



# UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

Misura della sezione d’urto  $pp \rightarrow b\bar{b} X$  con il rivelatore

CMS a LHC con piccola dipendenza dal modello

Relatore

Prof. Franco Simonetto

Laureando

Gaetano Maria Sifo

Anno Accademico 2021/2022



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Metodo di analisi</b>	<b>3</b>
1.1 Selezione degli eventi	3
1.1.1 Dati Utilizzati	3
1.2 Strategia di analisi	4
<b>2 Variabili discriminanti e preparazione dei campioni</b>	<b>6</b>
2.1 Muoni da stessa/diversa collisione	6
2.2 Determinazione di segnale da B e fondi	8
2.3 Definizione dei dati fittizi e verità Monte Carlo	10
2.3.1 Verità Monte Carlo: distribuzioni di $D_{xy}$	10
<b>3 Stima delle numerosità</b>	<b>12</b>
3.1 Stima di $N_1$	12
3.2 Stima di $N_2$	14
<b>4 Applicazione ai dati</b>	<b>18</b>
4.1 Stima della sezione d'urto	18
<b>Conclusioni</b>	<b>22</b>
<b>A Fit con modello completo per la stima di <math>N_1</math> ed <math>N_2</math></b>	<b>23</b>
<b>B Confronto delle estrapolazioni effettuate sul Monte Carlo e sui Dati</b>	<b>25</b>

# Introduzione

La misura della sezione d'urto di produzione  $pp \rightarrow b\bar{b} X$ , oltre a rappresentare un importante test della cromodinamica quantistica (QCD), risulta di fondamentale importanza per la misura dei rapporti di decadimento assoluti degli adroni con beauty, specialmente per  $B_S$  e  $\Lambda_b$  che non vengono prodotti nelle beauty-factories [5].

L'obiettivo di questa tesi è di sviluppare e testare una strategia di analisi che consenta di stimare la sezione d'urto in questione minimizzando al contempo la sensibilità della misura alle incertezze teoriche.

L'idea è quella di sfruttare l'approccio *tag & probe* per analizzare eventi nei quali sia già stato completamente ricostruito un mesone  $B^+$  (lato tag) e siano presenti ulteriori muoni non utilizzati per la ricostruzione (lato probe) e si mostrerà come, definendo specifici sotto-campioni sulla base delle caratteristiche dei muoni di probe e ricavandone la numerosità, sia possibile determinare tramite opportuni rapporti tutte le grandezze d'interesse.

In particolare, ispirandosi al lavoro [4], la misura qui presentata sfrutta il fatto che in un singolo evento di CMS vengono raccolte mediamente più di trenta collisioni indipendenti avvenute lungo l'asse dei fasci e spaziate all'incirca di un millimetro l'una dall'altra. Allora, come verrà spiegato successivamente, selezionando un campione di eventi con un "trigger" dal processo  $B^+ \rightarrow J/\psi K^+$  e ricercando un ulteriore muone da  $B$  nell'evento, nell'ipotesi in cui l'efficienza di selezione di muoni da  $B$  sia la stessa in tutte le condizioni, è possibile misurare la sezione d'urto di produzione delle coppie di beauty tramite il confronto tra il numero di eventi in cui il muone viene dalla stessa collisione in cui è prodotto il  $B^+$  e quello in cui esso proviene da una collisione diversa.

La tesi segue il seguente schema:

- nel Capitolo 1 verrà in primo luogo descritto il campione di dati utilizzato. Successivamente, verranno descritti i sotto-campioni e le equazioni necessarie ai fini della stima della sezione d'urto;
- nel Capitolo 2 verrà descritta l'analisi preliminare volta a stabilire criteri utili alla separazione degli eventi in sotto-campioni. In particolare, si mostrerà come tramite la proiezione longitudinale del parametro d'impatto sia possibile separare i muoni di probe provenienti dalla stessa/diversa collisione del  $B^+$  ricostruito. Infine, sfruttando le informazioni di generazione del Monte Carlo, verranno definite le categorie relative al segnale e ai fondi e verrà identificata una variabile discriminante;
- nel Capitolo 3 verranno descritte le equazioni e la procedura operativa per ricavare una stima delle numerosità dei sotto-campioni definiti nel primo capitolo. In particolare, si illustrerà come tramite il fit del Monte Carlo sui dati delle distribuzioni della proiezione trasversa del parametro d'impatto sia possibile stimare la frazione di muoni provenienti da  $B$ ;

- nel Capitolo 4 si applicherà l'intera procedura di analisi ad un campione di dati raccolti presso l'esperimento CMS ad LHC, giungendo così ad una stima effettiva della sezione d'urto di interesse.

# Capitolo 1

## Metodo di analisi

In questo capitolo si descrive dettagliatamente l'approccio ideato per giungere ad una misura della sezione d'urto di produzione  $pp \rightarrow b\bar{b} X$ .

Dopo aver introdotto il campione di dati utilizzato per l'analisi, si introdurranno le equazioni necessarie ai fini della stima della sezione d'urto.

### 1.1 Selezione degli eventi

L'intera strategia di analisi si basa sullo studio di muoni (detti di probe) raccolti in eventi nei quali sia stato completamente ricostruito un mesone  $B^+$  che funge da tag. Nello specifico, si è scelto di ricostruire il decadimento:

$$B^+ \rightarrow J/\psi K^+ \rightarrow \mu^+ \mu^- K^+$$

poiché, date le caratteristiche costruttive del CMS, esso risulta ben riconoscibile [1].

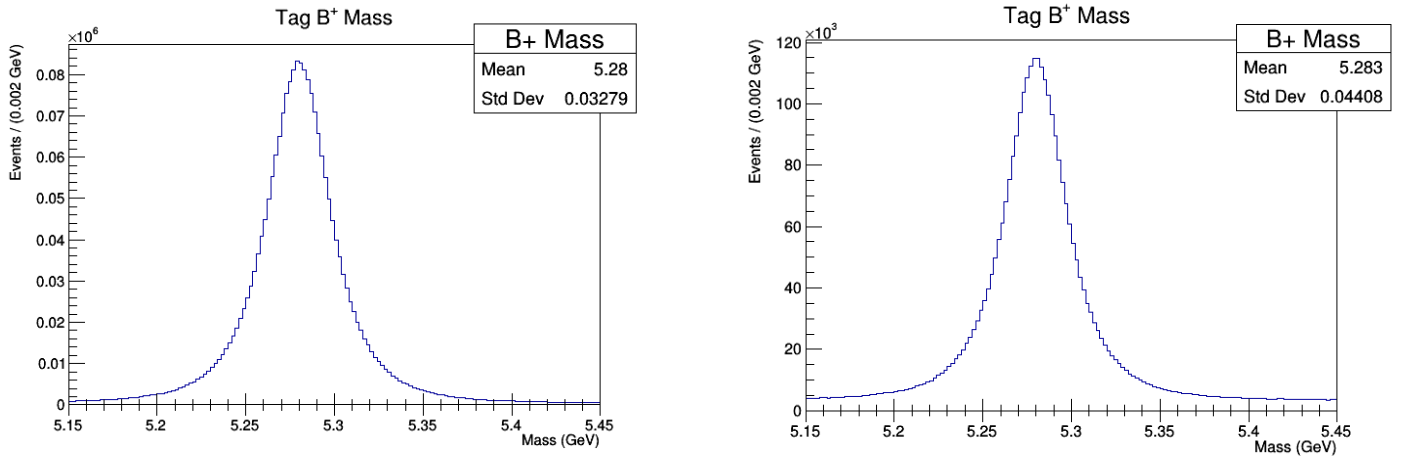
La terminologia *tag & probe* richiama l'omonimo metodo largamente diffuso nella fisica delle alte energie; in modo del tutto analogo, infatti, mentre i muoni di tag utilizzati per ricostruire il  $B^+$  devono soddisfare criteri molto stringenti (primo tra tutti quello sulla massa invariante che deve essere prossima a quella della  $J/\psi$ ), il lato probe si presenta come il generico rappresentante di un muone prodotto in una collisione  $pp$ , con il vincolo di esser stato prodotto nello stesso evento del  $B^+$  e altri requisiti minimi legati alla qualità della traccia. In particolare, si richiede che i muoni di probe abbiano momento trasverso  $p_T > 4$  GeV ed errore sul parametro d'impatto  $\sigma_{Dxy} < 50\mu\text{m}$  [4].

#### 1.1.1 Dati Utilizzati

L'intero studio di test è stato effettuato su un campione Monte Carlo di  $\approx 2 \cdot 10^6$  eventi con  $\sqrt{s} = 13$  TeV nei quali è presente un  $B^+$  completamente ricostruito. Di conseguenza, oltre alle grandezze cinematiche proprie degli studi di HEP, si dispone anche di numerose informazioni sulla genesi dei muoni analizzati.

Per l'analisi dati effettiva, è stata considerata una collezione di  $6.7 \times 10^5$  eventi raccolti a  $\sqrt{s} = 13$  TeV presso l'esperimento CMS con un  $B^+$  completamente ricostruito.

Si riportano le distribuzioni di massa del mesone  $B^+$  ottenute a partire da entrambi i campioni (dati e Monte Carlo) in figura [1.1].



**Figura 1.1:** Distribuzioni di massa del mesone  $B^+$  completamente ricostruito per i dati (sinistra) e per la simulazione Monte Carlo (destra)

## 1.2 Strategia di analisi

A partire dal campione appena descritto, sfruttando le informazioni relative ai muoni di probe, è possibile definire un certo numero di sotto-campioni e le rispettive numerosità<sup>1</sup>. In particolare, ricordando che tutti gli eventi sono caratterizzati da un  $B^+$  completamente ricostruito al quale è associato un vertice di collisione, si definiscono i seguenti campioni:

- eventi con un solo muone da  $B$  dallo stesso vertice del  $B^+$  (**N1**)<sup>2</sup>;
- eventi con almeno un muone da  $B$  prodotti nello stesso evento del  $B^+$  ma da un altro vertice di collisione (**N2**);
- eventi con due muoni da  $B$  dallo stesso vertice del  $B^+$  (**N3**);
- eventi con almeno due muoni da  $B$ , uno proveniente dallo stesso vertice del  $B^+$  ricostruito e almeno un altro da una diversa collisione (**N4**).

L'intera strategia adottata per stimare la sezione d'urto d'interesse si basa sul costruire rapporti tra le numerosità dei sotto-campioni così individuati. A partire da considerazioni del tutto generali e di natura puramente probabilistica, è possibile infatti trovare per le quattro quantità delle espressioni esplicite che prescindono dal modello che regola i processi di produzione sottostanti.

<sup>1</sup>Intese come il numero di muoni appartenenti al sotto-campione.

<sup>2</sup>Richiedendo la presenza di un solo muone da  $B$  nell'evento, **N1** conta tanto il numero di eventi quanto il numero di muoni

Nello specifico, si avrà:

$$\left\{ \begin{array}{l} N_1/N_0 = \left( \frac{\sigma_2}{\sigma_2 + \sigma_4} \epsilon + 3 \frac{\sigma_4}{\sigma_2 + \sigma_4} \epsilon (1 - \epsilon)^2 \right) \\ N_2/N_0 = \text{nP} \left( 2 \frac{\sigma_2}{\sigma_{ppX}} \epsilon + 4 \frac{\sigma_4}{\sigma_{ppX}} \epsilon \right) \\ N_3/N_0 = 6 \frac{\sigma_4}{\sigma_2 + \sigma_4} \epsilon^2 (1 - \epsilon) \\ N_4/N_0 = \text{nP} \left( \frac{\sigma_2}{\sigma_2 + \sigma_4} \epsilon + 3 \frac{\sigma_4}{\sigma_2 + \sigma_4} \epsilon (1 - \epsilon)^2 \right) \left( 2 \frac{\sigma_2}{\sigma_{ppX}} \epsilon + 4 \frac{\sigma_4}{\sigma_{ppX}} \epsilon \right) \end{array} \right.$$

dove sono state introdotte le seguenti grandezze:

- $N_0$ , numero totale di eventi in cui è stato ricostruito il mesone  $B^+$ ;
- $\sigma_2$ , sezione d'urto d'interesse, ovvero  $\sigma(pp \rightarrow b\bar{b} X)$ ;
- $\sigma_4$ , sezione d'urto collegata al cosiddetto *double gluon splitting* che porta alla produzione di due coppie  $b\bar{b}$ ;
- $\sigma_{ppX} = \sigma(pp \rightarrow X)$ , sezione d'urto totale di scattering inelastico  $pp$ ;
- $\epsilon$ , prodotto tra la branching ratio del decadimento del  $B$  in  $\mu$  (diretto o in cascata) e l'efficienza di rilevazione dei muoni, che si suppone esser la stessa in tutte le collisioni raccolte in ciascun evento;
- $\text{nP}$ , numero di vertici di collisione di pile-up registrati nell'evento (pari cioè al numero totale di vertici ricostruiti meno uno).

Essendo  $N_0$ ,  $\sigma_{ppX}$  e  $\text{nP}$  quantità note, il sistema risulta sovra-costretto e fornisce quindi soluzioni per i tre parametri incogniti.

Per semplificare il sistema di equazioni, si è scelto di trascurare la possibilità di produzione di due coppie  $b\bar{b}$  in quanto, essendo il fenomeno estremamente più raro e data la precisione attesa della misura, esso può esser trattato come effetto di ordine superiore.

Tale approssimazione si traduce nel considerare  $\sigma_4 \ll \sigma_2$ , giungendo a:

$$\left\{ \begin{array}{l} N_1/N_0 = \epsilon \\ N_2/N_0 = 2 \text{nP} \frac{\sigma_2}{\sigma_{ppX}} \epsilon \\ N_3/N_0 = 6 \frac{\sigma_4}{\sigma_2} \epsilon^2 (1 - \epsilon) \\ N_4/N_0 = 2 \text{nP} \frac{\sigma_2}{\sigma_{ppX}} \epsilon^2 \end{array} \right. \quad (1.1)$$

Ai fini della stima di  $\sigma_2$ , si è scelto di sfruttare le prime due equazioni costruendo il rapporto:

$$\frac{N_2}{N_1} = 2 \text{nP} \frac{\sigma_2}{\sigma_{ppX}} = 2 \text{nP} R \quad (1.2)$$

L'analisi dati è dunque volta alla determinazione di  $N_1$  ed  $N_2$ , il che, ricordando la definizione dei sotto-campioni corrispondenti, implica lo sviluppo di strategie per separare i muoni di probe provenienti dallo stesso/diverso vertice di collisione del  $B^+$  e per determinare la frazione di muoni originati da  $B$  rispetto agli altri fondi.



# Capitolo 2

## Variabili discriminanti e preparazione dei campioni

In questo capitolo si descrive come, attraverso lo studio del parametro d'impatto, sia possibile suddividere l'insieme degli eventi nei due sotto-campioni delineati nel capitolo precedente

Nella fisica delle alte energie, il *parametro d'impatto*  $D$  è definito come la minima distanza tra la traccia di una particella prodotta nella collisione e un certo punto di riferimento, in questo caso scelto come il vertice di produzione del  $B^+$  completamente ricostruito.

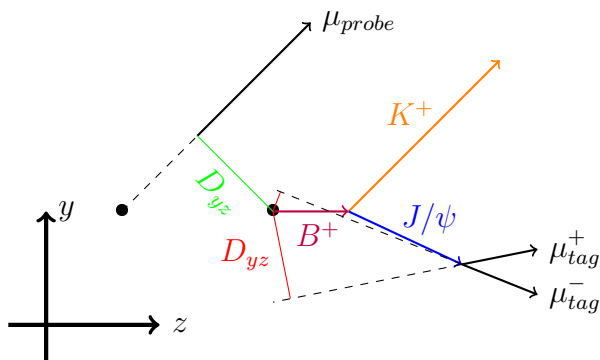
### 2.1 Muoni da stessa/diversa collisione

Per poter dividere i muoni di probe provenienti dallo stesso vertice di collisione del  $B^+$  da quelli originati in un vertice differente, è necessario identificare un metodo per quantificare la distanza del muone dal vertice ricostruito. Per sua natura, la *proiezione longitudinale del parametro d'impatto*  $D_z$  fornisce esattamente le informazioni necessarie a questo scopo (situazione rappresentata schematicamente in figura 2.1). Nello specifico, la differenza

$$\Delta D_z = D_z^{probe} - D_z^{tag}$$

fornisce una stima di quanto il muone di probe si allontani dal vertice ricostruito rispetto ai muoni di tag ( $\mu_{tag}^\pm$ ) che, per costruzione, provengono da quest'ultimo. Operativamente, si è scelto in primo luogo un criterio per la definizione del  $D_z^{tag}$ :

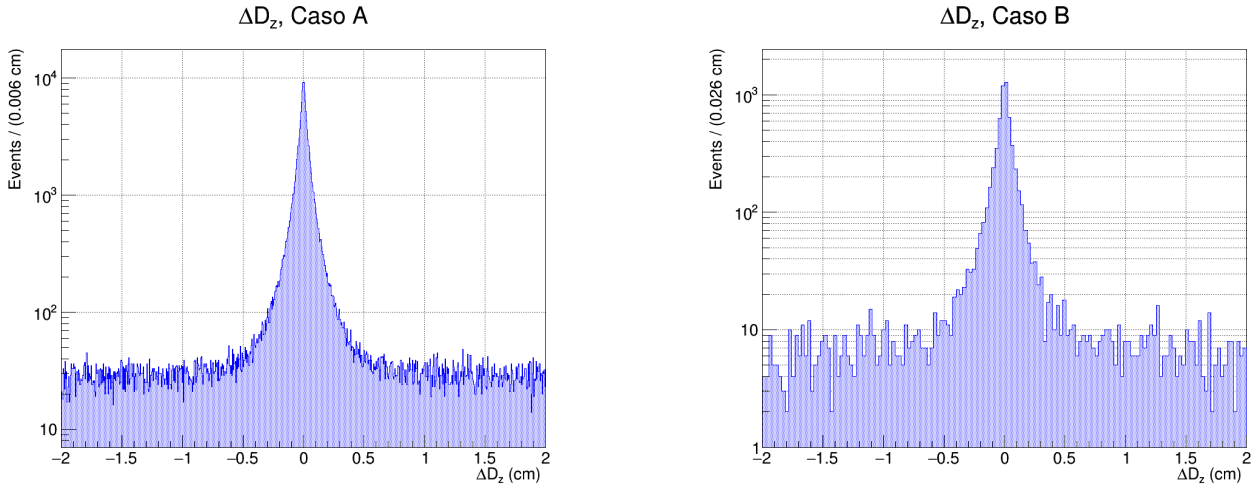
- **Caso A** : per entrambi i  $\mu_{tag}^\pm$  risulta correttamente misurato il valore di  $D_{z,\pm}^{tag}$ ,  $D_z^{tag}$  è definito come la media dei due;
- **Caso B** : solamente uno dei  $\mu_{tag}^\pm$  presenta la misura associata di  $D_{z,\pm}^{tag}$ ,  $D_z^{tag}$  prende quel valore;



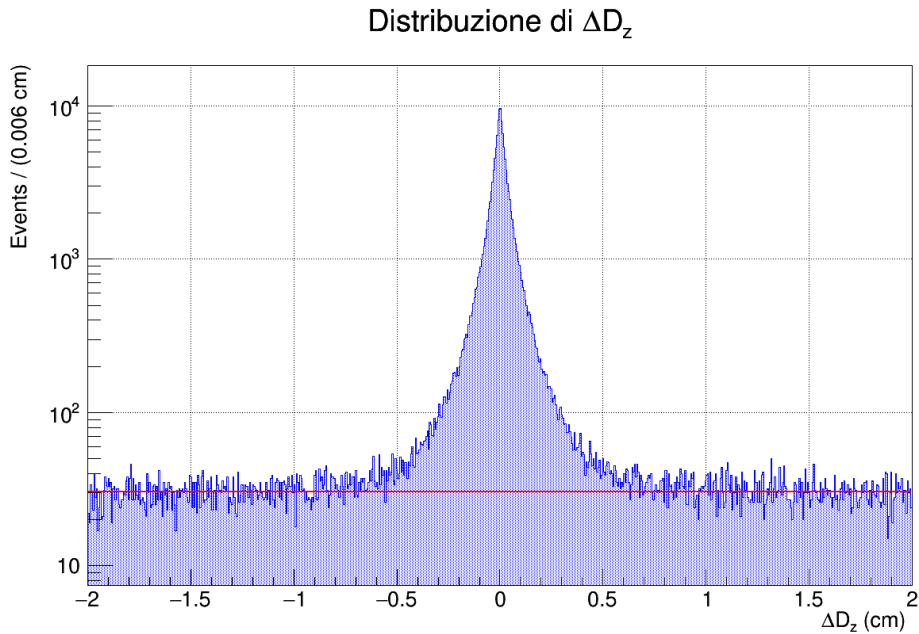
**Figura 2.1:** Rappresentazione schematica dei muoni di tag associati al  $B^+$  ricostruito e di un muone di probe da diversa collisione; in rosso e in verde le rispettive proiezioni sul piano yz del parametro d'impatto

- $D_{z,\pm}^{tag}$  non è misurato per nessun  $\mu_{tag}^\pm$ , l'evento si scarta<sup>1</sup>.

A questo punto, mantenendo separati Caso A/B, sono state create le distribuzioni di  $\Delta D_z$  per poter stabilire un criterio generale per la categorizzazione (riportate in figura 2.2).



**Figura 2.2:** Distribuzioni di  $\Delta D_z$  per entrambe le categorie delineate; a sinistra  $D_{z,\pm}^{tag}$  è misurato per entrambi i  $\mu_{tag}$ , a destra si dispone della misura di  $D_{z,\pm}^{tag}$  solo per uno di essi.



**Figura 2.3:** Distribuzione di  $\Delta D_z$  a campioni unificati; in rosso, modello costante per stimare il fondo di eventi da diversa collisione.

In entrambi i casi trattati, le distribuzioni presentano un pronunciato picco centrale corrispondente ai muoni da stessa collisione e un fondo costante per i muoni da un vertice diverso. Per separare correttamente le due componenti, si è tentato in primo luogo un fit del picco con una gaussiana, poi una interpolazione completa con un modello dato dalla somma di due gaussiane e una costante; non essendo però il picco propriamente gaussiano, i risultati di ambedue i fit presentano una forte dipendenza sia dal valore dei parametri utilizzati per inizializzare il fit sia dall'intervallo utilizzato.

<sup>1</sup>La percentuale di eventi scartati è minore dell' 1%

Constatata dunque la totale sovrapponibilità delle distribuzioni e non potendo separare analiticamente gli eventi tramite un fit, si è scelto di adottare su base puramente qualitativa le seguenti definizioni<sup>2</sup>:

- *stessa collisione*: muoni caratterizzati da  $|\Delta D_z| \leq 0.3$  cm
- *diversa collisione*: eventi caratterizzati da  $|\Delta D_z| \geq 0.6$  cm.

Inoltre, sempre in virtù dell'analogia tra le distribuzioni, il resto dell'analisi verrà svolto a campioni unificati, continuando a distinguere caso A/B solo ai fini della definizione di  $D_z^{tag}$ . Per poter stabilire correttamente l'entità del fondo e per evidenziare eventuali asimmetrie, si è scelto di fittare separatamente la side-band destra e sinistra con due funzioni costanti. Avendo ottenuto due valori compatibili ed esclusa l'ipotesi di asimmetria, si ottiene un'unica stima del fondo di muoni da vertice diverso (figura 2.3) pari a :

$$C_{div} = 30.5 \pm 0.2 \text{ Events} / (0.006 \text{ cm})$$

A partire da quest'ultima costante, è possibile ricavare due quantità necessarie ai fini delle analisi successive. In particolare, è necessario stimare:

- La frazione di muoni provenienti da diversa collisione presenti nel range  $|\Delta D_z| \leq 0.3$  cm sul totale di eventi nello stesso intervallo; si ottiene costruendo il rapporto tra l'integrale della costante e quello dell'istogramma in quell'intervallo e vale:

$$\epsilon_{div} = 0.0238 \pm 0.0002 \quad (2.1)$$

- il rapporto tra il numero di muoni da diversa collisione presenti nella regione esclusa ( $|\Delta D_z| \leq 0.6$  cm) stimati attraverso il fit costante e il numero di muoni da collisioni diverse nella side-band ( $|\Delta D_z| \geq 0.6$  cm). Esso vale:

$$\epsilon_{side-band} = 0.1016 \pm 0.0009 \quad (2.2)$$

## 2.2 Determinazione di segnale da B e fondi

Dopo aver elaborato una strategia per separare i muoni provenienti dallo stesso vertice di collisione del  $B^+$  completamente ricostruito da quelli originati in altre collisioni, è necessario stabilire un metodo per stimare la frazione di muoni di probe originati da  $B$  in qualsiasi momento della loro storia, includendo quindi sia i muoni diretti provenienti da decadimenti leptonici e semi-leptonici sia quelli prodotti in cascata (es.  $B^\pm \rightarrow J/\psi K \rightarrow \mu^+ \mu^- K^\pm$ ).

Sfruttando le informazioni di generazione del Monte Carlo, il primo passo è quello di suddividere gli eventi in categorie che definiscano il segnale da  $B$  e i vari fondi. Nello specifico, si è scelto di raggruppare gli eventi tramite le informazioni disponibili sull'insieme dei processi che hanno portato alla produzione di ciascun muone di probe, identificando 5 categorie (1 segnale + 4 fondi):

1. **Eventi da B**: tutti gli eventi con muoni provenienti da  $B$  in qualsiasi momento della loro vita (cioè per produzione diretta, per un processo in cascata  $b \rightarrow cX$ ,  $c \rightarrow \mu Y$  e anche per decadimenti di mesoni leggeri  $b \rightarrow \pi/KX$ ,  $\pi/K \rightarrow \mu Y$ );
2. **Eventi da Charm**: eventi provenienti dal decadimento di adroni charmati in cui il quark  $c$  non è stato prodotto dal decadimento del  $b$ ;
3. **Eventi Light-Quark**: eventi provenienti da Pioni e Kaoni non prodotti da  $b$ ;

---

<sup>2</sup>Valide sia per Caso A che per Caso B

4. **Eventi Else:** tutti gli altri eventi per cui esiste un'informazione in simulazione che possa mostrare che il muone non è prodotto da un decadimento del  $b$ ;
5. **Eventi Null:** eventi per i quali non esiste informazione di simulazione;

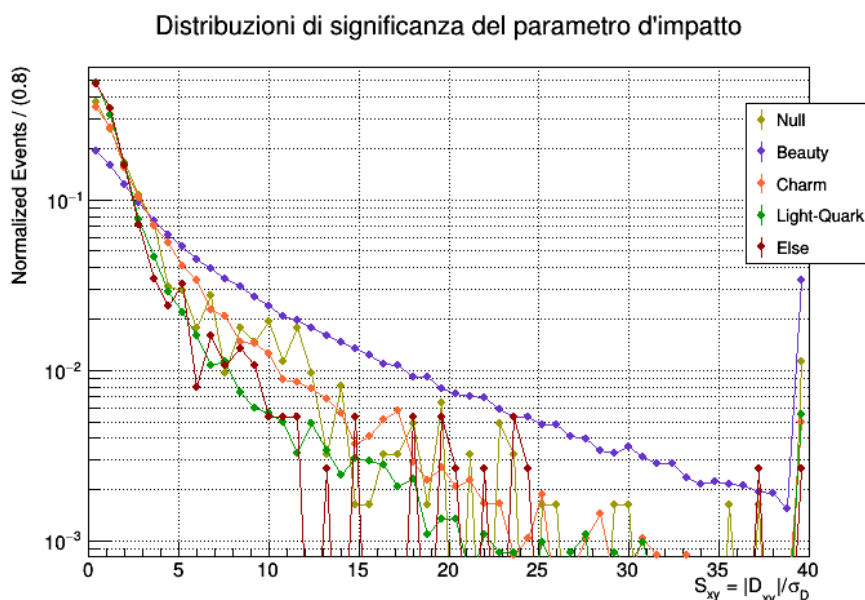
A questo punto, ai fini delle analisi successive volte a stabilire la frazione di eventi appartenenti a ciascuna categoria, si sceglie di sfruttare la *proiezione trasversa del parametro d'impatto*  $D_{xy}$  quale variabile discriminante. Essa, infatti, rappresenta una delle scelte più naturali nel contesto della cosiddetta *heavy-flavour jet identification* [2].

Affinché tale scelta risulti efficace per le analisi successive, si procede con la verifica della bontà dei dati di simulazione tramite lo studio delle distribuzioni di significanza per  $D_{xy}$ , ovvero graficando il rapporto di significanza

$$S_{xy} = \frac{|D_{xy}|}{\sigma_{D_{xy}}}$$

e normalizzando ciascun grafico all'unità (figura 2.4).

Gli andamenti delle distribuzioni di  $S_{xy}$  per le categorie da  $B$ ,  $C$  e light-quark riproducono correttamente quanto atteso (si veda ad esempio [2]); le altre due categorie, in quanto raccoglitori di eventi di natura anche molto diversa, non presentano un comportamento noto a priori.



**Figura 2.4:** Distribuzioni di significanza di  $D_{xy}$  normalizzate all'unità per ciascuna categoria individuata. L'ultimo bin contiene tutti gli overflow.

## 2.3 Definizione dei dati fittizi e verità Monte Carlo

Come verrà specificato meglio nei prossimi capitoli, tramite lo studio del  $D_{xy}$  è possibile costruire un modello complessivo di segnale e fondi con cui stimare la frazione di eventi da  $B$  adoperando un fit del Monte Carlo sui dati. Di conseguenza, per poter effettivamente testare l'approccio sviluppato, è stato introdotto un criterio generico<sup>3</sup> attraverso il quale separare la simulazione in due sotto-campioni. Nello specifico, si è scelto di sfruttare il numero di evento associato a ciascuna entrata del database utilizzato, definendo come "dati" tutti gli eventi con eventNumber dispari e lasciando ai restanti la funzione di Monte Carlo.

### 2.3.1 Verità Monte Carlo: distribuzioni di $D_{xy}$

Delineata la porzione di simulazione da usare come dati fittizi e dovendo costruire un modello con cui effettuare il fit, è necessario ricavare a partire dal Monte Carlo le distribuzioni della proiezione trasversa del parametro d'impatto  $D_{xy}$  per ciascuna categoria di segnale/fondi individuata.

In particolare, dovendo stimare le quantità  $N_1$  ed  $N_2$ , bisogna considerare separatamente i sotto-campioni:

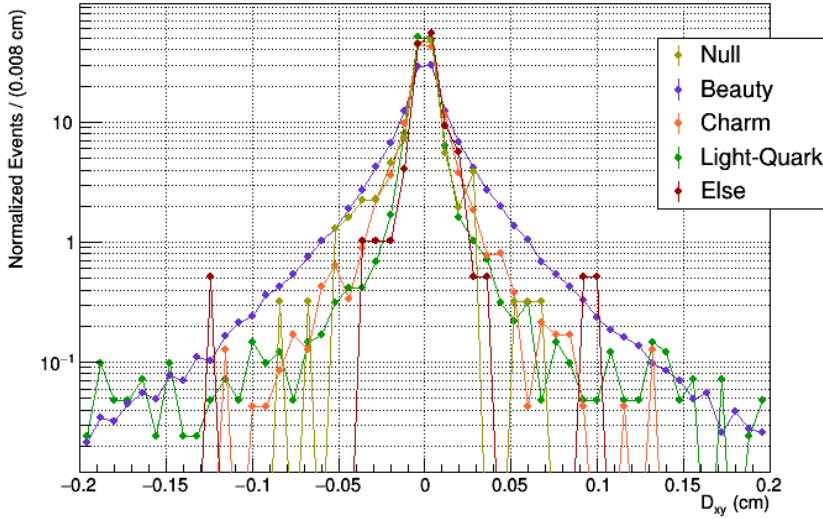
- eventi con un solo muone proveniente dalla stessa collisione del  $B^+$  ricostruito;
- eventi con almeno un muone prodotto in un vertice di collisione diverso da quello del  $B^+$ .

Allora, sfruttando le informazioni di generazione della simulazione, si ricavano per ciascuno le distribuzioni di  $D_{xy}$  correttamente normalizzate all'unità e le numerosità di ciascuna categoria di fondo/segnale (risultati riportati in figura [2.5](#) e [2.6](#) ed in tabella [2.1](#) e [2.2](#)).

---

<sup>3</sup>Il criterio è stato scelto in modo da introdurre il maggior grado possibile di casualità nella suddivisione, cercando quindi di non generare alcun tipo di distorsione nei dati.

Distribuzioni  $D_{xy}$  (Monte Carlo)

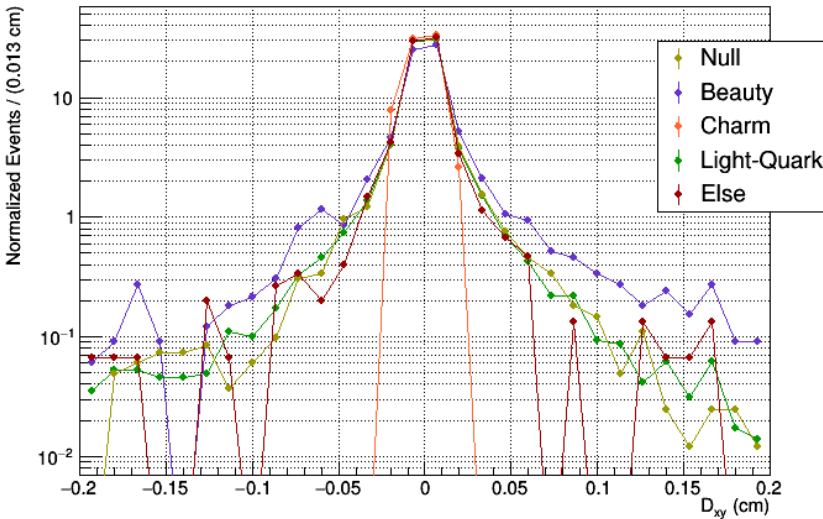


**Figura 2.5:** Distribuzioni Monte Carlo normalizzate all'unità della proiezione trasversa del parametro d'impatto per gli eventi con un solo muone da stessa collisione. I bin agli estremi raccolgono gli overflow.

Categorie	Events
Beauty	57493 (87.1%)
Charm	2858 (4.3%)
Light Quark	5022 (7.61%)
Else	216 (0.33%)
Null	380 (0.58%)
<b>Totale</b>	<b>65969</b>

**Tabella 2.1:** Numerosità di ciascuna categoria per gli eventi Monte Carlo con un solo muone proveniente dallo stesso vertice del  $B^+$  ricostruito.

Distribuzioni  $D_{xy}$  (Monte Carlo)



**Figura 2.6:** Distribuzioni Monte Carlo normalizzate all'unità della proiezione trasversa del parametro d'impatto per i muoni da diversa collisione. I bin agli estremi raccolgono gli overflow.

Categorie	Events
Beauty	2593 (8.06%)
Charm	31 (0.09%)
Light Quark	22108 (68.7%)
Else	1151 (3.57%)
Null	6289 (19.5%)
<b>Totale</b>	<b>32172</b>

**Tabella 2.2:** Numerosità di ciascuna categoria per gli eventi Monte Carlo con almeno un muone proveniente da un vertice diverso da quello del  $B^+$  ricostruito.

# Capitolo 3

## Stima delle numerosità

In questo capitolo si testa l'intera procedura di analisi dati ideata ai fini della stima della sezione d'urto tramite un fit di Monte Carlo su Monte Carlo. In particolare, si illustra come sia possibile giungere ad una stima per le quantità  $N_1$  ed  $N_2$  sfruttando le informazioni raccolte nelle analisi preliminari e tramite lo studio delle distribuzioni della proiezione trasversa del parametro d'impatto.

### 3.1 Stima di $N_1$

Ricordando che per definizione  $N_1$  indica il numero di eventi con un solo muone da  $B$  proveniente dallo stesso vertice del  $B^+$  ricostruito, è possibile ricavare una formula esplicita che ne consenta la stima sulla base di informazioni ricavabili dai dati. In particolare, possiamo scrivere:

$$N_1 = N_{tot}^{same} (1 - \epsilon_{div}) f_B \quad \sigma_{N_1}^2 \approx N_1^2 \left[ \left( \frac{\sigma_{fb}}{f_B} \right)^2 + \left( \frac{\sigma_{\epsilon_{div}}}{1 - \epsilon_{div}} \right)^2 \right] \quad (3.1)$$

dove  $N_{tot}^{same}$  indica il numero totale di eventi con un solo muone proveniente dallo stesso vertice di collisione ricostruito ed  $f_B$  è la frazione di eventi provenienti da  $B$ . Si ricorda inoltre che  $\epsilon_{div}$ , come ricavato in [2.1](#), indica la frazione di eventi da collisioni diverse inclusi erroneamente nel campione selezionato.

Individuata la formula esplicita per  $N_1$ , è necessario innanzitutto selezionare tra i dati, tramite le strategie sviluppate nel precedente capitolo, i muoni di probe provenienti dalla stessa collisione<sup>1</sup> del  $B^+$  e ricavare la distribuzione del parametro d'impatto  $D_{xy}$  associata (riportati in figura [3.1](#)).

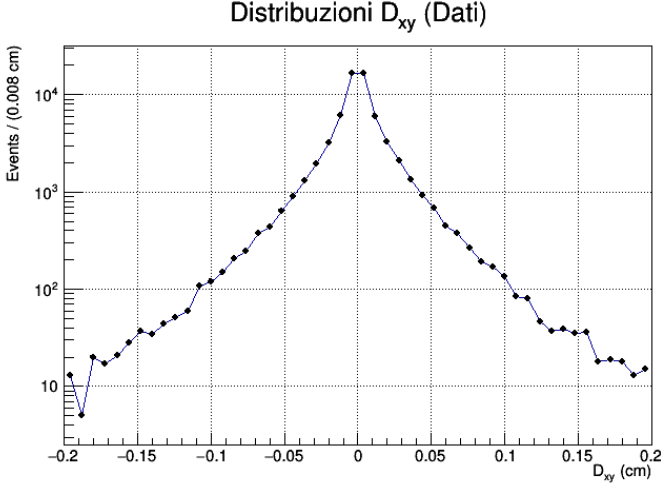
Successivamente, combinando opportunamente le distribuzioni Monte Carlo ricavate nel precedente capitolo e riportate in figura [2.5](#), è possibile definire un modello non parametrico con il quale effettuare un fit sui dati e stimare di conseguenza la frazione da  $B$  ricercata. Nello specifico, dette  $T_i$  le distribuzioni di ciascuna categoria ed  $f_i$  le frazioni di eventi associate, definiamo un modello  $M$  debitamente normalizzato come:

$$\begin{cases} M = \sum_{i=0}^{N_{cat}} f_i T_i \\ \sum_{i=0}^{N_{cat}} f_i = 1 \end{cases} \quad (3.2)$$

---

<sup>1</sup>Con la accortezza aggiuntiva di considerare esclusivamente eventi in cui un solo muone soddisfa tale condizione

in cui i parametri da fittare sono rappresentati proprio dalle  $f_i$ , ciascuna inizializzata alla corrispondente frazione ricavata dal Monte Carlo riportata in tabella [2.1](#).



**Figura 3.1:** Distribuzione della proiezione trasversa del parametro d’impatto dei dati per i muoni da stessa collisione del  $B^+$ . I bin agli estremi raccolgono gli overflow ( $N_{tot}^{same} = 65953$ )

risultati appare subito evidente la prevista instabilità del fit a cinque componenti. Le frazioni relative ai fondi, infatti, assumono valori poco affidabili e, in certi casi, del tutto inverosimili. Per cercare quindi di migliorare la stabilità del fit e per ottenere risultati più attendibili, si è scelto di ridurre il numero di parametri liberi nel modello fissando le due componenti meno significative (i.e. fondo da eventi *Null* ed *Else*) ai valori stimati a partire dal Monte Carlo (risultati del fit in figura [3.2](#)).

Seppur il  $\chi^2$  risulti elevato a causa di alcune fluttuazioni che, come si evince dai residui, interessano soprattutto le code della distribuzione, il fit con il modello semplificato risolve in parte i problemi sopra evidenziati. Con questo approccio, infatti, sia la frazione da B che i fondi assumono valori plausibili e confrontabili con quanto ottenuto dai dati simulati. Per queste ragioni, si sceglie di proseguire l’analisi sfruttando  $f_B$  ricavato con il modello semplificato<sup>2</sup> pari a:

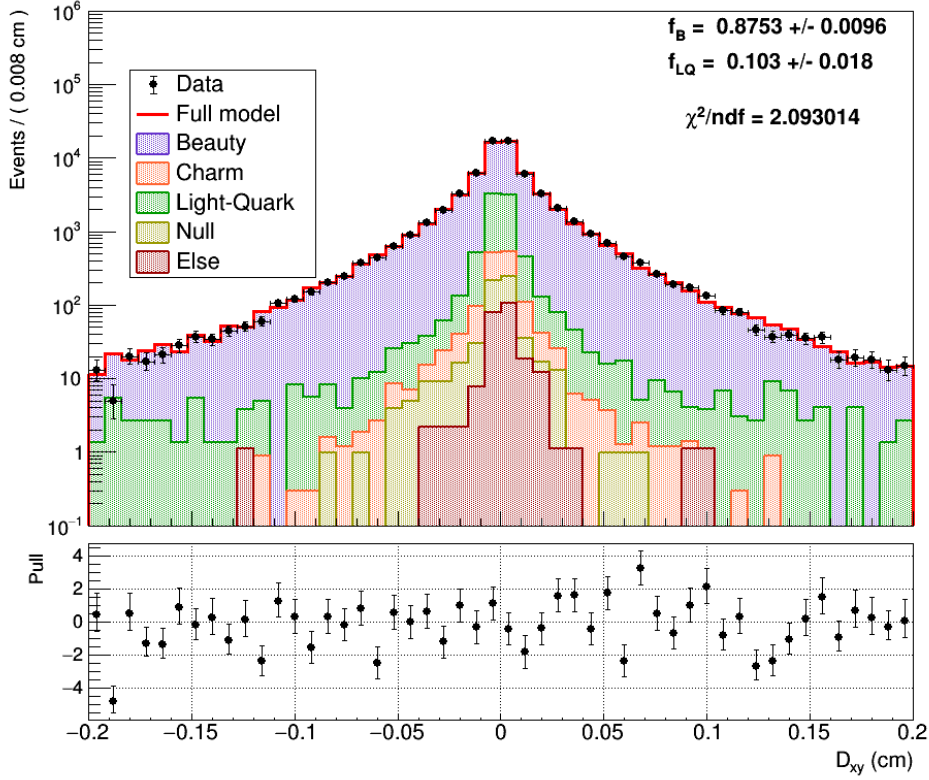
$$f_B = 0.87 \pm 0.01$$

Disponendo allora di tutti i parametri necessari, è possibile stimare  $N_1$  tramite la formula [3.1](#), ottenendo

$$N_1 = (57.7 \pm 0.6) \times 10^3$$

<sup>2</sup>Si noti in ogni caso come le frazioni di eventi da B ottenuti a partire dalle due stime si discostino per meno di una deviazione standard





**Figura 3.2:** Fit con modello semplificato della distribuzione della proiezione trasversa del parametro d'impatto per la stima di  $N_1$ . Le componenti Null ed Else sono fissate al valore ricavato dal Monte Carlo mentre la frazione da Charm non riportata si intende determinata per normalizzazione a 1.

## 3.2 Stima di $N_2$

Pur riprendendo in parte l'approccio seguito nella sezione precedente, la stima di  $N_2$  presenta alcune peculiarità che richiedono una trattazione specifica. In particolare, ricordando che per definizione  $N_2$  conta tutti i muoni da  $B$  provenienti da una collisione diversa da quella in cui è stato ricostruito il  $B^+$ , esso dipenderà necessariamente dal numero di vertici di pile-up (nP) registrato in ogni evento.

Prima di procedere con la stima della frazione di eventi da  $B$  sarà quindi necessario analizzare la dipendenza da nP e sviluppare un metodo per trattare correttamente il problema. In particolare, è necessario esaminare l'andamento della quantità  $N_{div}(nP)$ , ovvero del numero totale di muoni provenienti da una diversa collisione in funzione del numero di vertici di pile-up ricostruiti.

A tal proposito, sulla base di semplici considerazioni probabilistico-combinatorie, detto nEv(nP) il numero totale di eventi con un dato nP ed  $f$  il numero medio di muoni di pile-up, ci si attende che  $N_{div}(nP)$  sia esprimibile come:

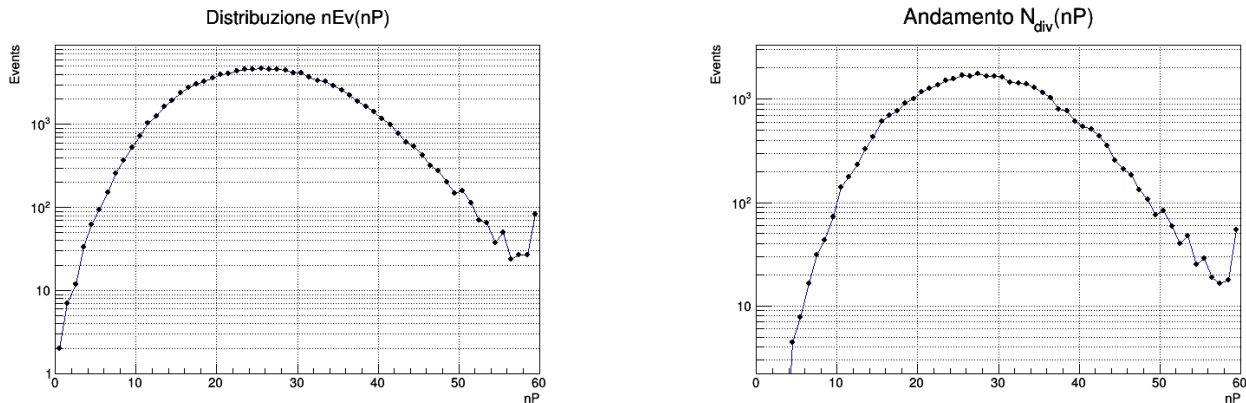
$$N_{div}(nP) = nEv(nP) f nP \quad (3.3)$$

ovvero ci si aspetta che la quantità:

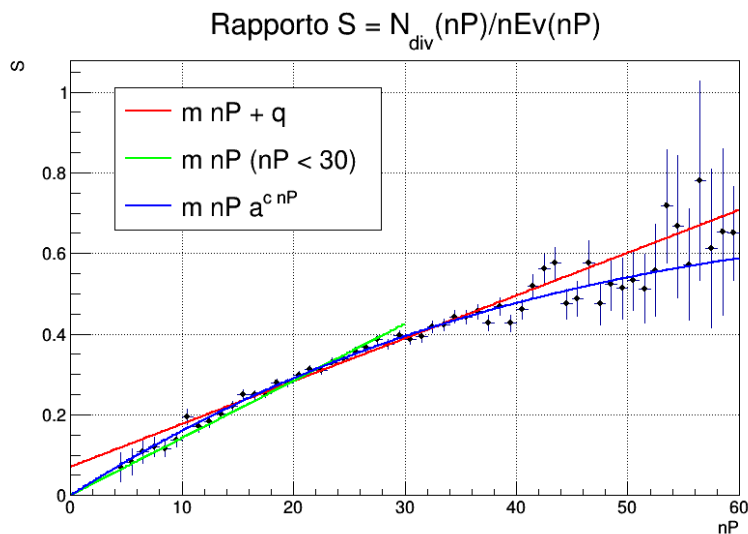
$$S(nP) = \frac{N_{div}(nP)}{nEv(nP)} = f nP \quad (3.4)$$

abbia un andamento lineare.

Per verificare tale ipotesi e per stimare  $f$ , si procede con la costruzione delle distribuzioni delle due grandezze  $N_{div}(nP)$  ed  $nEv(nP)$  (figura 3.3); si sottolinea come nella costruzione dell'istogramma di  $N_{div}(nP)$  sia stata già effettuata la correzione per il fattore  $(1 + \epsilon_{side-band})$  ricavato in (2.2) per conteggiare correttamente gli eventi da diversa collisione. Si riporta infine in figura 3.4 il rapporto  $S(nP)$ .



**Figura 3.3:** Distribuzioni di  $nEv$  (a sinistra) e di  $N_{div}$  (a destra) in funzione del numero di vertici di pile-up  $nP$ . I valori di  $N_{div}$  nell'istogramma di destra sono già corretti per la frazione  $\epsilon_{side-band}$ .



**Figura 3.4:** Andamento del rapporto  $S(nP)$  in funzione del numero di vertici di pile-up  $nP$ . Si riportano, inoltre, l'interpolazione lineare a due parametri (rosso), l'interpolazione lineare ad intercetta fissata e con dominio limitato (verde) e il modello con correzione esponenziale (blu).

Nonostante dal modello sviluppato ci si aspetti un andamento lineare ad intercetta nulla, per non introdurre vincoli esterni si è scelto in primo luogo di effettuare un fit lineare a due parametri (del tipo  $m nP + q$ , in rosso nella figura) su tutto l'intervallo considerato e di verificare a posteriori tale ipotesi. Come è possibile apprezzare dal grafico, però, per  $nP > 30$  i dati sembrano allontanarsi via via dall'ipotesi addotta ( $\chi^2/ndf = 73.2/54$ ) e, in aggiunta, il fit restituisce  $q = 0.091 \pm 0.008$ , incompatibile con zero per più di  $10\sigma$ .

Tale andamento anomalo non risulta giustificabile sulla base di conoscenze note e potrebbe testimoniare un'inefficienza che si presenta in eventi ad alta energia e con altissima molteplicità e che meriterebbe uno studio specifico e sistematico.

A questo stadio, per poter proseguire con la stima di  $N_2$ , si sceglie di procedere con un approccio empirico-osservativo, individuando cioè una serie di modelli che possano parametrizzare al

meglio l'andamento dei dati. In particolare, sono stati introdotti un fit lineare con intercetta fissata a zero e ristretto al dominio in cui i dati seguono l'andamento lineare atteso ( $nP \leq 30$ ) e un fit con un modello che parametrizza l'inefficienza tramite un termine moltiplicativo esponenziale (riportati rispettivamente in verde e in blu in figura 3.4). Si sottolinea come, per piccoli valori di  $nP$ , i due modelli considerati producono risultati simili.

Studiata la dipendenza da  $nP$ , è possibile giungere ad una espressione esplicita per la quantità  $N_2$ . Essa, infatti, sarà esprimibile come:

$$N_2 = \left( \sum_{nP} N_{div}(nP) f_B(nP) \right) = \sum_{nP} nEv(nP) \xi(nP) f_B(nP) \quad (3.5)$$

dove  $\xi(nP)$  indica il generico modello adottato precedentemente per i fit e  $f_B(nP)$  rappresenta la frazione di muoni da  $B$  sul totale  $N_{div}(nP)$  per un dato  $nP$ .

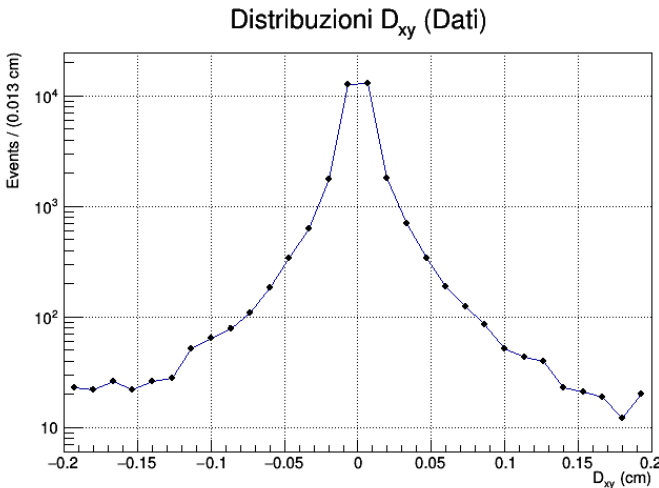
Per poter effettivamente stimare  $N_2$  tramite la stima di  $f_B$  seguendo lo stesso approccio adoperato per  $N_1$ , si sceglie di lavorare nell'assunzione in cui la popolazione di ciascun campione di muoni non dipenda da  $nP$ . In questo modo sarà possibile considerare  $f_B(nP) = f_B$  (costante) e quindi, ripercorrendo le analisi precedentemente svolte, giungere ad una formula esplicita per la stima di  $N_2$  a partire dai dati.

In particolare, effettuando una estrapolazione a  $nP = 1$ , ovvero considerando  $nEv(1) = nEv_{tot}$  ( $= \sum_{nP} nEv(nP)$ ) si ottiene:

$$N_2 = nEv_{tot} \xi(nP = 1) f_B \quad (3.6)$$

con incertezza ricavata tramite propagazione opportuna per il modello considerato.

Ottenuta quindi una formula esplicita per la stima di  $N_2$  a partire dai dati, è necessario procedere con la determinazione della frazione di muoni da  $B$  nel campione considerato<sup>3</sup> seguendo lo stesso schema già sfruttato per  $N_1$ .



**Figura 3.5:** Distribuzioni della proiezione trasversa del parametro d'impatto per i muoni da diversa collisione. I bin agli estremi raccolgono gli overflow

plicato (risultati in figura 3.6). Quest'ultimo, così come nel caso di  $N_1$ , produce risultati giudicati plausibili se confrontati ai valori di riferimento del Monte Carlo, ottenendo:

$$f_B = 0.088 \pm 0.016$$

<sup>3</sup>Muoni da diversa collisione

Dunque operativamente, dopo aver ricavato dai dati la distribuzione del  $D_{xy}$  (figura 3.5) per i muoni da diversa collisione, è stato costruito il modello complessivo di segnale più fondi con cui effettuare il fit a partire dalle distribuzioni Monte Carlo ricavate nel capitolo precedente (figura 2.6).

Ripercorrendo poi tutte le considerazioni sulla stabilità delle stime già adottate nello studio di  $N_1$ , dopo un primo tentativo non riuscito di fit con il modello completo (riportato in appendice in figura A.2), si è deciso di fissare le due frazioni relative alle categorie meno numerose (i.e. Else e Charm) al valore del Monte Carlo (tabella 2.2) e procedere con il fit di un modello semplificato.

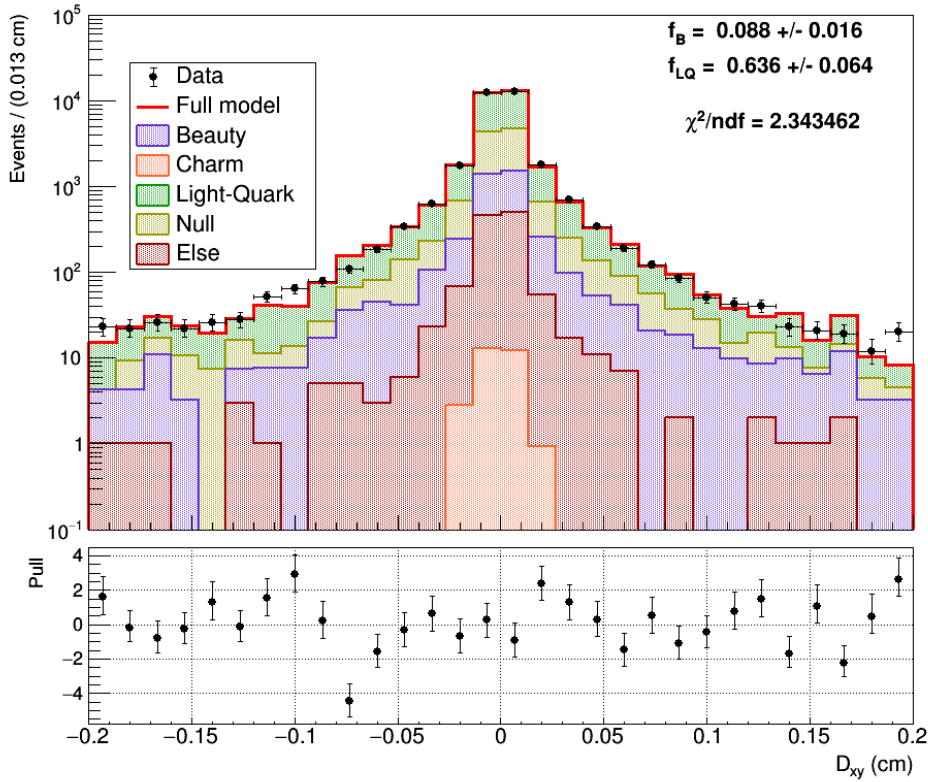
Si sottolinea come in questo caso, essendo la componente da  $B$  non maggioritaria, la precisione sulla stima di  $f_B$  risulti peggiore rispetto al caso di  $N_1$ .

Disponendo a questo punto di tutte le informazioni necessarie, tramite la formula [3.6](#), si riportano le stime di  $N_2$  al variare del modello  $\xi(nP)$  considerato:

Modello Estrapolazione	$N_2$
$m nP$ ( $nP \leq 30$ )	$(1.5 \pm 0.3) \times 10^2$
$m nP a^{c nP}$	$(1.5 \pm 0.3) \times 10^2$

**Tabella 3.1:** Stime di  $N_2$  al variare del modello considerato.

Come atteso, i due modelli producono risultati del tutto equivalenti alla luce della loro somiglianza per piccoli valori di  $nP$ .



**Figura 3.6:** Fit con modello semplificato della distribuzione della proiezione trasversa del parametro d'impatto per la stima di  $N_2$ . Le componenti Charm ed Else sono fissate al valore ricavato dal Monte Carlo mentre la frazione da Null non riportata si intende determinata per normalizzazione a 1.

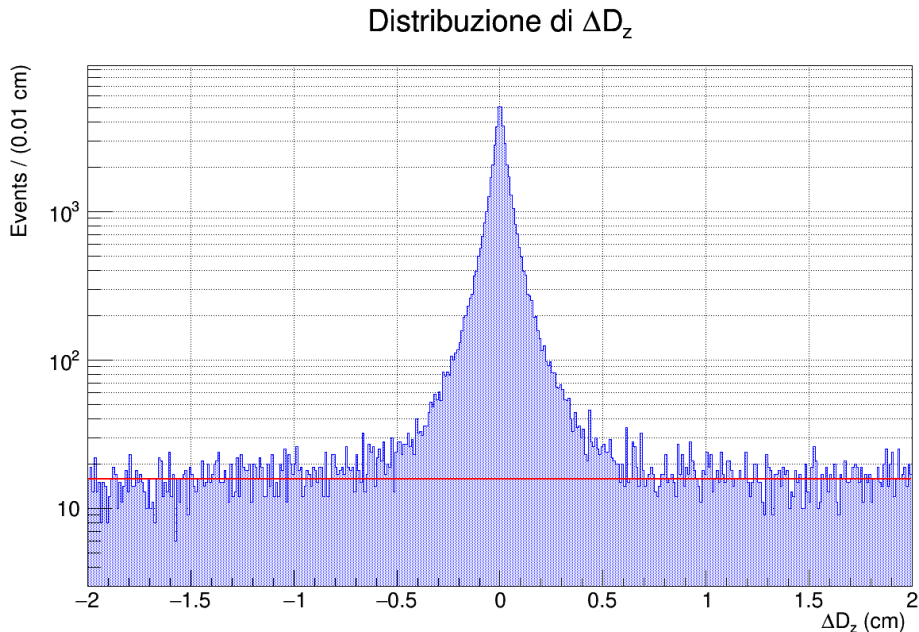
# Capitolo 4

## Applicazione ai dati

In questo capitolo si effettua l'analisi<sup>1</sup> su un campione raccolto nell'ambito dell'esperimento CMS ad LHC, giungendo ad una stima della sezione d'urto di produzione di coppie  $b\bar{b}$ .

### 4.1 Stima della sezione d'urto

Appurata la fattibilità dell'analisi tramite il test Monte Carlo su Monte Carlo effettuato nei capitoli precedenti, si può procedere con un'effettiva misura della sezione d'urto  $\sigma(pp \rightarrow b\bar{b} X)$  applicando ad un campione reale l'intero approccio sviluppato.



**Figura 4.1:** Distribuzione di  $\Delta D_z$  per i dati CMS; in rosso, modello costante per stimare il fondo di eventi da diversa collisione.

Seguendo gli step già percorsi, si è in primo luogo verificato che le distribuzioni della differenza  $\Delta D_z$  ricavate a partire dai dati consentissero di mantenere le definizioni di *stessa/diversa* collisione adoperate per il Monte Carlo e si è stimata l'entità del fondo tramite fit costante per

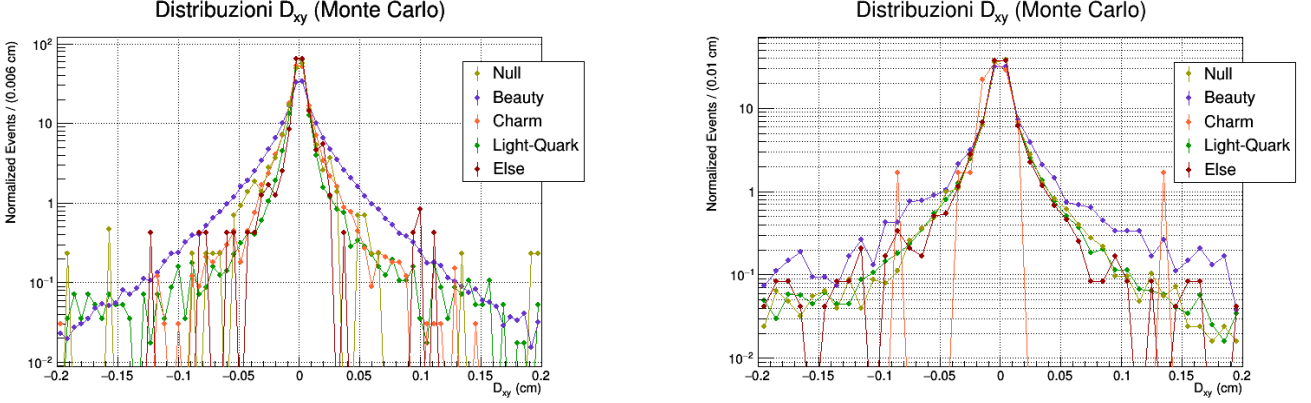
---

<sup>1</sup>A tal proposito, per evitare ripetizioni, si riportano solamente i passi salienti dell'intera procedura, essendo essa del tutto analoga a quanto descritto nei capitoli precedenti.

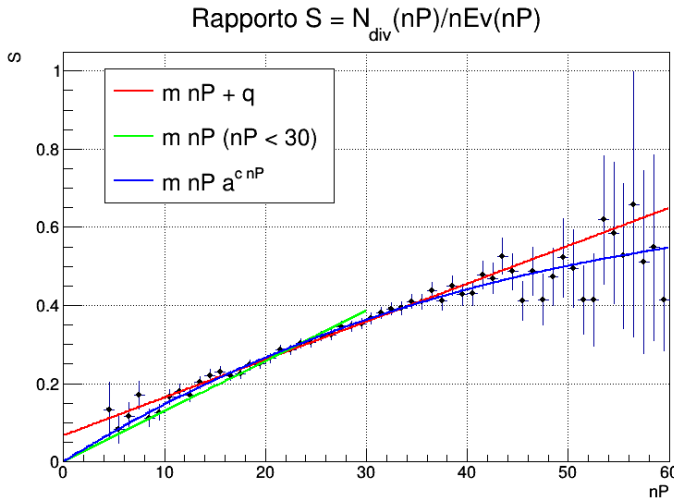
poter ricavare le due quantità necessarie nelle analisi successive (distribuzione e interpolazione riportati in figura 4.1). In particolare, si ottiene:

$$C_{div} = 15.7 \pm 0.2 \text{ Events} / (0.01 \text{ cm}) \quad \epsilon_{div} = (20.2 \pm 0.3) \times 10^{-3} \quad \epsilon_{side-band} = 0.097 \pm 0.001$$

Successivamente, per giungere alla stima di  $N_1$  ed  $N_2$ , sono state ricavate a partire dal Monte Carlo le distribuzioni per la componente trasversa  $D_{xy}$  del parametro d'impatto per i muoni di entrambi i sotto-campioni considerati (figura 4.2).



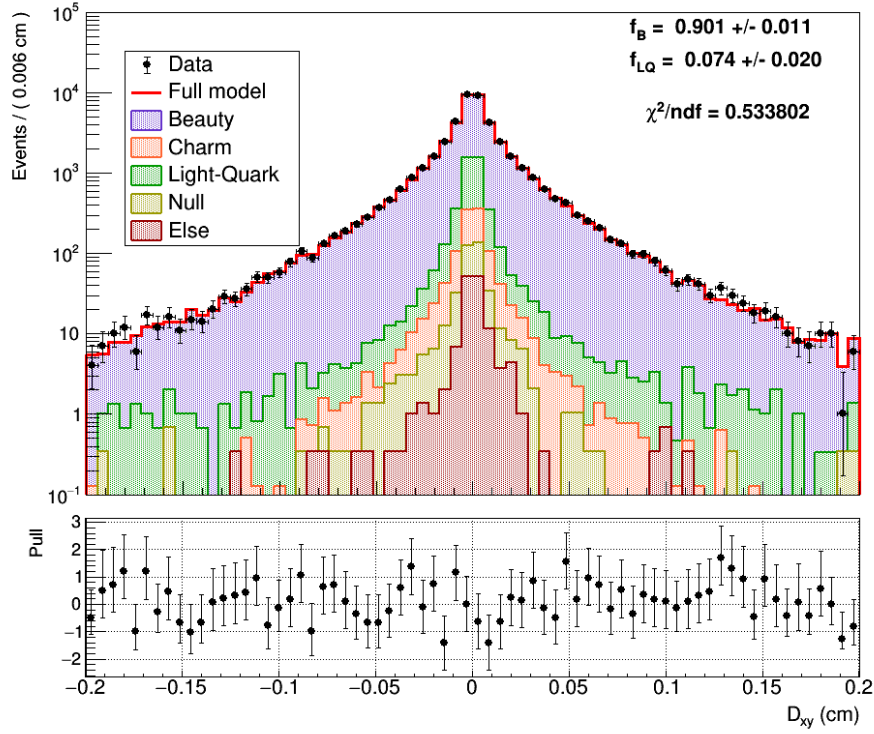
**Figura 4.2:** Distribuzioni Monte Carlo della proiezione trasversa del parametro d'impatto per ciascuna categoria, normalizzate all'unità. A sinistra eventi con muone dalla stessa collisione del  $B^+$ , a destra quelli da vertice differente. I bin agli estremi raccolgono gli overflow.



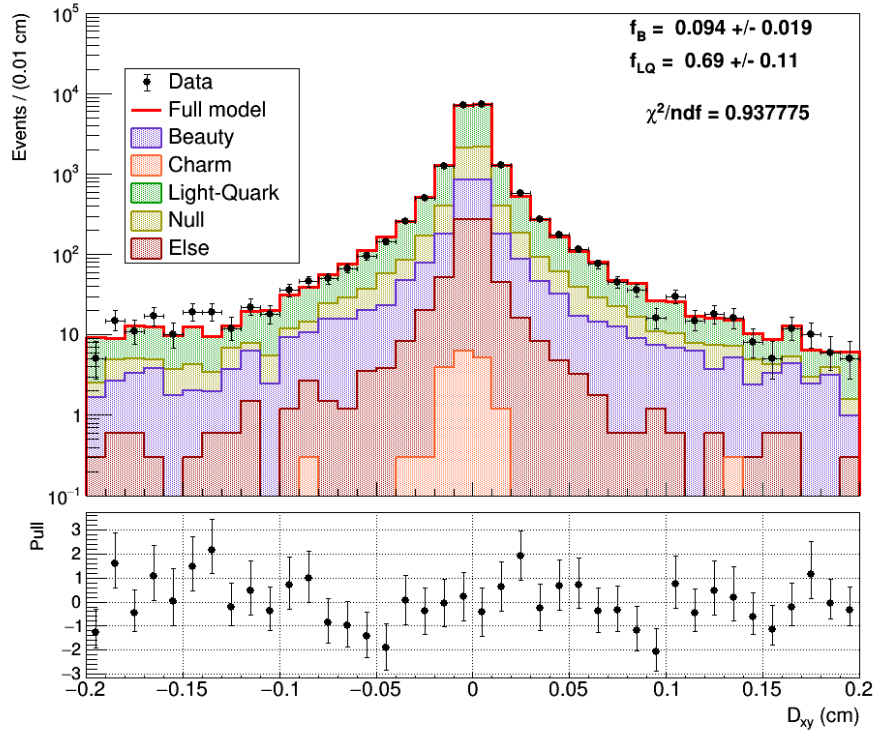
**Figura 4.3:** Andamento del rapporto  $S(nP)$  in funzione del numero di vertici di pile-up  $nP$  ricavato dai dati. Si riportano, inoltre, l'interpolazione lineare a due parametri (rosso), l'interpolazione lineare ad intercetta fissata e con dominio limitato (verde) e il modello con correzione esponenziale (blu).

Inoltre, per poter effettuare correttamente l'estrapolazione ad  $nP = 1$  sviluppata per  $N_2$ , è stato costruito il rapporto  $S(nP)$  definito in 3.4 e sono state effettuate le dovute interpolazioni (figura 4.3). Nello specifico, si sottolinea come anche in questo caso l'interpolazione lineare non riesca a descrivere perfettamente l'intero andamento di  $S(nP)$ , presentando lo stesso comportamento anomalo per alto numero di collisioni già registrato nell'analisi del Monte Carlo. Analogamente a quanto già svolto, allora, si sfruttano gli stessi modelli per descrivere l'andamento osservato dei dati (confronto tra i risultati del fit sui dati e sul Monte Carlo riportato in appendice in figura B.1).

Disponendo di tutte le informazioni necessarie, sono stati effettuati i fit del Monte Carlo sui dati giungendo così ad una stima della frazione di muoni da  $B$  per gli eventi da stessa e diversa collisione. In particolare, così come avvenuto nel caso dei test sulle simulazioni, i fit sono stati effettuati costruendo un modello semplificato, ovvero fissando le componenti meno significative al valore del Monte Carlo. Si riportano i risultati in figura 4.4 e 4.5.



**Figura 4.4:** Fit dei dati con modello semplificato della distribuzione della proiezione trasversa del parametro d'impatto per la stima di  $N_1$ . Le componenti Null ed Else sono fissate al valore ricavato dal Monte Carlo mentre la frazione da Charm non riportata si intende determinata per normalizzazione a 1.



**Figura 4.5:** Fit dei dati con modello semplificato della distribuzione della proiezione trasversa del parametro d'impatto per la stima di  $N_2$ . Le componenti Charm ed Else sono fissate al valore ricavato dal Monte Carlo mentre la frazione da Null non riportata si intende determinata per normalizzazione a 1.

Al termine della procedura di analisi, dunque, tramite le formule [3.1](#) e [3.6](#), si ottiene:

$$N_1 = (40.4 \pm 0.5) \times 10^3$$

Modello Estrapolazione	$N_2$
$m \text{ nP } (\text{nP} \leq 30)$	$(0.8 \pm 0.2) \times 10^2$
$m \text{ nP } a^{c \text{ nP}}$	$(1.0 \pm 0.2) \times 10^2$

Considerando allora la formula [1.2](#) introdotta nel primo capitolo e propagando correttamente l'incertezza, ricordando l'estrapolazione ad  $\text{nP} = 1$ , possiamo scrivere:

$$R = \frac{\sigma(pp \rightarrow b\bar{b}X)}{\sigma(pp \rightarrow X)} = \frac{1}{2\text{nP}} \frac{N_2}{N_1} \Big|_{\text{nP}=1} = \frac{N_2}{2N_1} \quad \sigma_R^2 \approx R^2 \left[ \left( \frac{N_1}{\sigma_{N_1}} \right)^2 + \left( \frac{N_2}{\sigma_{N_2}} \right)^2 \right] \quad (4.1)$$

ottenendo:

Modello Estrapolazione	$R$
$m \text{ nP } (\text{nP} \leq 30)$	$(1.0 \pm 0.2) \times 10^{-3}$
$m \text{ nP } a^{c \text{ nP}}$	$(1.2 \pm 0.2) \times 10^{-3}$

**Tabella 4.1:** Stime di  $R$  al variare del modello considerato.

Infine, sfruttando la misura di  $\sigma(pp \rightarrow X)$  a  $\sqrt{s} = 13$  TeV effettuata presso CMS [3](#), si giunge a

$$\sigma(pp \rightarrow b\bar{b}X) = \begin{cases} (69 \pm 14) \mu b & (m \text{ nP } (\text{nP} \leq 30)) \\ (85 \pm 17) \mu b & (m \text{ nP } a^{c \text{ nP}}) \end{cases} \quad (4.2)$$

Per confronto, si riporta la stima effettuata presso l'esperimento LHCb [5](#), pari a :

$$\sigma(pp \rightarrow b\bar{b}X) = (144 \pm 1 \pm 21) \mu b$$

Entrambi i valori ottenuti a partire dai due modelli considerati si discostano da quest'ultima per meno di tre deviazioni standard.



# Conclusioni

In questa tesi è stato sviluppato e testato un nuovo approccio per la stima della sezione d'urto di produzione  $pp \rightarrow b\bar{b} X$  con ridotta dipendenza dal modello.

Dopo aver introdotto la strategia di misura e le equazioni associate, sono state effettuate analisi preparatorie volte alla determinazione di criteri utili alla separazione dei dati in sotto-campioni tramite l'utilizzo del parametro d'impatto. L'intera analisi dati è stata poi svolta su un campione Monte Carlo, verificandone la fattibilità. In seguito, dopo aver testato l'approccio sulla simulazione, è stato possibile applicare l'intera procedura d'analisi ad un campione di misure raccolte presso l'esperimento CMS ad LHC, ottenendo una stima di  $\sigma(pp \rightarrow b\bar{b} X)$  da confrontare con altre misure indipendenti.

Per poter migliorare quanto svolto e rendere più solido l'intero approccio sviluppato, sarà fondamentale sviluppare una nuova strategia di estrapolazione per i muoni provenienti da diversa collisione, effettuando ulteriori analisi volte allo studio delle anomalie riscontrate per eventi ad alto numero di collisioni ricostruite.

Sarà inoltre possibile individuare nuove variabili discriminanti da affiancare al parametro d'impatto per consentire una più efficace separazione del segnale dai fondi, migliorando la precisione della misura.

Tutte queste questioni aperte rappresentano un'interessante sfida per lavori futuri.

# Appendice A

## Fit con modello completo per la stima di $N_1$ ed $N_2$

Si riportano i tentativi di fit con modello completo dato dal segnale da B più le quattro tipologie di fondo (Null, Else, Charm, Light Quark) sia per la stima di  $N_1$  che di  $N_2$  effettuati nell'ambito dei test Monte Carlo su Monte Carlo.

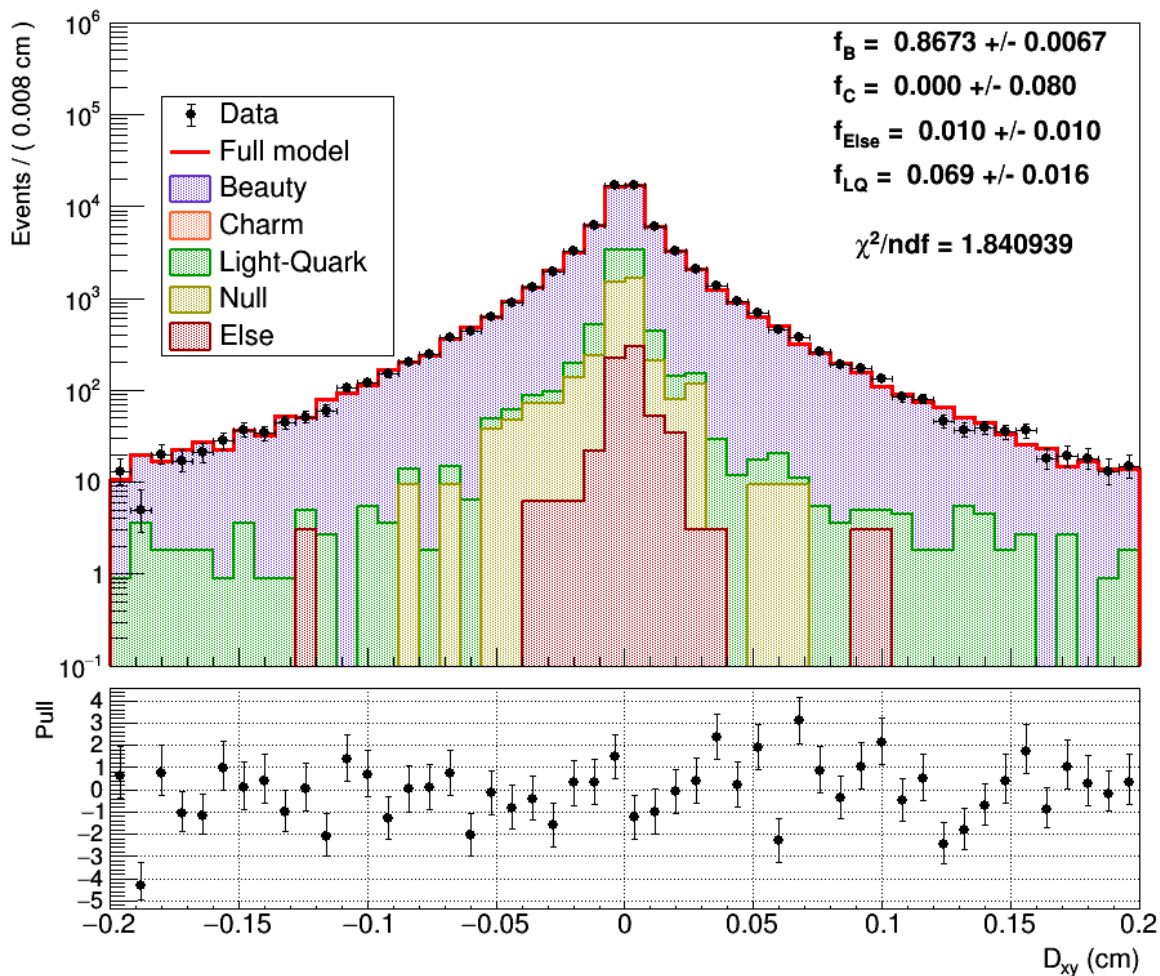
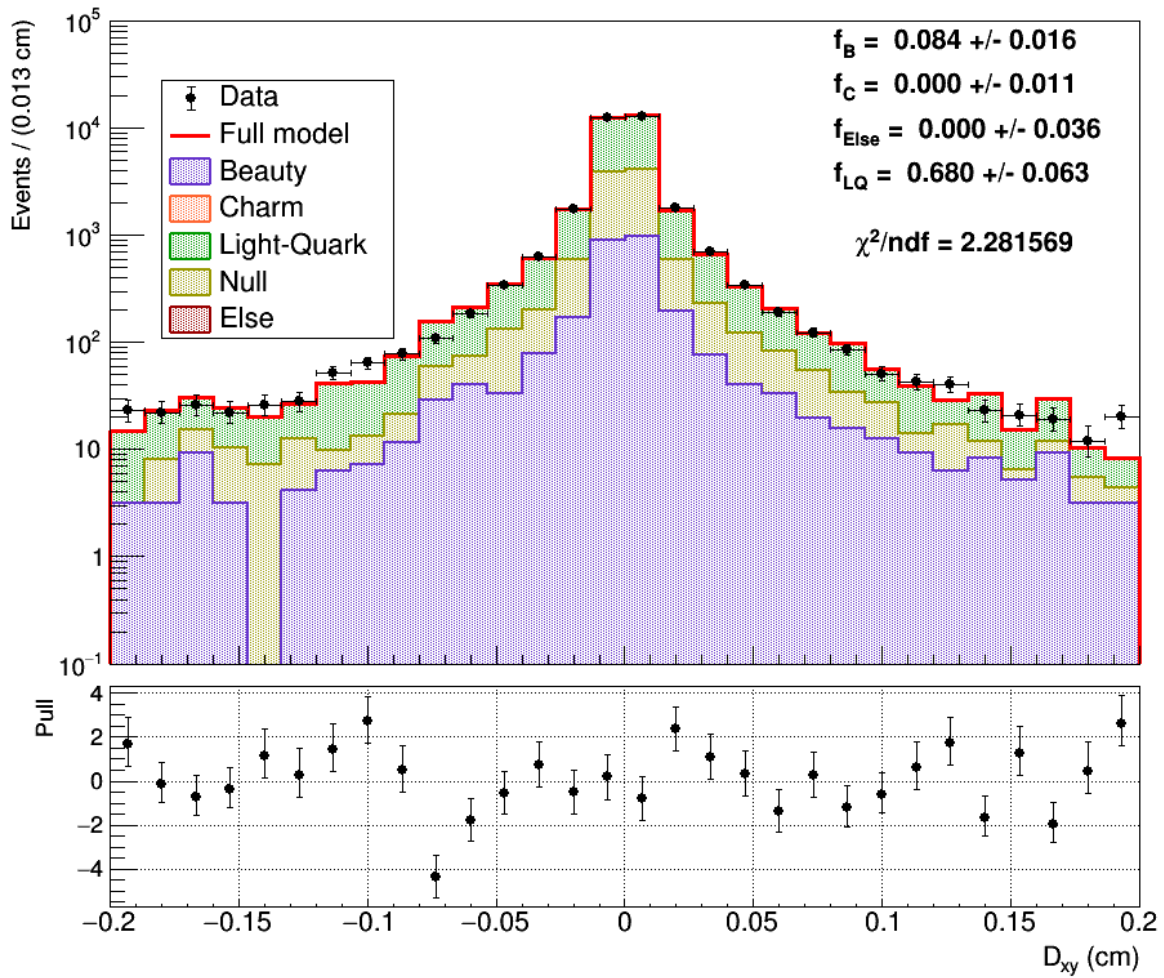


Figura A.1: Fit con modello completo della distribuzione della proiezione trasversa del parametro d'impatto per la stima di  $N_1$ ; la quinta frazione non riportata si intende determinata per normalizzazione a 1.

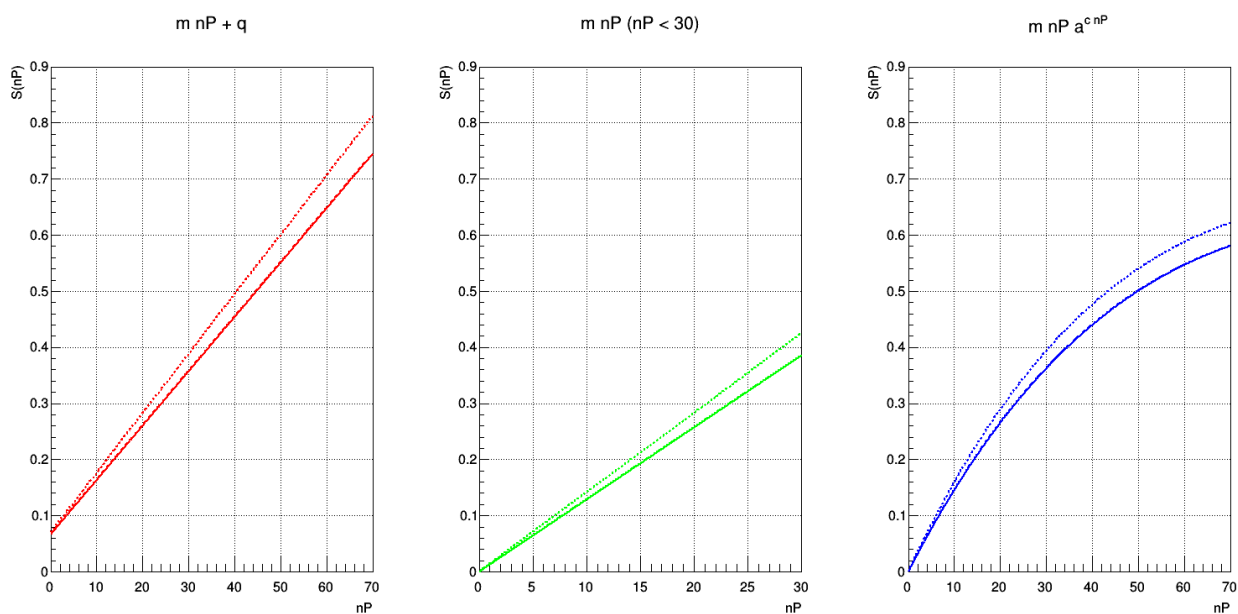


**Figura A.2:** Fit con modello completo della distribuzione della proiezione trasversa del parametro d'impatto per la stima di  $N_2$ ; la quinta frazione non riportata si intende determinata per normalizzazione a 1.

# Appendice B

## Confronto delle estrapolazioni effettuate sul Monte Carlo e sui Dati

Si riporta un confronto tra i fit del rapporto  $S(nP)$  effettuati sul campione Monte Carlo (tratteggiato) e sui dati (linea continua) per evidenziare eventuali differenze tra dati reali e simulazione.



**Figura B.1:** Risultati di ciascuna interpolazione effettuata sul rapporto  $S(nP)$  sul campione Monte Carlo (linea tratteggiata) e sui dati (linea continua).

# Bibliografia

- [1] **Chatrchyan, S. and others.** The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008. [doi:10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [2] **CMS Collaboration.** Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *Journal of Instrumentation*, 13(05):P05011–P05011, may 2018. [doi:10.1088/1748-0221/13/05/p05011](https://doi.org/10.1088/1748-0221/13/05/p05011).
- [3] **CMS Collaboration.** Measurement of the inelastic proton-proton cross section at  $\sqrt{s} = 13\text{TeV}$ . *Journal of High Energy Physics*, 2018. [doi:10.1007/JHEP07\(2018\)161](https://doi.org/10.1007/JHEP07(2018)161).
- [4] **Enrico Lusiani.** Measurement of  $b\bar{b}$  production in  $s = 13\text{ TeV}$  pp collisions with the CMS experiment. *Università degli Studi di Padova*, A.A. 2018/2019. URL: <http://hdl.handle.net/20.500.12608/27755>.
- [5] **LHCb Collaboration.** Measurement of the  $b$ -quark production cross section in 7 and 13 TeV  $pp$  collisions. *Phys. Rev. Lett.*, 118:052002, Feb 2017. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.118.052002>, [doi:10.1103/PhysRevLett.118.052002](https://doi.org/10.1103/PhysRevLett.118.052002).