Misura del rapporto di efficienza di identificazione del fotone tra dati e simulazione Montecarlo per il Run2 dell'esperimento ATLAS ad LHC

Tesi di Federico Marletta

Relatore: Prof. Leonardo Carlo Carminati

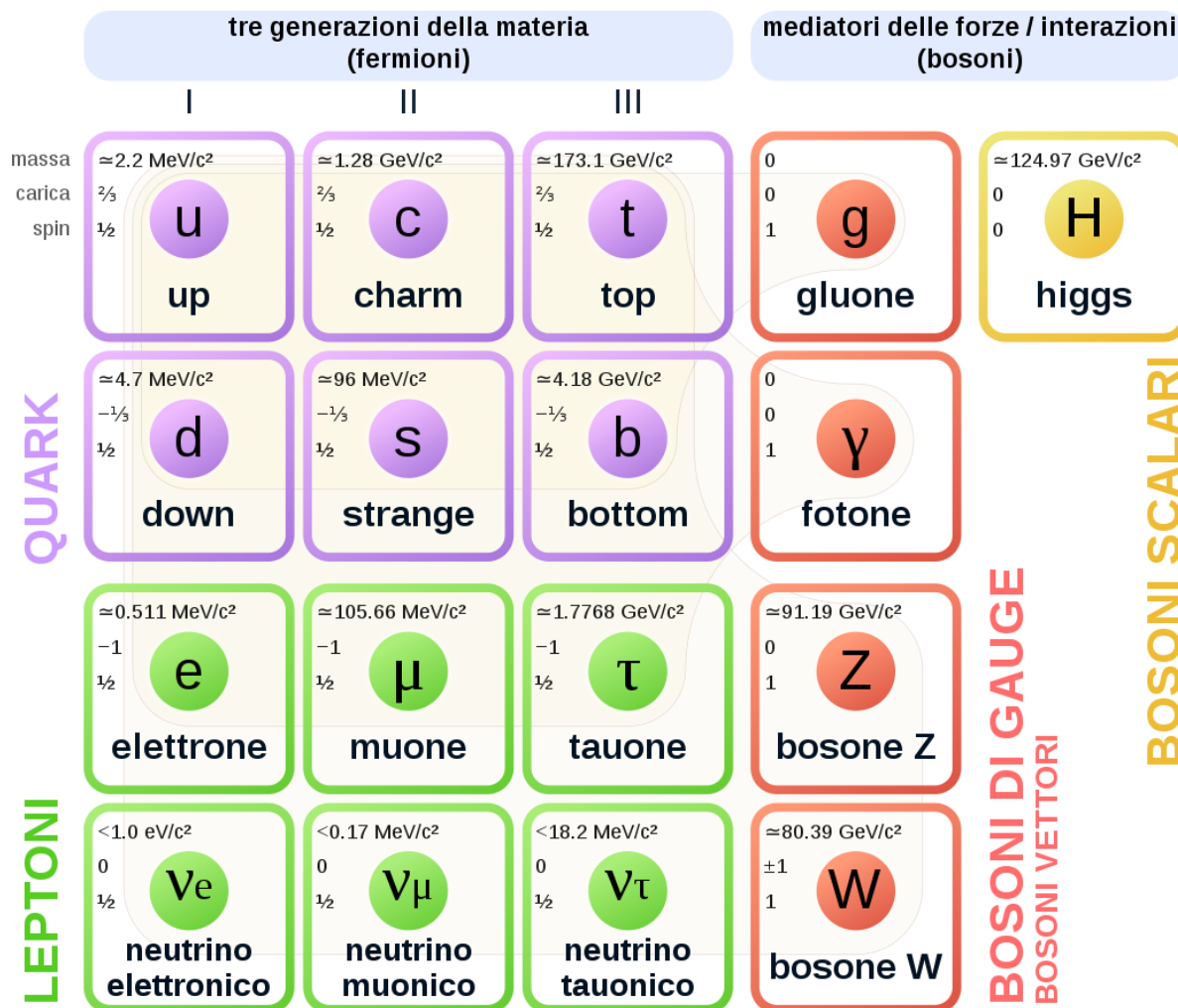
Correlatori: Dott. Ruggero Turra, Dott. Laura Nasella

Anno accademico 2024/2025



Contesto Teorico

Modello Standard delle Particelle Elementari



Bosoni

Mediatori delle interazioni fondamentali:

- γ : interazione elettromagnetica
- g: interazione forte
- W/Z: interazione debole
- H: conferisce massa

Fermioni

- Costituiscono la materia
- Spin $\frac{1}{2}$
- Accompagnati dalla propria antiparticella
- Divisi in Leptoni e Quark

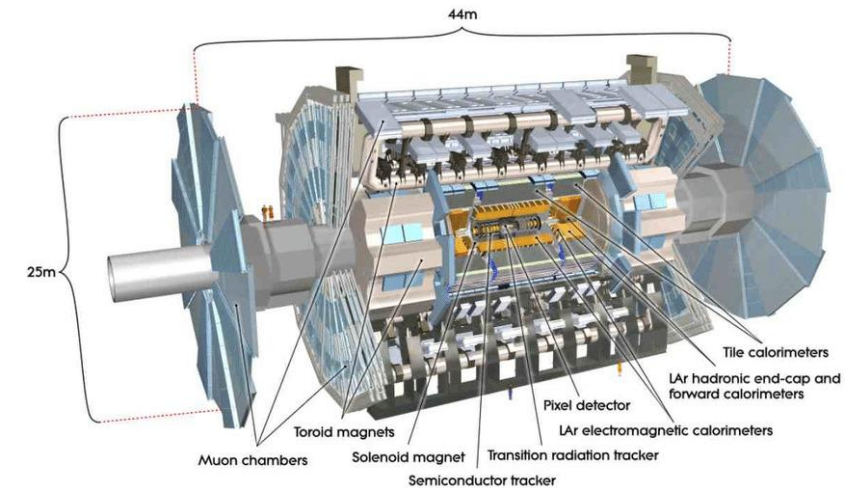


LHC

- Uno dei 7 acceleratori del CERN
- Collisioni Protone-Protone a $\sqrt{s} = 13,6$ TeV ➡ Studio processi fisici descritti da SM
- Avvengono in presenza dei rivelatori

ATLAS

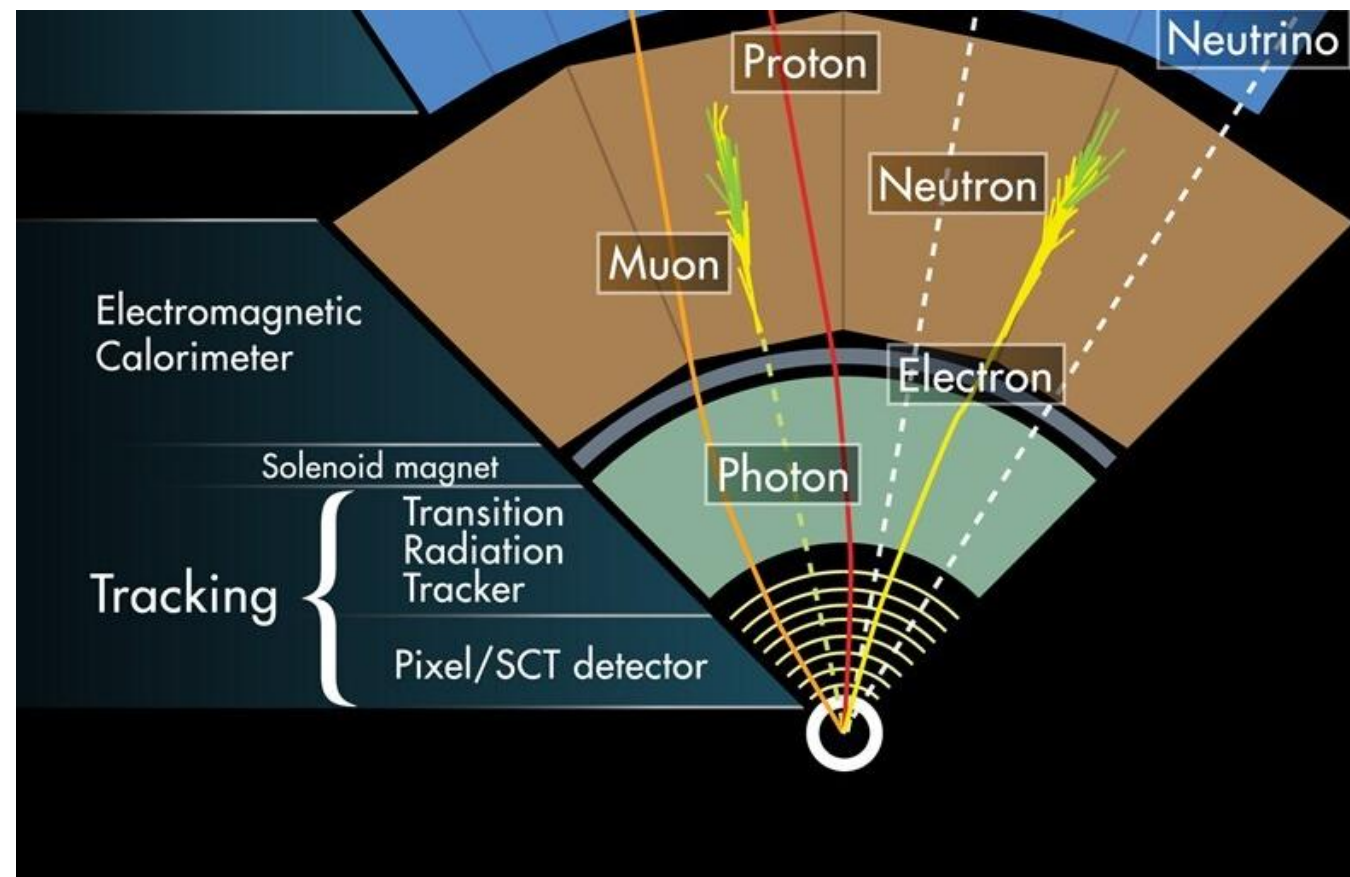
- Le collisioni avvengono al centro e le particelle si propagano in tutte le direzioni
- Sistema di rivelatori
 - Inner detector: momento e traiettoria delle particelle,
 - Calorimetri: energia delle particelle
 - Muon Spectrometer: momento dei muoni





Ricostruzione di Elettroni e Fotoni

- Ricostruzione di depositi energetici nelle celle dei calorimetri
- Ricostruzione delle tracce
- Abbinamento tracce o vertici di conversione con cluster di energia
- e^- :
 - definito come un oggetto costituito da un cluster con una traccia associata
- γ può essere di due tipi :
 - **Fotone convertito**: cluster associato ad un vertice di conversione
 - **Fotone non convertito**: cluster non associato né ad una traccia, né ad un vertice di conversione.





Obiettivi:

Analisi del decadimento: $Z \rightarrow l^+ l^- \gamma$

1. Misurare l'efficienza di Identificazione del fotone: $\epsilon = \frac{Y_{ield}^{riv}}{Y_{ield}^{tot}}$
2. Misura degli Scale Factors: $SF = \frac{\epsilon_{data}}{\epsilon_{MC}}$

Dati utilizzati (RUN2)

- merged_data_final.root
- merged_sig_final.root
- merged_bkg_final.root



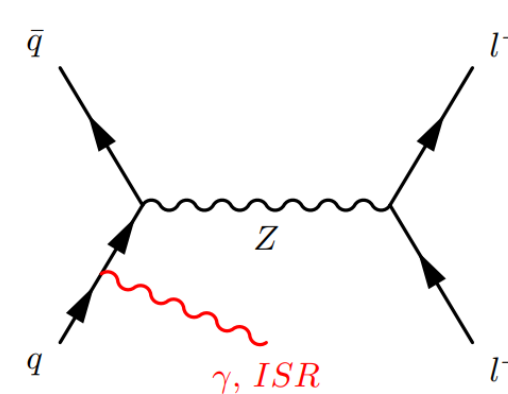
Selezioni applicate ai dataset

- $P_T^Y > 10 \text{ GeV}$
- $|\eta_\gamma| \in [0, 1.37] \cup [1.52, 2.37]$
- $\text{ph.conflag} = 3$
- $m_{ll} < 85 \text{ GeV}$
- $m_{ll\gamma} \in [50, 120] \text{ GeV}$



PROBLEMA:

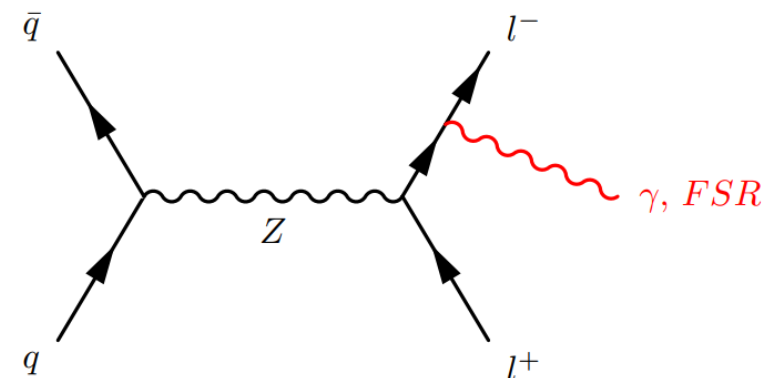
Rimane una frazione di background che bisogna riuscire a rimuovere per effettuare la misura dell'efficienza



Caso ISR:

$$m_{ll} \approx m_Z \approx 91 \text{ GeV}$$

$$m_{ll\gamma} > m_Z$$



Caso FSR:

$$m_{ll} < m_Z \approx 91 \text{ GeV}$$

$$m_{ll\gamma} \approx m_Z$$

Tipologia di
 $Z \rightarrow l^+ l^- \gamma$



Signal Region e Control Region

Definite tramite le seguenti variabili booleane:

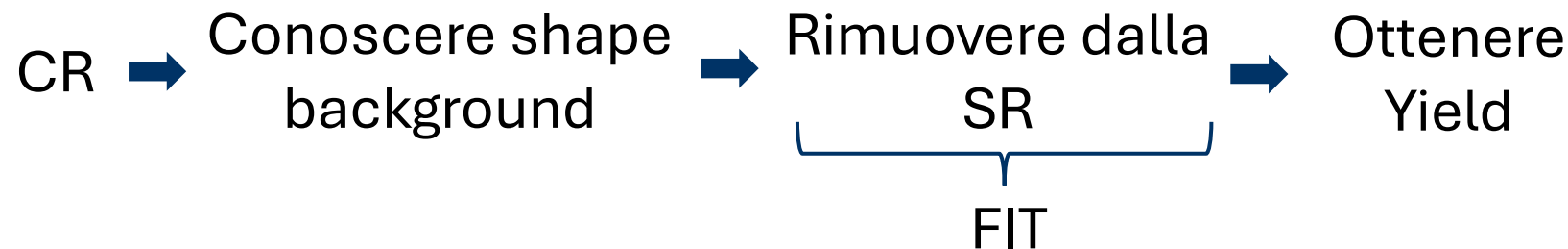
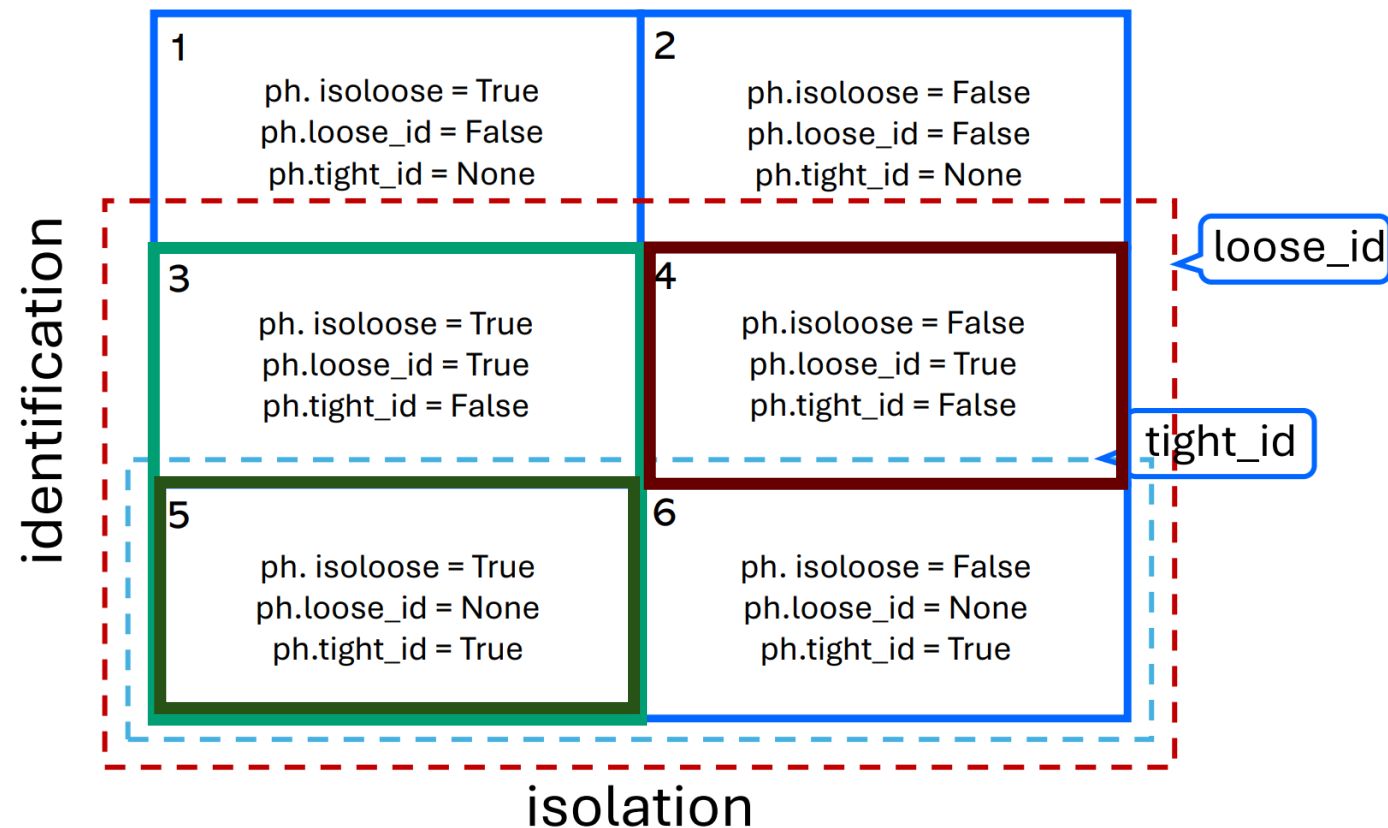
- ph.isoloose
- ph.loose_id
- ph.tight_id

Signal Region:

- Numeratore:
 - `ph.tight_id = True`
 - `ph.isoloose = True`
- Denominatore:
 - `ph.isoloose = True`
 - `ph.loose_id = True`

Control Region:

- `ph.tight_id = False`
- `ph.isoloose = False`
- `ph.loose_id = True`





Fit simultanei in SR e CR

- Confronto tra:

- bkg MC in SR

- bkg MC in CR e

- dati nella CR (con il leakage rimosso)

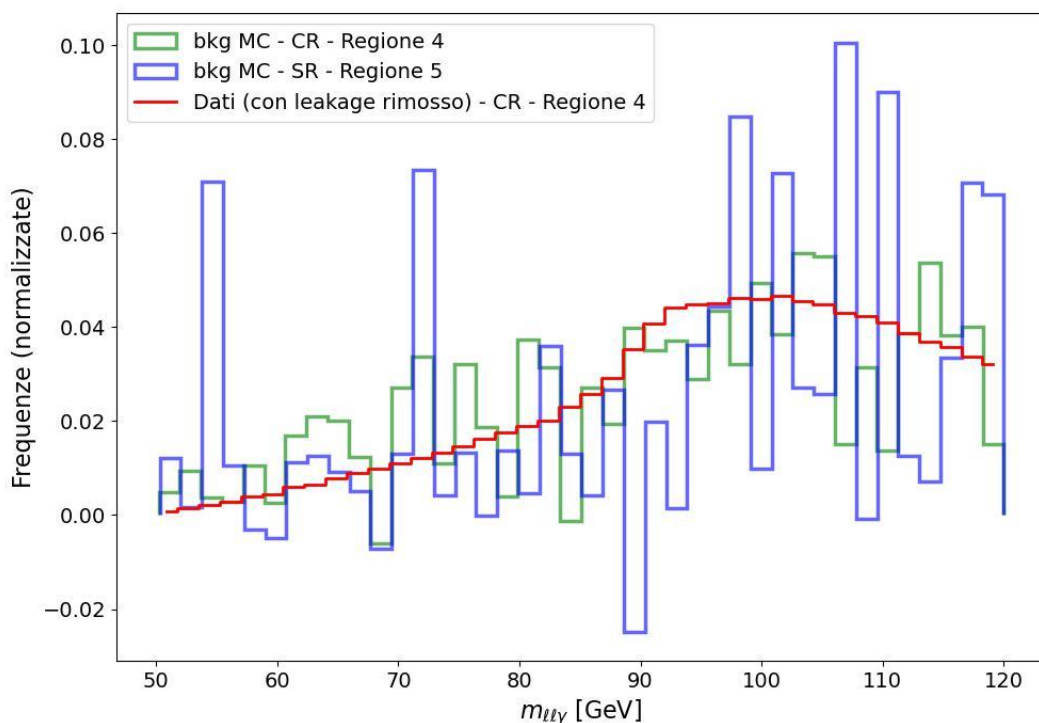


dati in CR descrivano
il bkg in SR

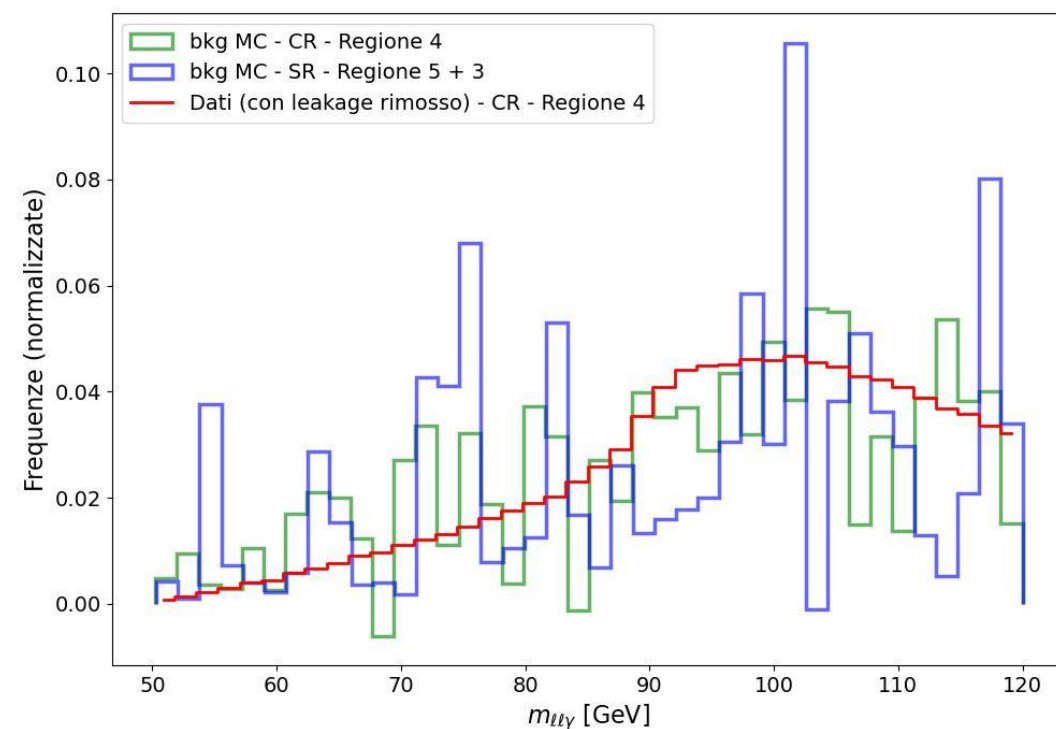


Utilizzato come
background nel Fit

Numeratore



Denominatore





Fit simultanei in SR e CR

Binning in η e P_T

$\eta = (0, 0.6), (0.6, 1.37), (1.52, 1.81), (1.81, 2.37)$

$P_T = (10, 15), (15, 20), (20, 25), (25, 30), (30, 35), (35, 40), (40, 50), (50, 70), (70, 150)$ GeV

Modello statistico SR:

$$N_{SR,i}^{model} = \mu_{SR} \cdot \underbrace{N_{sig,SR,i}^{MC}}_1 + \kappa_{bkg} \cdot \gamma_i \cdot \underbrace{N_{CR,i}^{data}}_3$$

Modello statistico CR:

$$N_{CR,i}^{model} = \mu_{CR} \cdot \underbrace{N_{sig,CR,i}^{MC}}_2 + \gamma_i \cdot \underbrace{N_{CR,i}^{data}}_3$$

- Channels: SR e CR
- Samples: segnale e background
- Modifiers:
 - Normfactor: μ_{SR}, μ_{CR} e κ_{bkg}
 - Shapefactor: γ_i

Istogrammi importati:

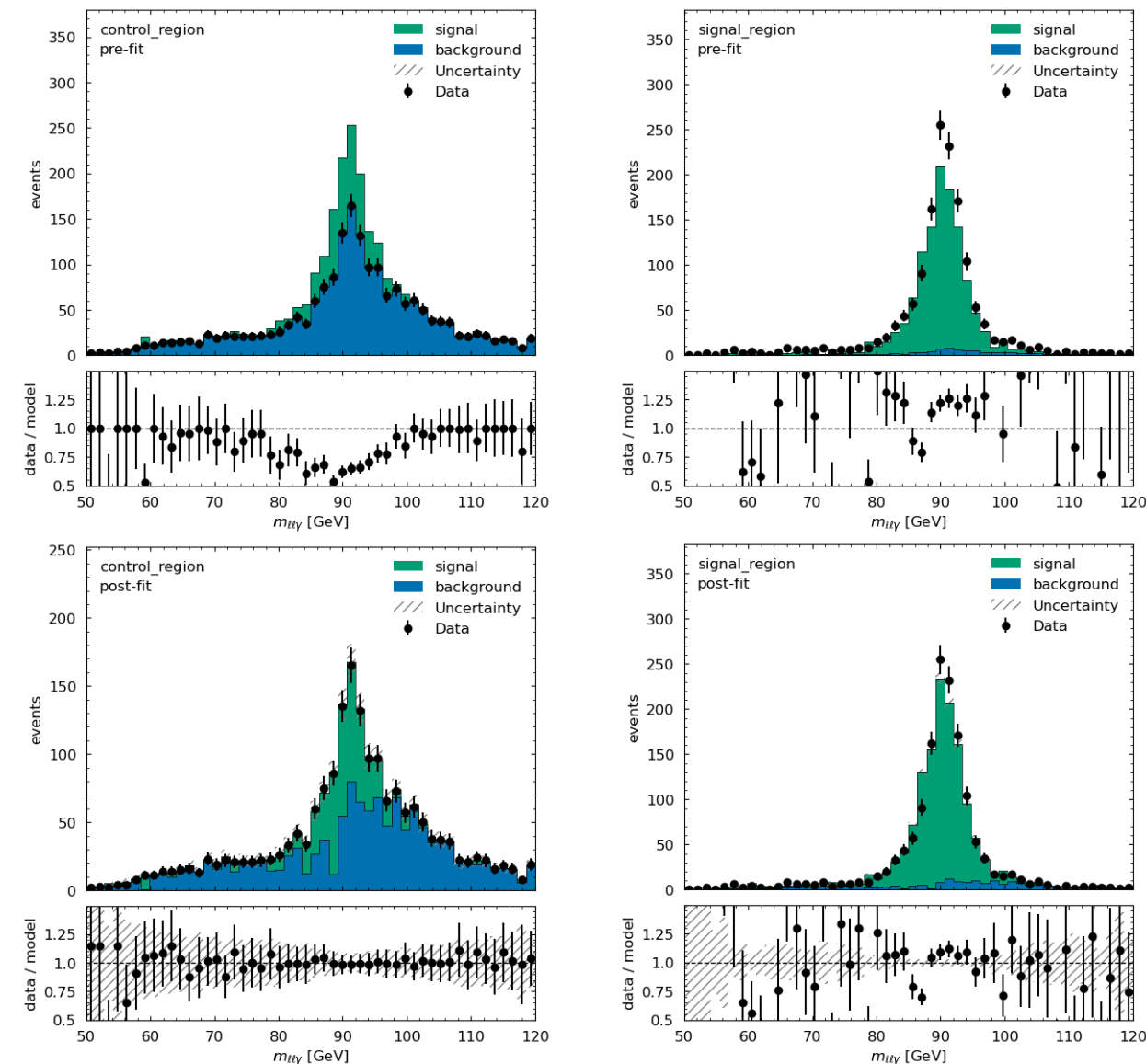
- $N_{sig,SR,i}^{MC}$: MC in SR ■ **1**
- $N_{sig,CR,i}^{MC}$: MC in CR ■ **2**
- $N_{CR,i}^{data}$: Dati in CR ■ **3**
- $N_{SR,i}^{data}$: Dati in SR ■ **4**

Confronto tra Data e Model:

$$4 \mid \frac{N_{SR,i}^{data}}{N_{SR,i}^{model}} \approx 1 \quad \Rightarrow \quad \text{Il modello descrive i dati}$$



Fit simultanei in SR e CR – esempio bin $\eta = [0, 0.6]$ e $P_T = [10, 15]$



PREFIT			
Canale	Campione	Yield	Errore
CR	Background	1855.0000	± 43.0697
CR	Segnale	581.2748	± 24.1096
SR	Background	92.7500	± 9.6306
SR	Segnale	1138.8173	± 33.7463

POSTFIT			
Canale	Campione	Yield	Errore
CR	Background	1275.0542	± 43.0632
CR	Segnale	581.2748	± 24.1096
SR	Background	187.9898	± 23.3245
SR	Segnale	1269.1621	± 40.3163

$$\mu_{SR} = 1.1145 \pm 0.0354$$

$$\kappa_{bkg} = 0.1474 \pm 0.0186$$



Efficienza dei dati

- Calcolo Efficienza: $\epsilon_{data,i} = \frac{Y_{ield,i}^{num}}{Y_{ield,i}^{den}}$
- Errore statistico attraverso metodo di Clopper-Pearson $\rightarrow \sigma_{data,i}^{stat}$

Errori sistematici:

- $m_{ll\gamma} \in [50, 120] \text{ GeV} \rightarrow m_{ll\gamma} \in [55, 115] \text{ GeV} \rightarrow \epsilon_{data}^{int} \rightarrow \sigma_{sist,i}^{int} = |\epsilon_{data,i}^{int} - \epsilon_{data,i}|$
- Regione 4 come CR $\rightarrow$ Regione 6 come CR $\rightarrow \epsilon_{data}^{CR} \rightarrow \sigma_{sist,i}^{CR} = |\epsilon_{data,i}^{CR} - \epsilon_{data,i}|$

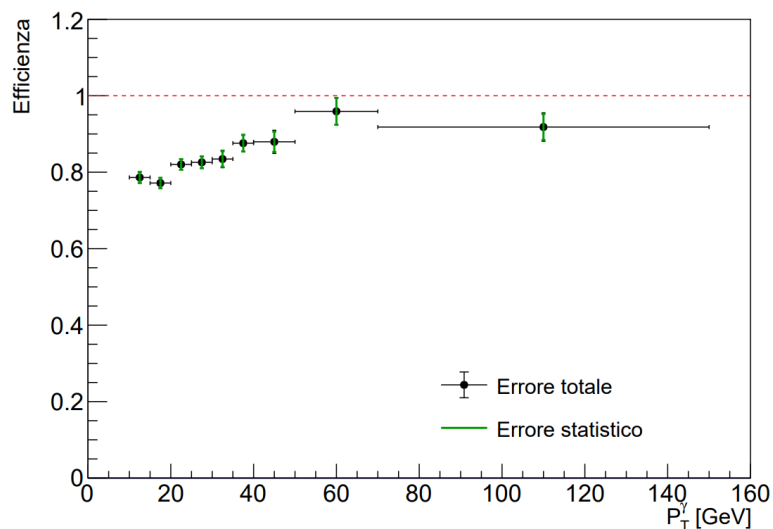
$$\sigma_{data,i}^{sist} = \sqrt{(\sigma_{sist,i}^{int})^2 + (\sigma_{sist,i}^{CR})^2}$$

$$\rightarrow \sigma_{data,i}^{tot} = \sqrt{(\sigma_{data,i}^{sist})^2 + (\sigma_{data,i}^{stat})^2}$$

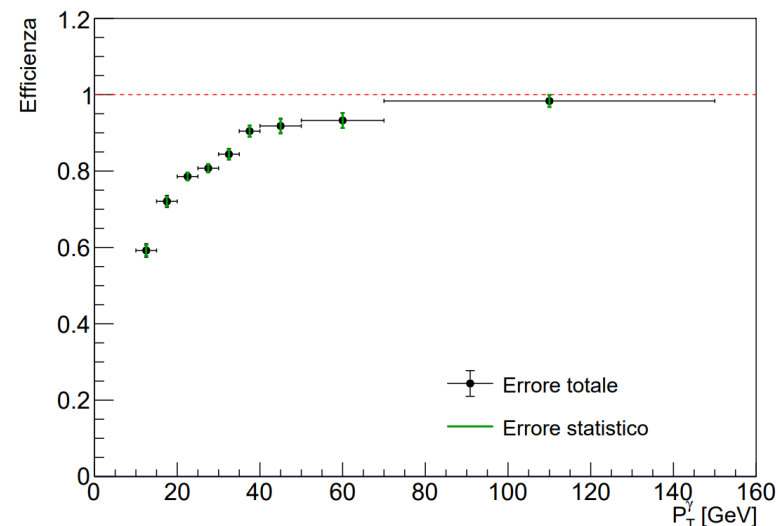


Efficienza dei dati

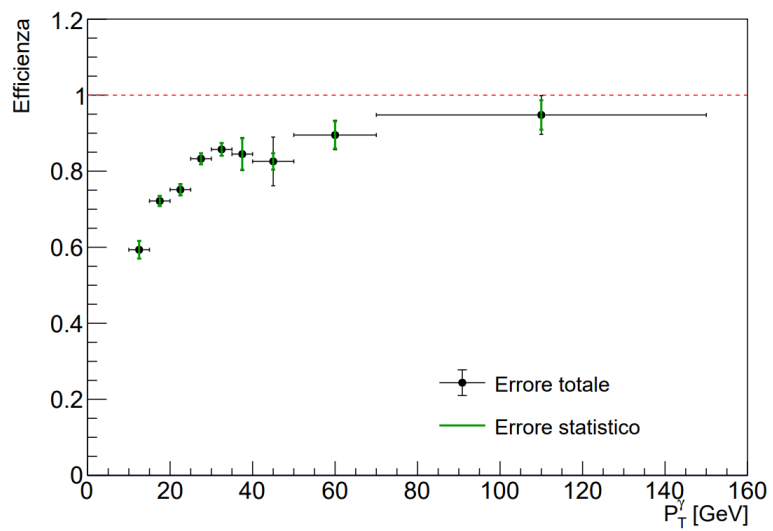
Efficienza dei dati dal fit per $\eta \in [0 - 0.6]$



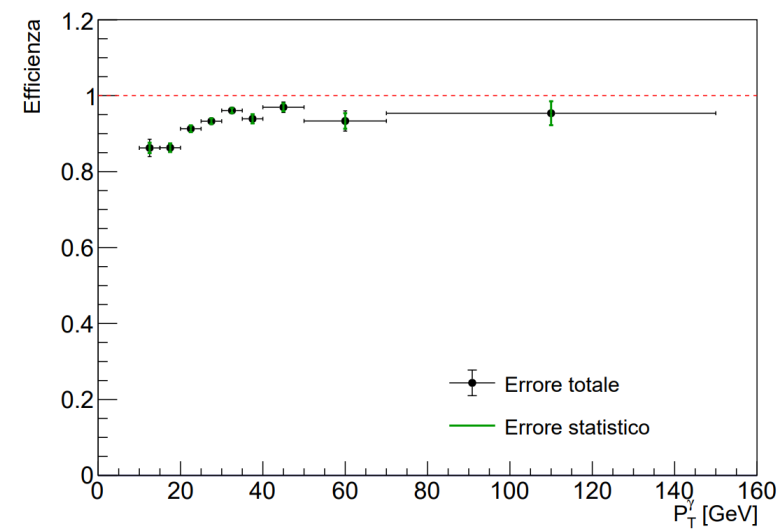
Efficienza dei dati dal fit per $\eta \in [0.6 - 1.37]$



Efficienza dei dati dal fit per $\eta \in [1.52 - 1.81]$



Efficienza dei dati dal fit per $\eta \in [1.81 - 2.37]$





Efficienza del Monte Carlo

Eventi di decadimento simulati con MC



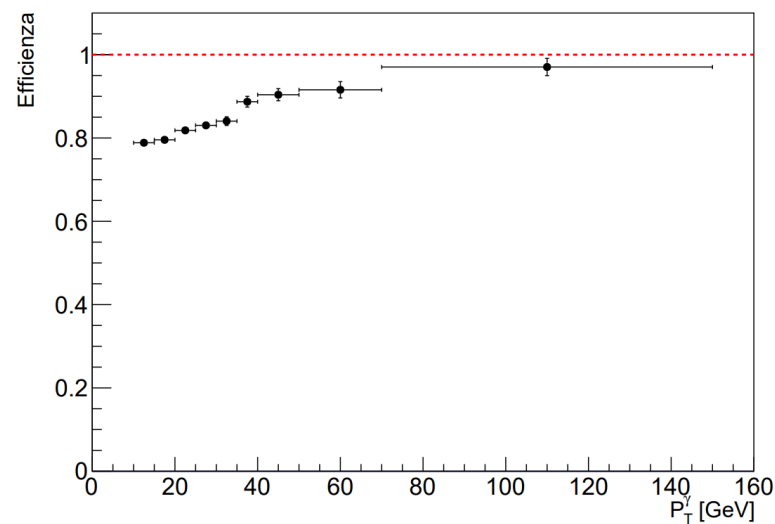
SR a numeratore: $Y_{ield,i}^{num,MC}$

SR a denominatore: $Y_{ield,i}^{den,MC}$

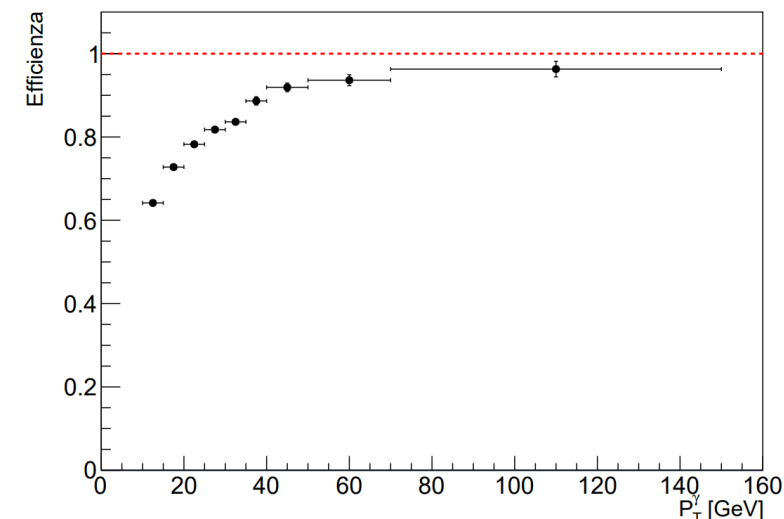
$$\epsilon_{MC,i} = \frac{Y_{ield,i}^{num,MC}}{Y_{ield,i}^{den,MC}}$$

→ $\sigma_{MC,i}^{stat}$ Clopper-Pearson

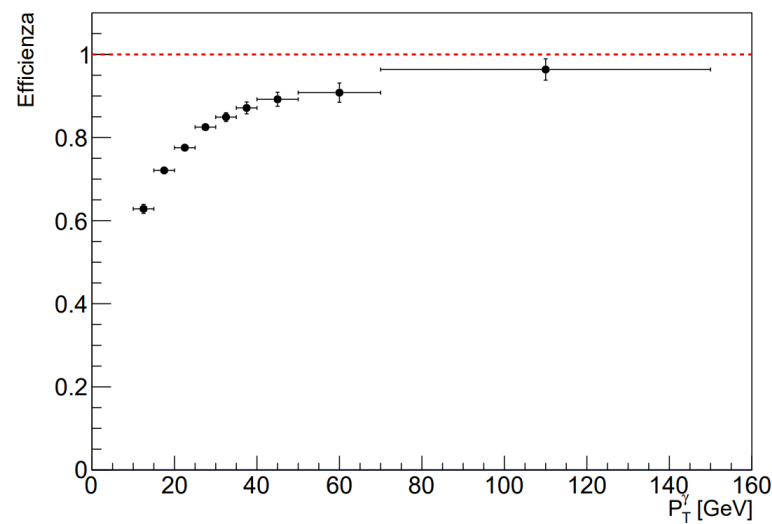
Efficienza di sgn MC per $\eta \in [0 - 0.6]$



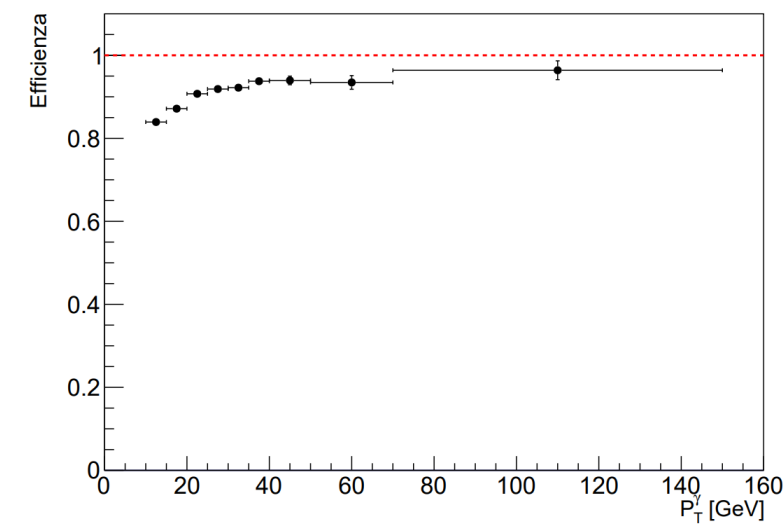
Efficienza di sgn MC per $\eta \in [0.6 - 1.37]$



Efficienza di sgn MC per $\eta \in [1.52 - 1.81]$



Efficienza di sgn MC per $\eta \in [1.81 - 2.37]$





Misura Scale Factors

$$SF_i = \frac{\epsilon_i^{data}}{\epsilon_i^{MC}}$$

Errori su SF:

$$\sigma_{data,i}^{stat,SF} = \frac{\sigma_{data,i}^{stat}}{\epsilon_{data,i}} \cdot SF_i$$

$$\sigma_{data,i}^{sist,SF} = \frac{\sigma_{data,i}^{sist}}{\epsilon_{data,i}} \cdot SF_i$$

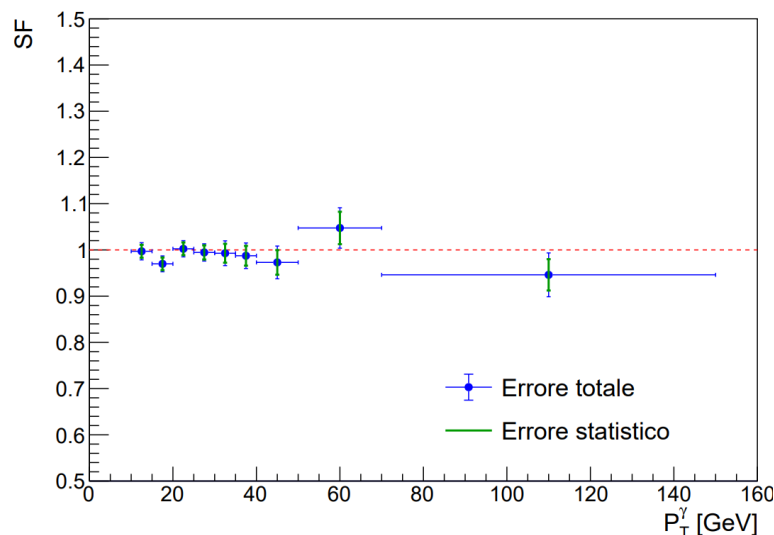
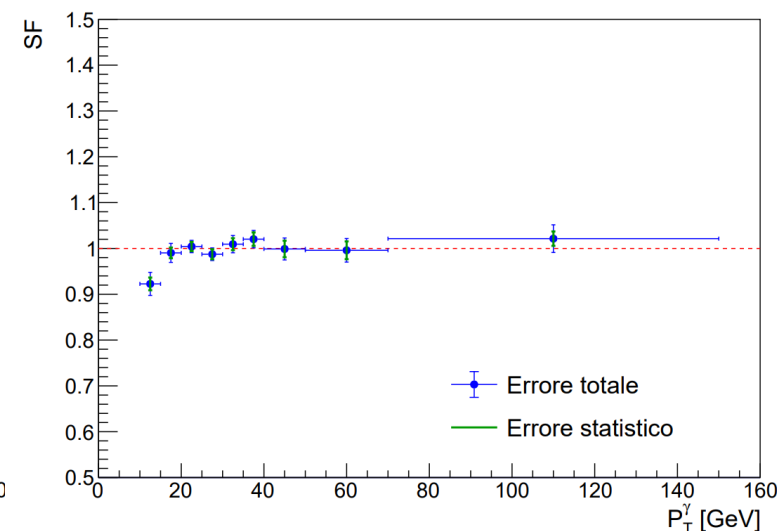
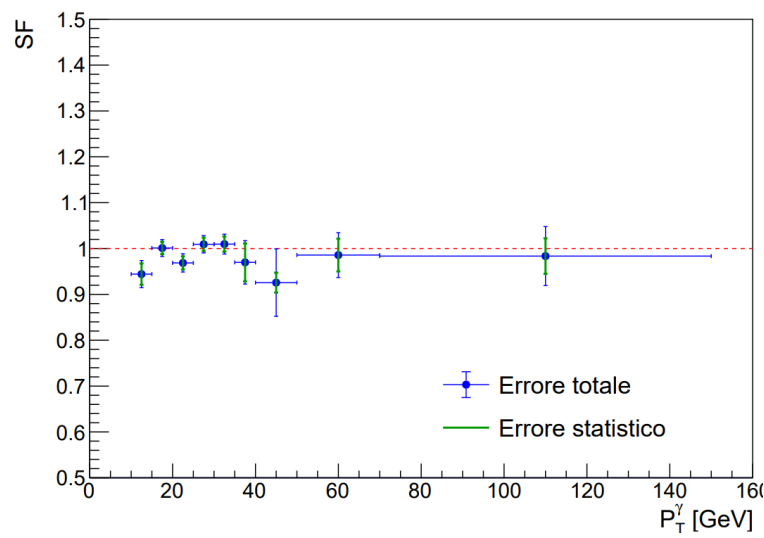
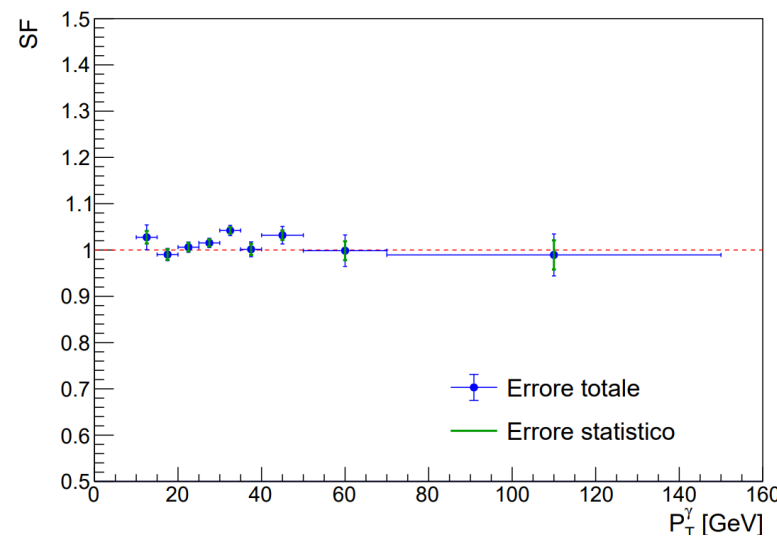
$$\sigma_{MC,i}^{stat,SF} = \frac{\sigma_{MC,i}^{stat}}{\epsilon_{MC,i}} \cdot SF_i$$

$$\sigma_{sist,i}^{SF} = \sqrt{(\sigma_{data,i}^{sist,SF})^2 + (\sigma_{MC,i}^{stat,SF})^2}$$

$$\sigma_{tot,i}^{SF} = \sqrt{(\sigma_{sist,i}^{SF})^2 + (\sigma_{data,i}^{stat,SF})^2}$$



Calcolo Scale Factors

Scale Factor per $\eta \in [0 - 0.6]$ Scale Factor per $\eta \in [0.6 - 1.37]$ Scale Factor per $\eta \in [1.52 - 1.81]$ Scale Factor per $\eta \in [1.81 - 2.37]$ 



Conclusioni

- Dati relativi al decadimento: $Z \rightarrow l^+ l^- \gamma$
- Misura ϵ_{data} e dell'errore composto dai contributi σ_{data}^{stat} e σ_{data}^{sist}
- Misura di SF con il proprio errore composto dai contributi:
 - valori compatibili con 1 entro $\sigma^{tot,SF}$
 - $\sigma^{tot,SF}$ varia da 0.66% - 7.75%
 - $\sigma^{tot,SF}$ composto da:
 - $\sigma^{stat,SF}$: dato dal σ_{data}^{stat} sull'efficienza, risulta dominante a basso P_T
 - $\sigma^{sist,SF}$: dato dal σ_{data}^{sist} e σ_{MC}^{stat} , risulta dominante a alto P_T

Possibili applicazioni:

- Analisi che coinvolgono fotoni nello stato finale
- Ricerca della verifica sperimentale di $HH \rightarrow \gamma\gamma b\bar{b}$, previsto da MS

Grazie per l'attenzione

Backup Slides

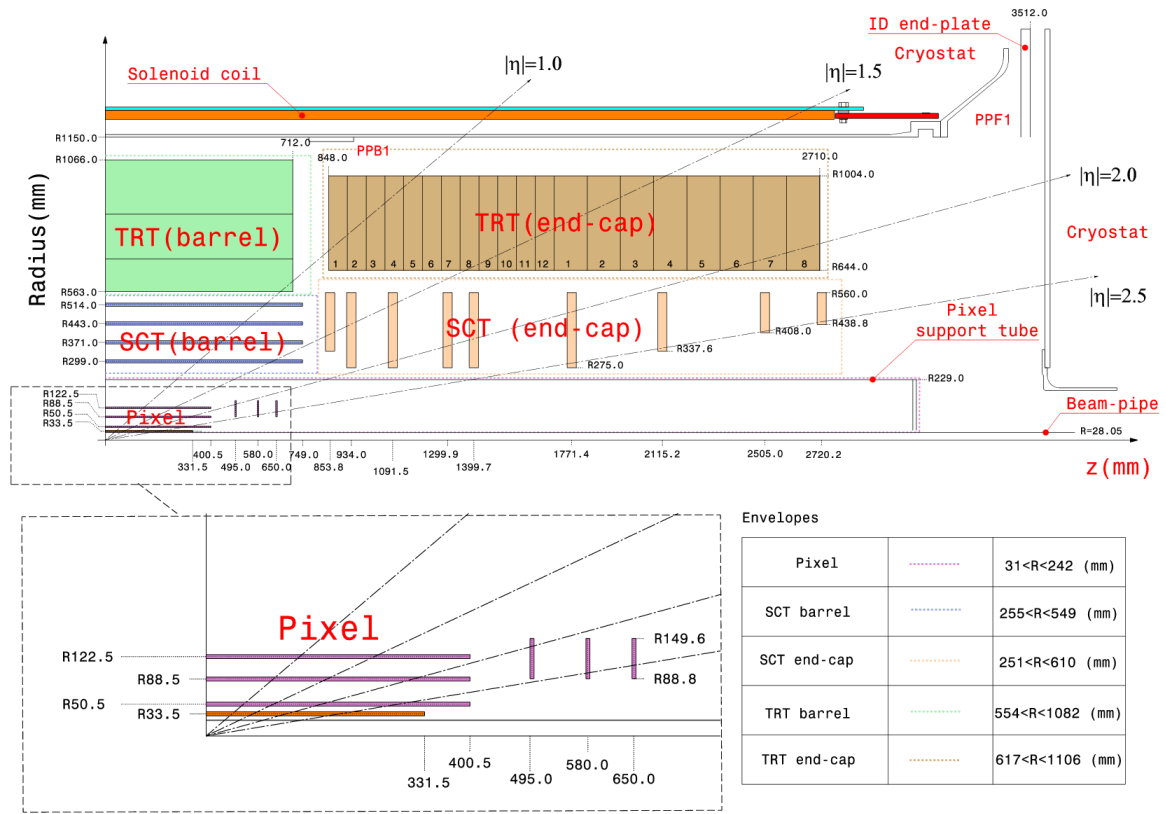
INDICE

- 1) Il large Hadron Collider - LHC
- 2) Il rivelatore ATLAS
- 3) Ricostruzione di Elettroni e Fotoni con il rivelatore ATLAS
- 4) Algoritmi di classificazione con Machine Learning
- 5) Classificazione di Elettroni e Fotoni
- 6) Conclusioni



Inner Detector

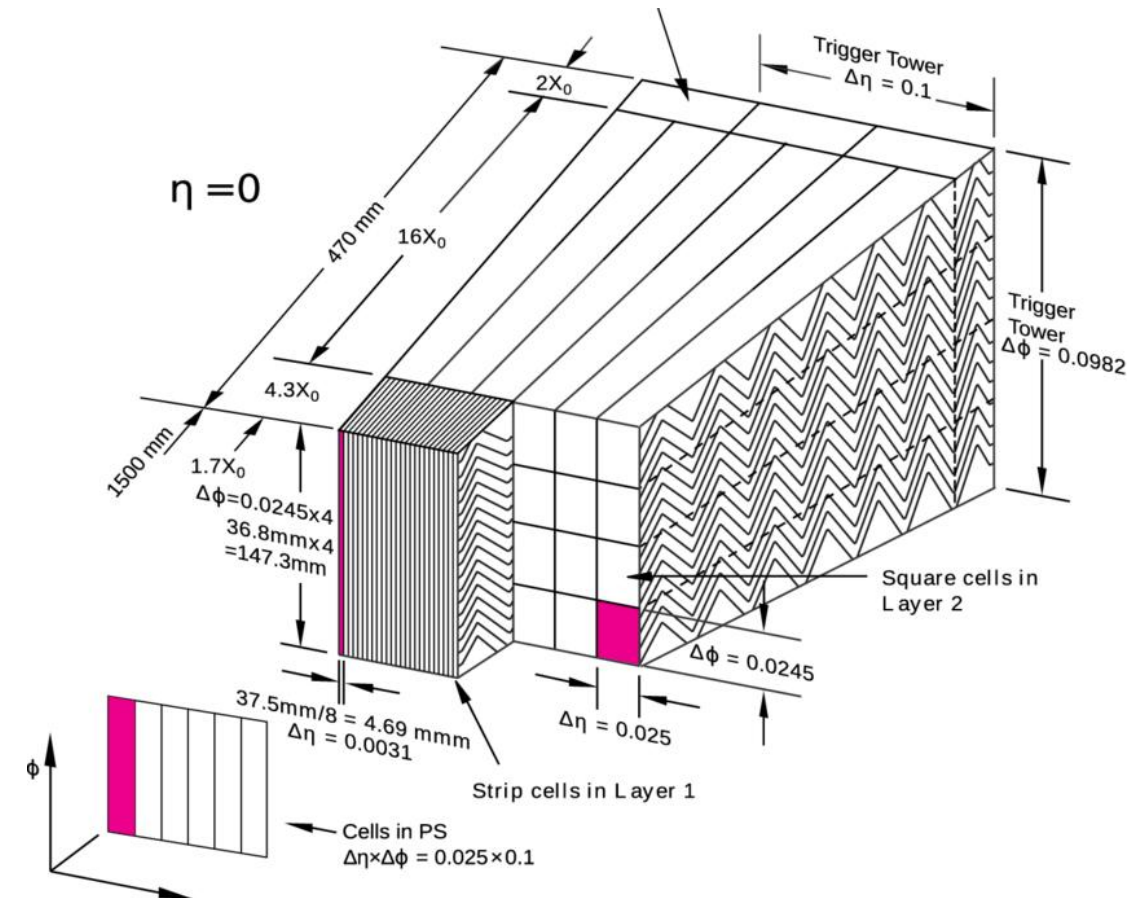
- Pixel:
 - Quando una particella carica passa attraverso il silicio, ionizza gli atomi del materiale, creando coppie elettrone-lacuna. Questi segnali elettrici vengono raccolti dai pixel e convertiti in dati digitali che rappresentano il percorso della particella
- SCT
 - Simile al Pixel Detector, le particelle ionizzano il silicio creando coppie elettrone-lacuna. Le strisce di silicio, disposte perpendicolarmente tra loro, raccolgono questi segnali per tracciare il percorso della particella.
- TRT
 - Quando una particella carica passa attraverso i tubi, provoca ionizzazione del gas. I segnali elettrici generati vengono raccolti dai fili centrali nei tubi. Inoltre, le particelle relativistiche producono radiazione di transizione quando attraversano i materiali di rivestimento, che può essere rilevata per distinguere tra particelle leggere e pesanti.



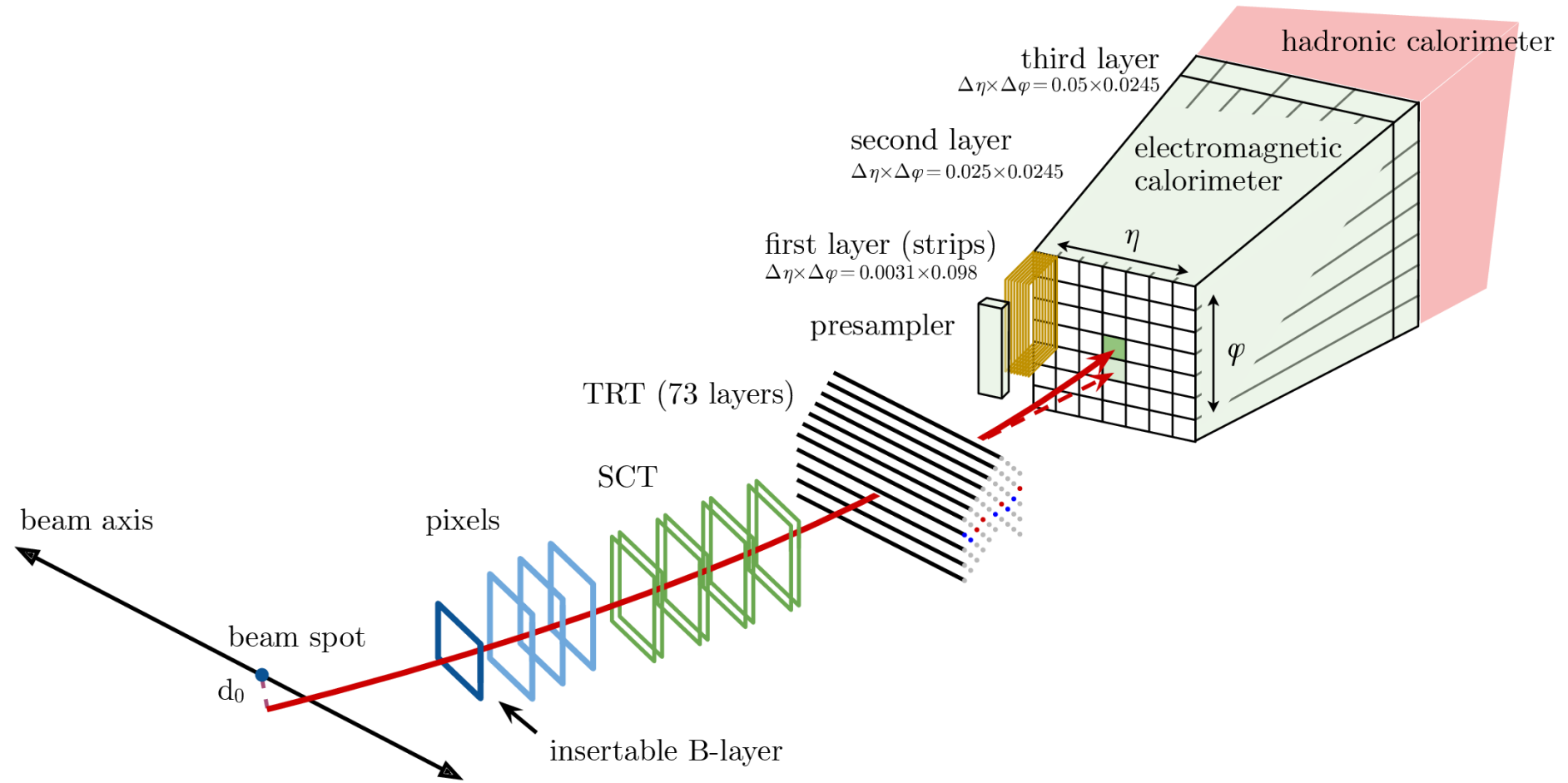
Rappresentazione schematica del Layout del Tracciatore Interno – Inner Detector

Calorimetro Elettromagnetico

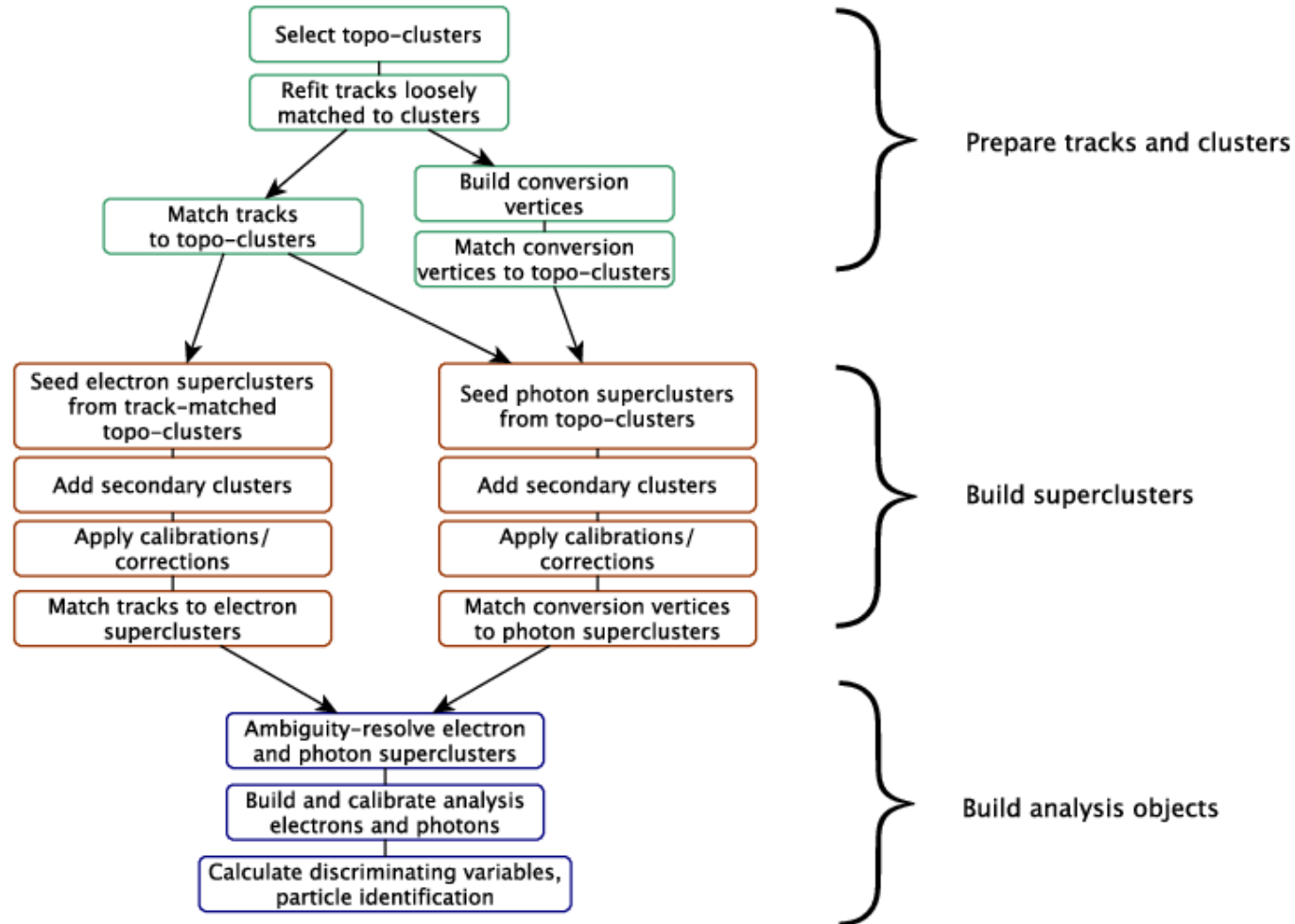
- Quando una particella elettromagnetica entra nel calorimetro, interagisce con il materiale assorbitore (Pb), producendo una cascata di particelle secondarie.
- Queste particelle secondarie ionizzano il materiale attivo, producendo segnali (luce o elettricità).
- Il materiale attivo è composto da argon liquido (LAr).
- La forma a fisarmonica consente una copertura quasi continua senza grandi spazi vuoti ("zone morte") tra i moduli del calorimetro



Ricostruzione di una Particella



Ricostruzione di Elettroni e Fotoni



- Ricostruzione: molto complicata.
- e^- definito come un oggetto costituito da un cluster con una traccia associata
- γ può essere di due tipi :

Fotone convertito: cluster associato ad un vertice di conversione

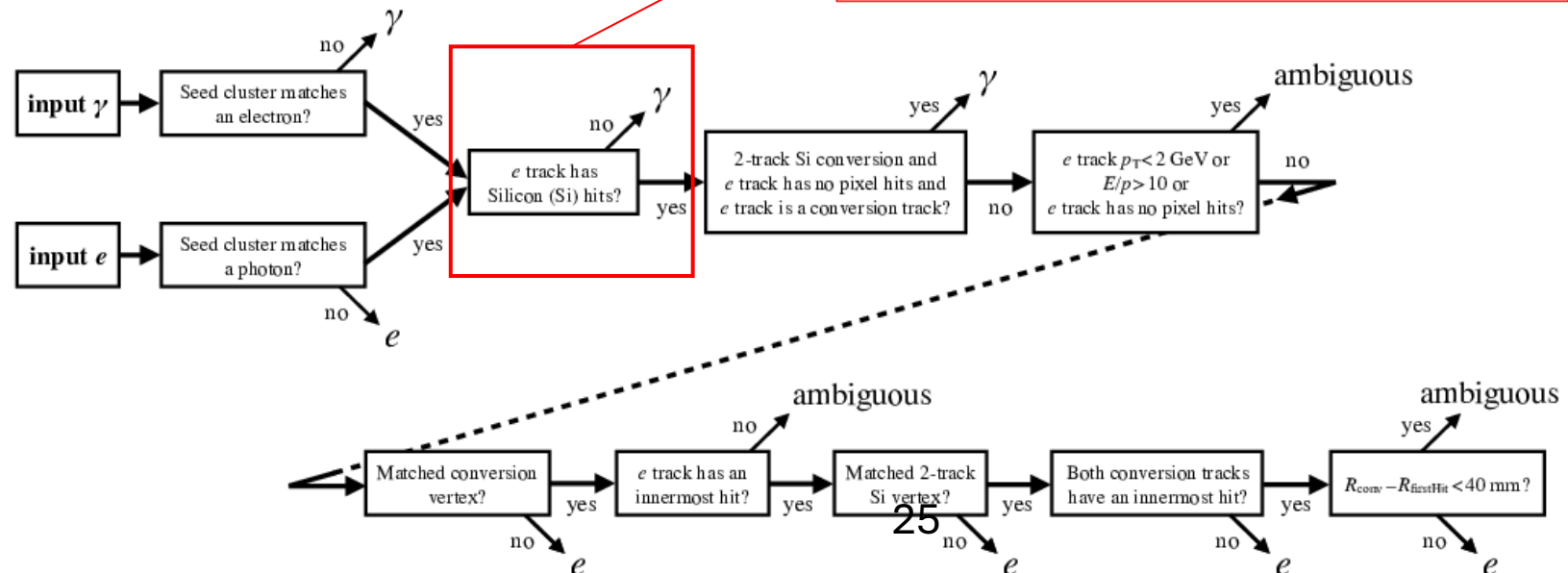
Fotone non convertito: cluster non associato né ad una traccia, né ad un vertice di conversione.

Ambiguity Resolver

- Dopo il rilevamento e l'abbinamento di tracce/ vertici di conversione e cluster di energia: come ricostruisco questi segnali?
- Algoritmo cut-based che classifica i candidate elettroni e fotoni come:
 - Elettroni
 - Fotoni
 - Ambigui, che vengono ricostruiti sia come elettroni che come fotoni

Se una particella non ha tracce nel rivelatore SCT dell'ID difficilmente sarà un elettrone

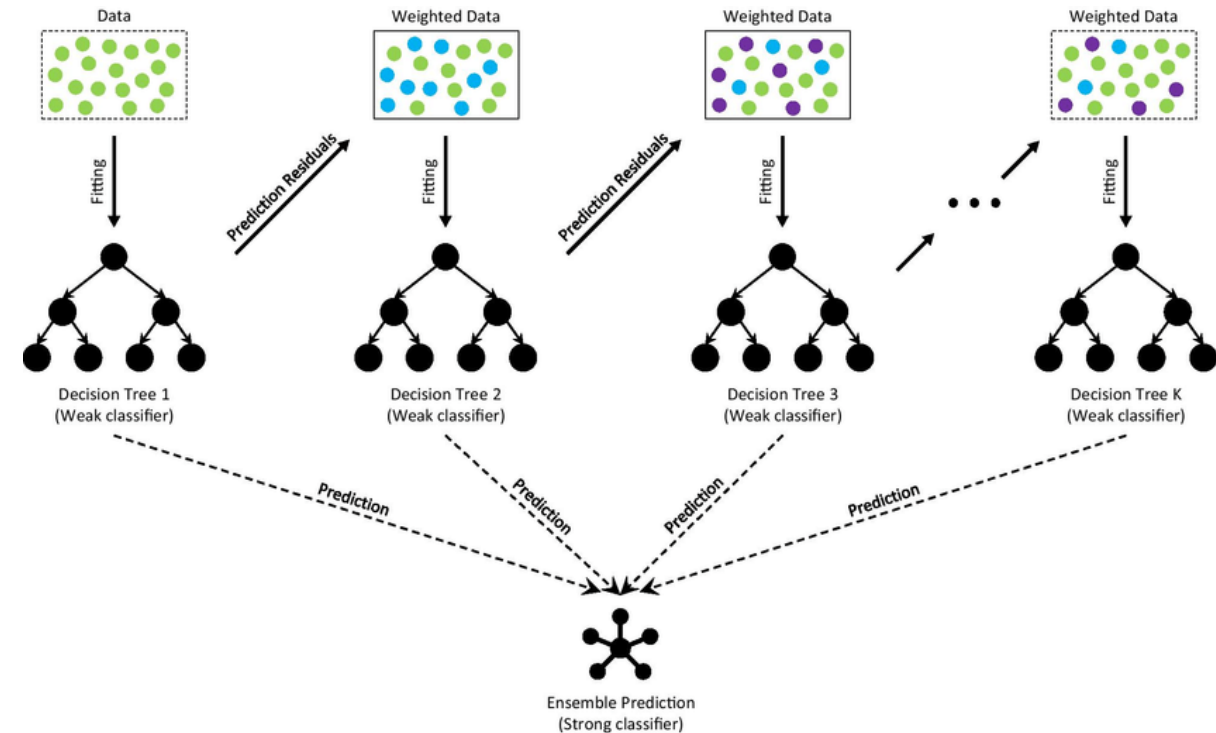
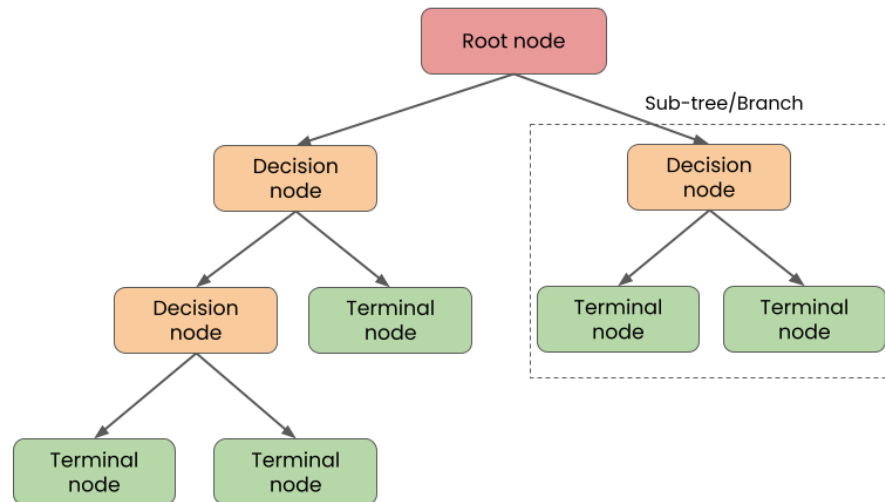
A questo livello, in un campione di Veri Fotoni si trova ~35% di ambigui.
In un campione di Veri Elettroni ~8.7%



Gradient Boosted Decision Tree (GBDT)

- Tecnica di Machine learning che prevede siano note:
 - Variabili di input (distintive)
 - Vero output (etichette)
- Supervised Learning**

$$\mathcal{D} = \{(x_i, y_i)\} \quad (|\mathcal{D}| = n, x_i \in \mathbb{R}^n, y_i \in \mathbb{R})$$



- Il GBDT si basa sul lavoro coordinato di un insieme di **alberi di decisione**

Gradient Boosted Decision Tree (GBDT)

- Cosa vuol dire **Gradient Boosted**?

Gli alberi decisionali non sono costruiti in maniera casuale ma seguendo il gradiente della **funzione obiettivo regolarizzata** $\mathcal{L}$

$$\mathcal{L} = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

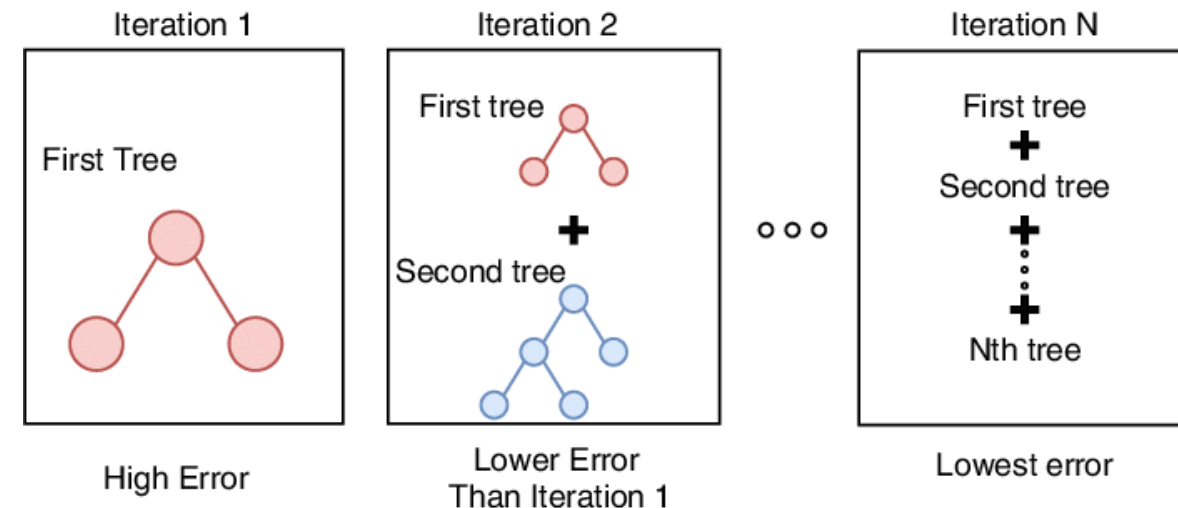
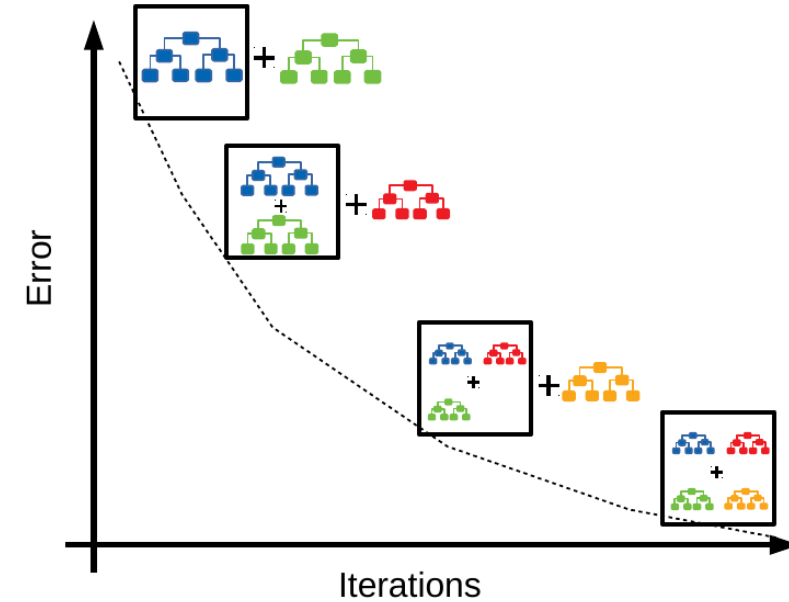
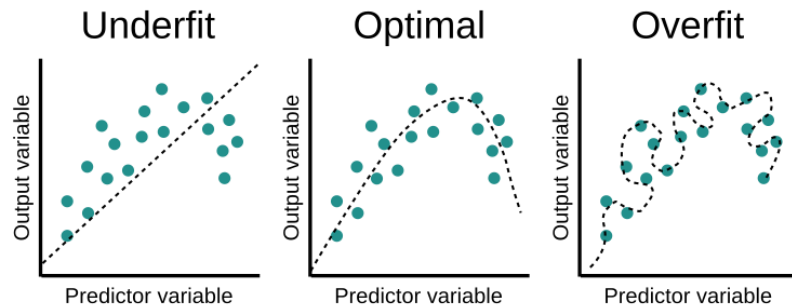
$$\Omega = \gamma T + \frac{1}{2} \lambda_k ||\omega_k||^2$$

Ω :

- Penalizza la complessità del modello
- Previene overfitting

$l(\hat{y}_i, y_i)$:

- *Loss function*
- Errore previsione-vero output ('residuo')





Cross Entropy

- Funzione di costo (*loss function*)
- Fornisce una quantificazione diretta della "distanza" tra le probabilità predette e quelle reali, permettendo ai modelli di apprendimento automatico di migliorare la loro accuratezza attraverso l'ottimizzazione iterativa durante il processo di addestramento.
- Se consideriamo due distribuzioni di probabilità P (la distribuzione vera) e Q (la distribuzione predetta dal modello), la cross entropy $H(P, Q)$ è definita come:

$$H(P, Q) = - \sum_x P(x) \log Q(x)$$

Classificazione di elettroni e fotoni

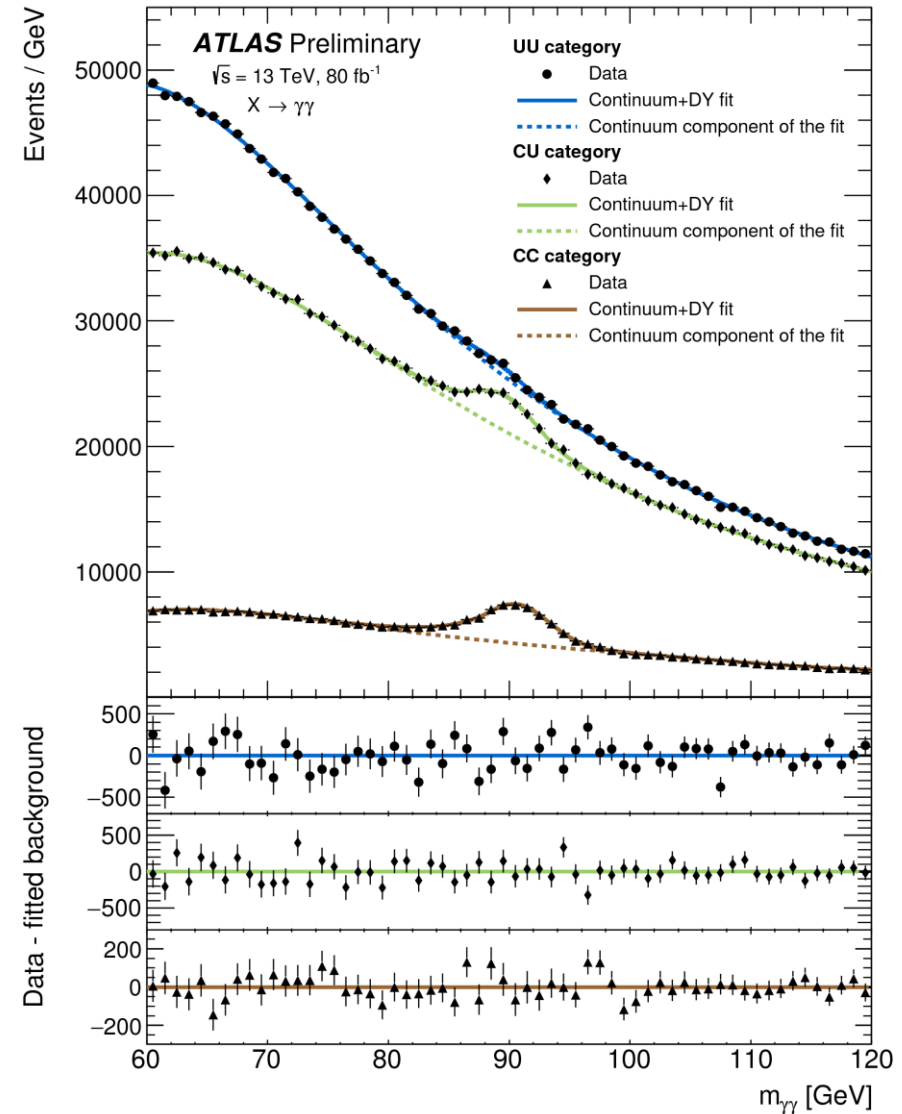
- **Perché è utile classificare elettroni e fotoni?**
 - Eccesso di eventi potenzialmente compatibili con $H \rightarrow \gamma \gamma$ a 95 GeV (bosone di Higgs addizionale)
 - Contaminazione da parte di $Z \rightarrow e^+ e^-$ a 90 GeV

→ Necessaria migliore classificazione

→ **Risposta della mia tesi:**

Gradient Boosted Decision Tree è uno strumento potenzialmente molto utile per rispondere a questo bisogno.

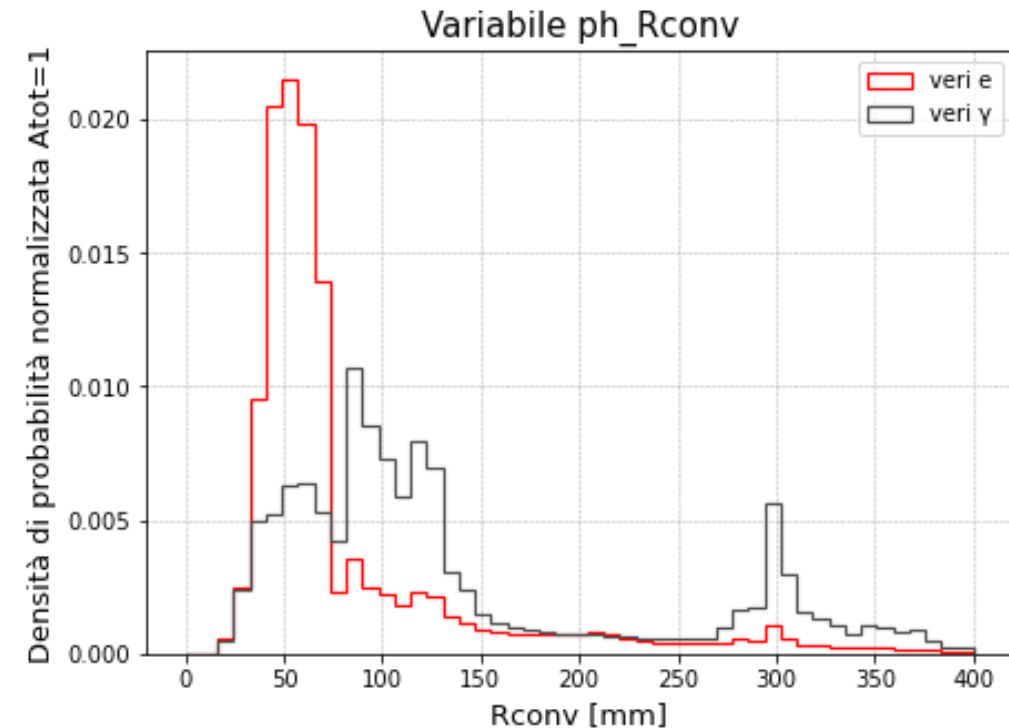
→ Applicazione dell'algoritmo su un campione di particelle simulate con metodi MonteCarlo (MC).



Ulteriore esempio di variabile discriminante

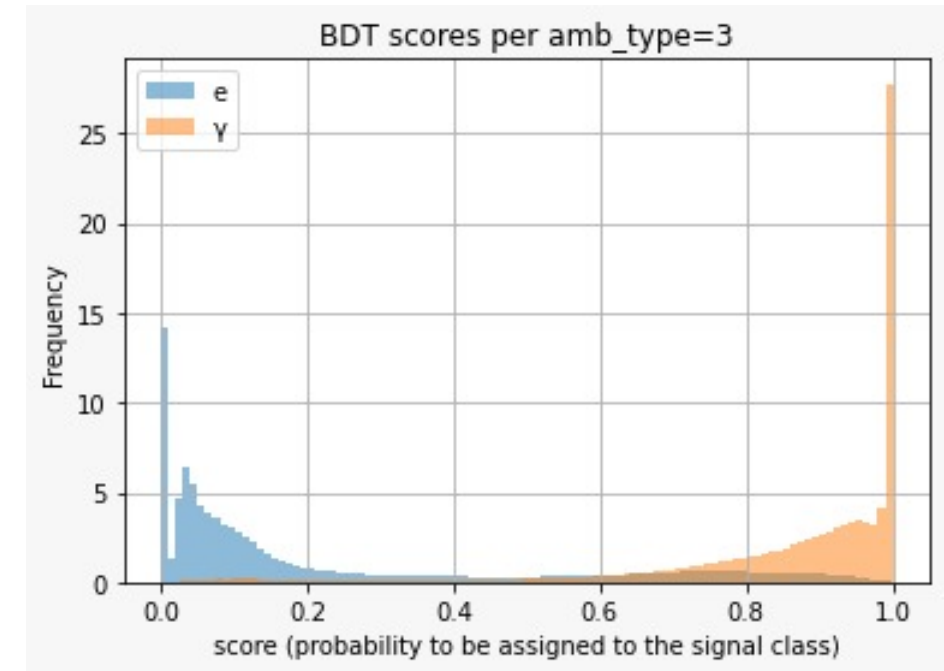
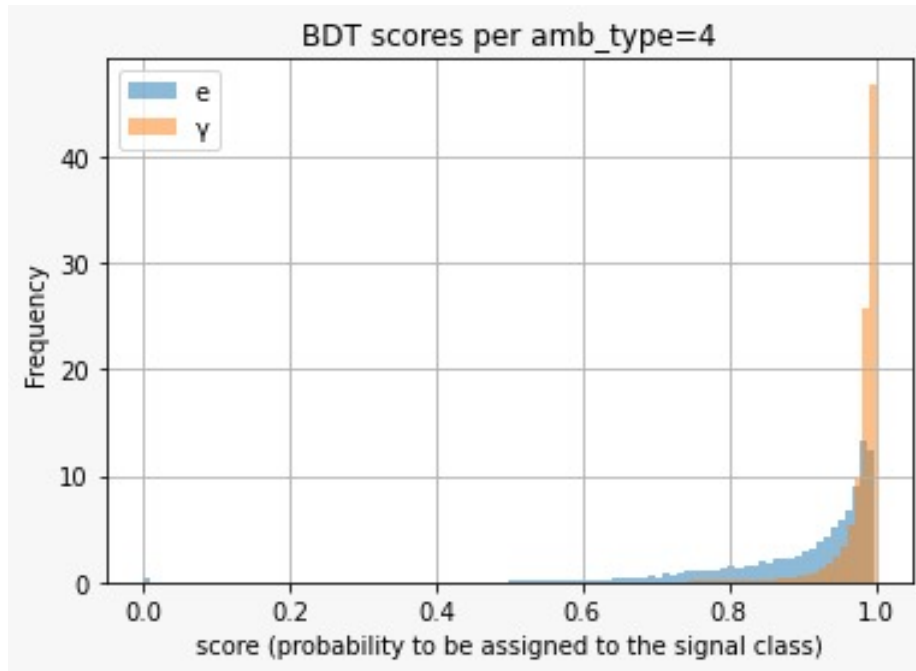
21 Variabili Discriminanti:

- Variabili generali (cl_eta, cl_pt)
- Variabili calcolate nella ricostruzione di elettroni
 - Hits nei rivelatori Pixel ed SCT
 - Qualità della traccia
- Variabili calcolate nella ricostruzione di fotoni
 - Variabili associate alla conversione
 - Hits e P_T delle tracce associate al fotone convertito

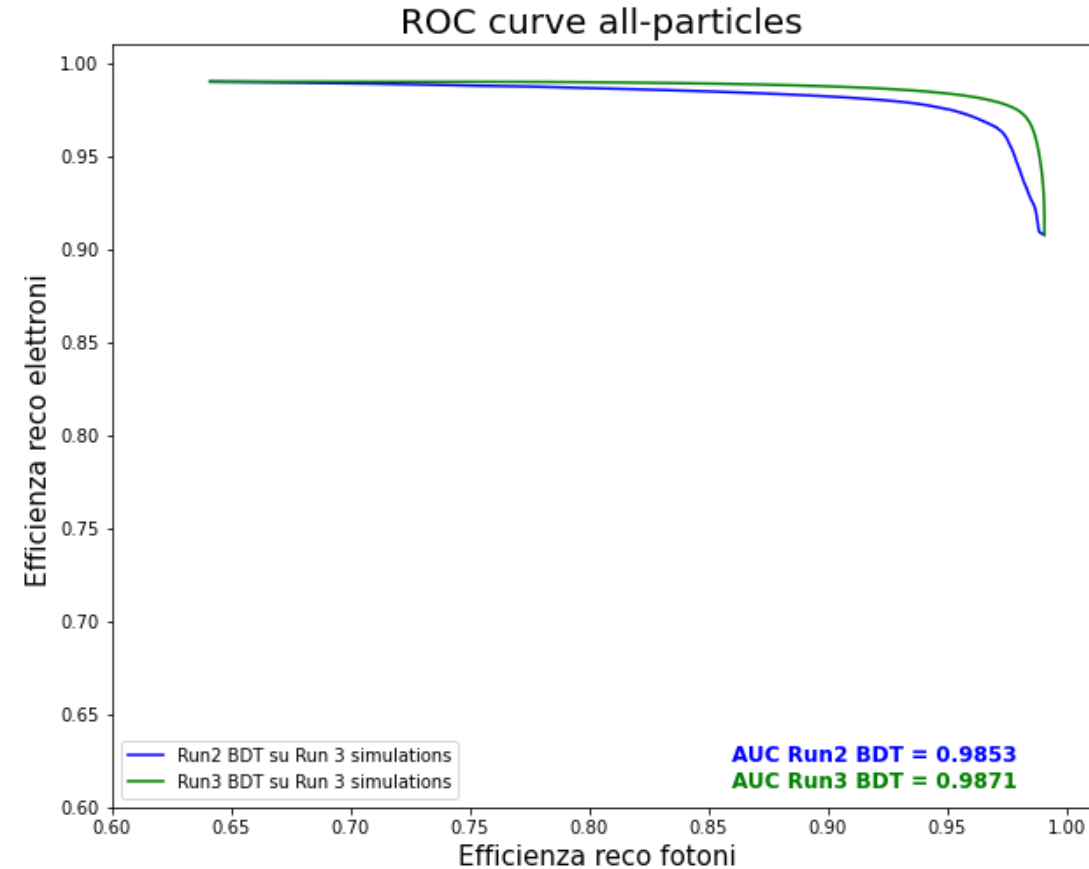
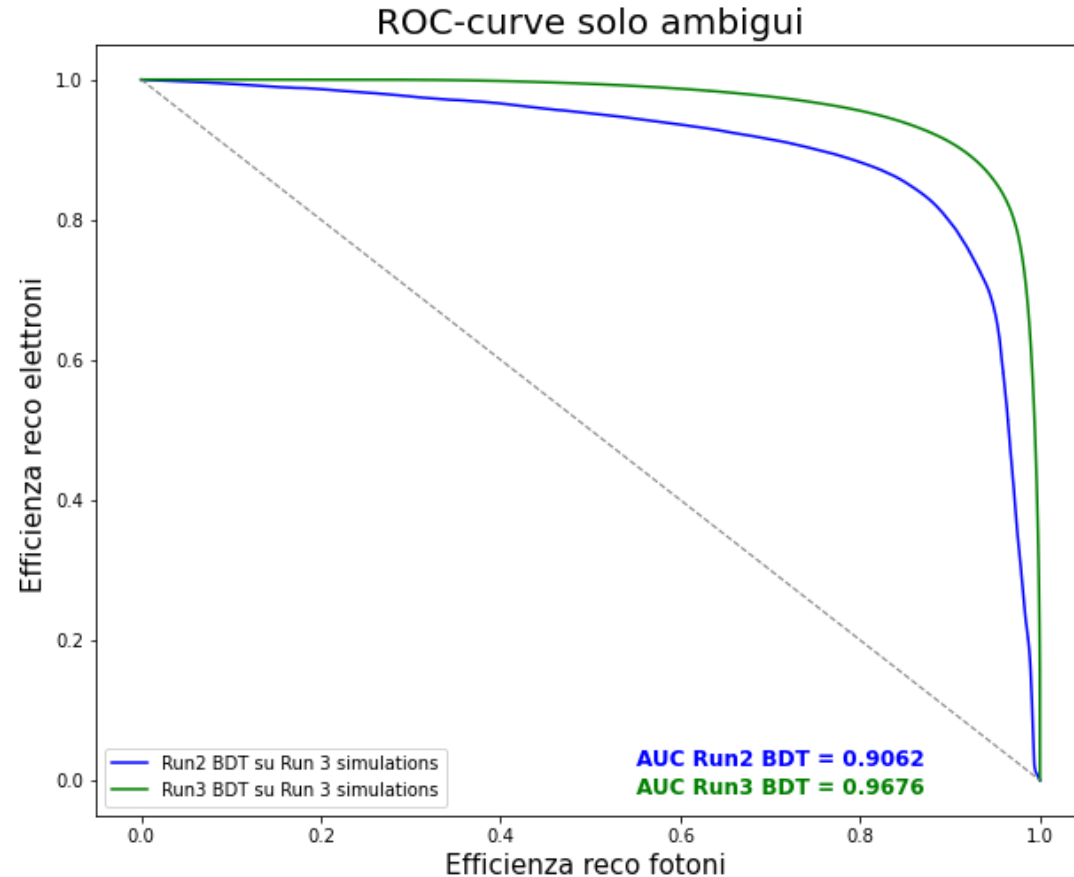


Posizione radiale della conversione

BDT score per amb_type



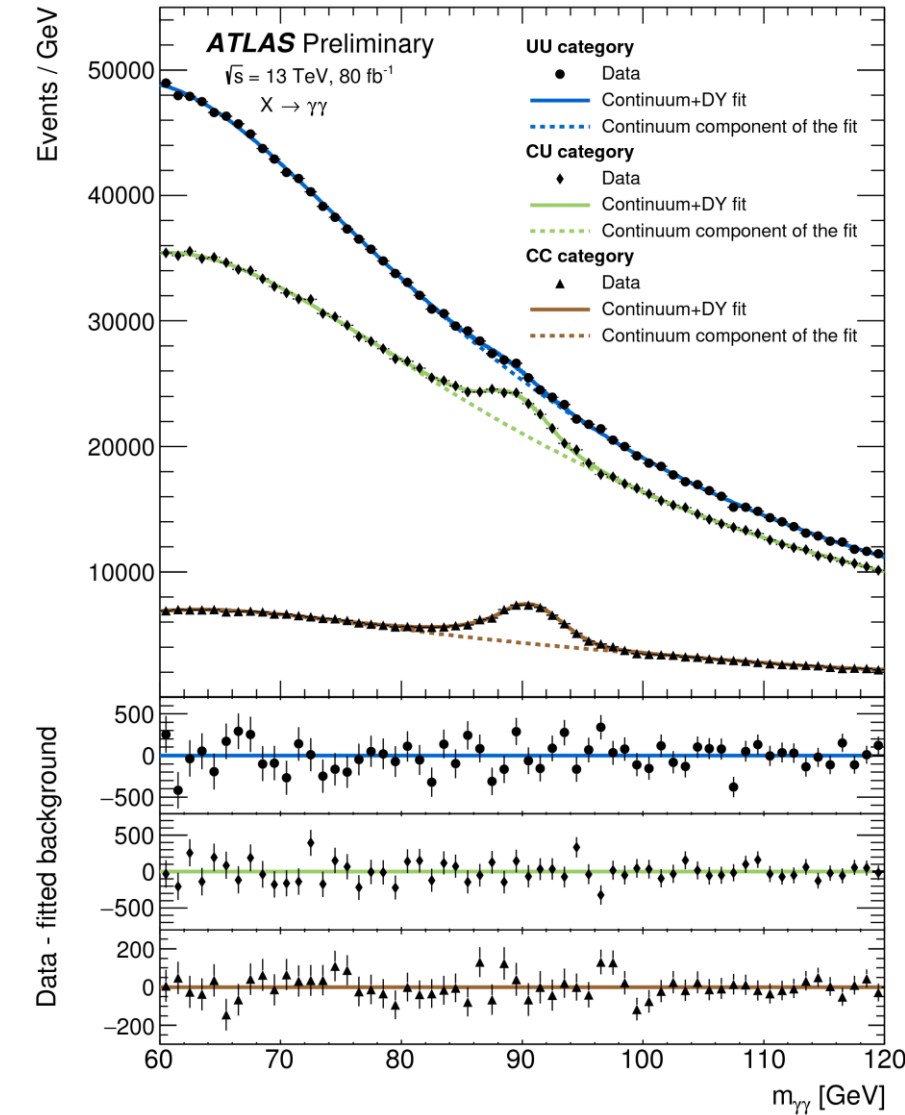
Classificazione di elettroni e fotoni



- **ROC Curve** (Receiver Operating Characteristic): a seconda della soglia che scelgo per la classificazione binaria ottengo un'efficienza per le particelle classificate come elettroni o come fotoni. Valutazione con **AUC**(Area Under the Curve).
- **Run 2 BDT/ Run 3 BDT** : modello allenato rispettivamente su campione MC con parametri della presa dati Run 2 o di Run 3.

Conclusioni

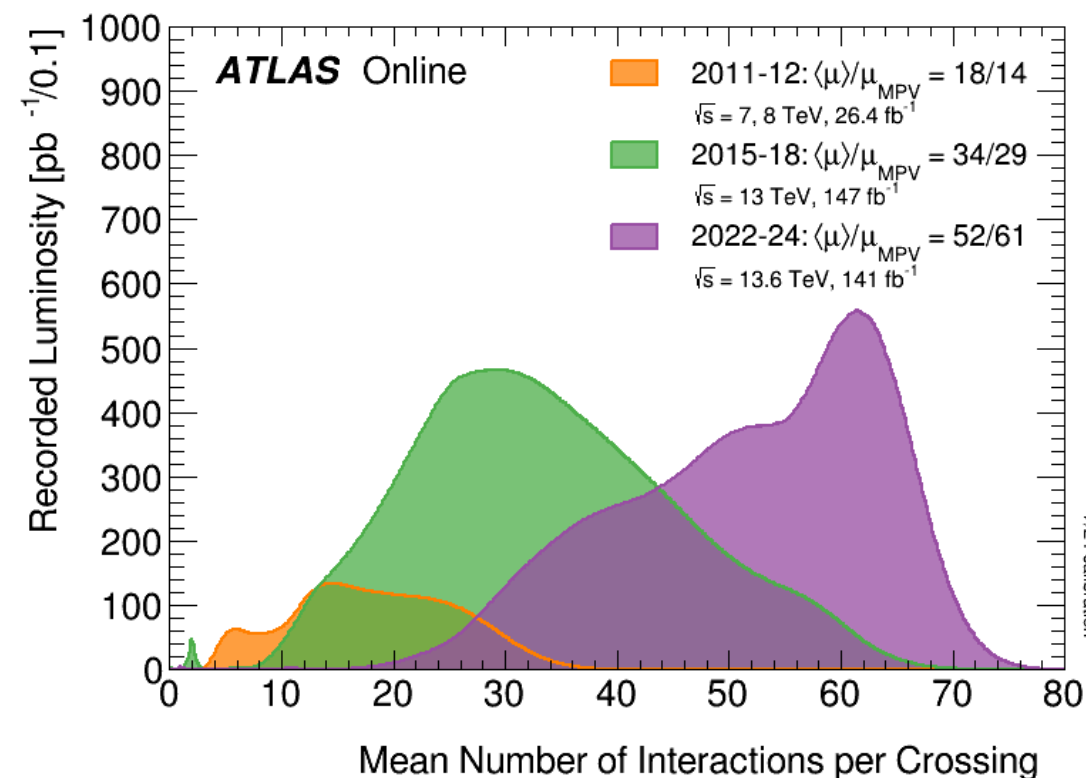
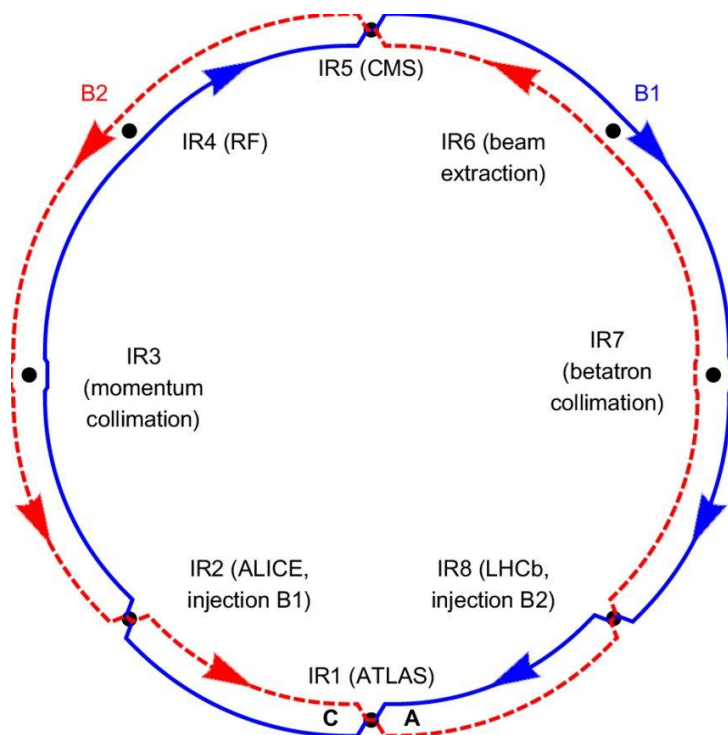
- Ulteriori sviluppi nello studio dell'argomento trattato dalla mia tesi potrebbero concentrarsi sull'applicazione del nuovo modello:
- Il grafico a destra mostra un picco di eventi ricostruiti come γ a 90 GeV. Questo picco deriva dalla frazione di elettroni ricostruiti come fotoni provenienti dal decadimento $Z \rightarrow ee$
- L'effetto della BDT dovrebbe essere quello di diminuire l'altezza del picco classificando correttamente una buona parte degli elettroni che producono questo picco.





Large Hadron Collider -LHC

- Ginevra, Svizzera. È un acceleratore di protoni e ioni pesanti installato nel tunnel di 27 km del LEP
 - Sequenza di magneti per curvare e focalizzare il fascio
 - Collisioni ad energia del centro di massa di 13.6 TeV



- Campagne di prese dati:
 - Run 1 → 2009 - 2013
 - Run 2 → 2015 - 2018
 - Run 3 → 2022 - in corso
- Condizioni sperimentali molto diverse



ATLAS

Sistema di magneti: 3 magneti toroidali ed uno solenoidale producono il campo magnetico interno ad ATLAS.

Nella parte esterna si trova lo **spettrometro per muoni**

Interaction Point (IP)

L'Inner Detector ricostruisce le tracce delle particelle cariche. Composto da due sotto rivelatori a Silicio di fine granularità, Pixel Detector ed SCT, ed un rivelatore a radiazione di transizione (TRT).

Il sistema di Calorimetri: misura l'energia di elettroni, fotoni e jet adronici. Calorimetro elettromagnetico (e^- , γ) ed uno adronico.

