

## Prossimi passi per misura W/Z:

1. Stato attuale
2. Campioni MC
3. Ntuple
4. Versione di Athena
5. Efficienze
6. Risoluzioni
7. PDF
8. Convoluzione risoluzioni/efficienze con segnale MC

## Stato attuale (in ordine sparso):

- 1) Selezione Z: Cosenza + Pavia
- 2) Efficienze trk+trig: Pavia + Roma1 + Frascati + Roma2
- 3) Scala Emiss: Milano
- 4) PDF: Roma2
- 5) Risoluzioni scale di impulso: Frascati+Roma3

- Tra le cose principali rimane scoperta la selezione del W e l'efficienza delle tracce dell'ID.
- Studi di fondi ecc. ecc. probabilmente arriveranno in un secondo momento

## Dalla discussione di martedì siamo giunti ad almeno tre conclusioni:

- 1) Necessità di avere NTUPLE per permettere di lavorare agli 'over-40' che non digeriscono Athena, ma anche ai laureandi.
- 2) Cominciare a mettere insieme i vari pezzi.
- 3) Eventuale produzione locale di sample MC, eventualmente in maniera ufficiale sui Tier2.

## Campioni MC

Direi come obiettivo di riprodurre una misura con almeno una statistica pari a  $100 \text{ pb}^{-1}$ .

Segnale:

$Z \rightarrow \mu\mu$ : 150 kevts

$W \rightarrow \mu\nu$ : 1,3 Mevts

Low-Mass  $\mu\mu$ : 3 Mevt

Alcuni fondi:

$Z \rightarrow \tau\tau$ : 150 kevts

$W \rightarrow \tau\nu$ : 1,3 Mevts

Low-Mass  $\tau\tau$ : 3 Mevts

$t\bar{t}$ : 15 kevts

heavyQ (bb, cc)( $\mu 20X$ ):  $2 \times 13$  kevts (13 Mevts, errore a causa di tabelle errate su sito web di FDR2)

Jets (Dijets  $J \geq 3 > 70$  GeV,  $\sim 6\mu\text{b}$ ): 600 Mevts (!)

Tot evets (esclusi jets)  $\sim 10$  Mevts  $\rightarrow$  2 Gsec/CPU (gen+sim+rec)  $\sim 23$  kday/CPU (1 mese con 250 CPU/tier2?)

Jets con ATLFast? Se si guadagna un fattore 10 in tempo si potrebbe anche fare.

## Spazio disco (AOD)

10 Mevts  $\rightarrow$   $150 \text{ kB} \times 10^7 = 1.5 \text{ TB}$

per questi sui tier2 non credo ci siano problemi, è previsto fino ad 1 TB per user (mi pare).

600 Mevts  $\rightarrow$  90 TB mmmm..., si potrebbero tenere però solo quelli con almeno un trigger di  $\mu$ . Allora sarebbe almeno un fattore  $10^3$  di meno.

## Tempo di processamento

Quanto impiega Athena a girare sugli AOD? 30 ms/evt (a occhio e croce)? In tal caso servirebbero 3-4 giorni/job per girare su tutto il campione MC. Se ogni volta che si vuole aggiungere una variabile ad un'entupla bisogna aspettare 3-4 giorni non si finisce più. **DPD necessari per ridurre il tempo di processamento.** Prendendo solo eventi preselezionati (1  $\mu$ 20 almeno) si potrebbe arrivare a 2 Mevts riducendo di un fattore 5 solo per il numero di eventi.

## MC data-like

Come 'dati' potremmo usare MC scalibrati. Potremmo anche prendere campioni MC per i vari samples con rapporto di numero di eventi lievemente diverso rispetto a quello del campione MC vero e proprio.

## Ntuple

Vanno discusse le informazioni da mettere nell'entupla.

Le entuple devono andare bene per:

- studi di selezione
- studi di efficienze
- studi di calibrazione

Possibilmente vanno inserite le informazione sul match dei  $\mu$  con il trigger e con la verità MC.

Va salvata l'informazione su quale pre-selezione hanno passato gli eventi?

Soddisfano la selezione dell Z? Soddisfano quella del W?

Per i  $\mu$  direi che servono tutte le informazioni, comprese le matrici di covarianza, e per tutti i tipi di ricostruzione (SA, MUID, STACO, LOW-Pt, Calo). Cerchiamo di fare in modo che sia facile passare mboy a moore senza dover duplicare il codice.

Per le tracce dell'ID potremmo salvare solo carica impulso e parametro di impatto ed eventuale match con un  $\mu$ .

Idee per jets, taus, elettroni, Etmis?

**Importante cominciare anche solo con una versione preliminare.**

Bisognerebbe anche organizzare le informazioni in blocchi in modo che sia possibile ‘accendere’ e ‘spegnere’ blocchi di informazioni da python e sia altrettanto semplice definire e riempire nuovi blocchi.

Potremmo anche creare entuple corrispondenti a runs in modo che sia facile mescolare i vari canali nel rapporto corretto.

## Spazio disco

Riducendo gli eventi a 2 Mevts e riducendo la dimensione degli eventi dovremmo poter mettere tutti gli eventi in max 100 GB, ma magari anche meno. Comunque una cosa gestibile anche su un portatile.

Le entuple possono anche essere ridotte a mini-entuple in un secondo momento con root per velocizzare l’analisi.

## Versione di Athena?

Con alcune delle ultime versioni di Athena abbiamo avuto problemi con la configurazione dell'HLT. Dovremmo comunque scegliere una versione accettabile e cercare di rimanere su quella e minimizzare le continue migrazioni. Proposte?

## Efficienze

Andrebbero prodotti istogrammi 2-D (pt,eta) delle seguenti efficienze:

TRK+cuts

TRK+cuts+TRG

per MS e ID, per i vari algoritmi (Staco, Muid)

e sia per i sample MC (con tutti i fondi) che per i sample MC data-like.

I tagli aggiuntivi (cuts) vanno tunati in base alla selezione.

C'è una discussione in corso poi sul formato. Si veda la nota di M.Schott a riguardo.

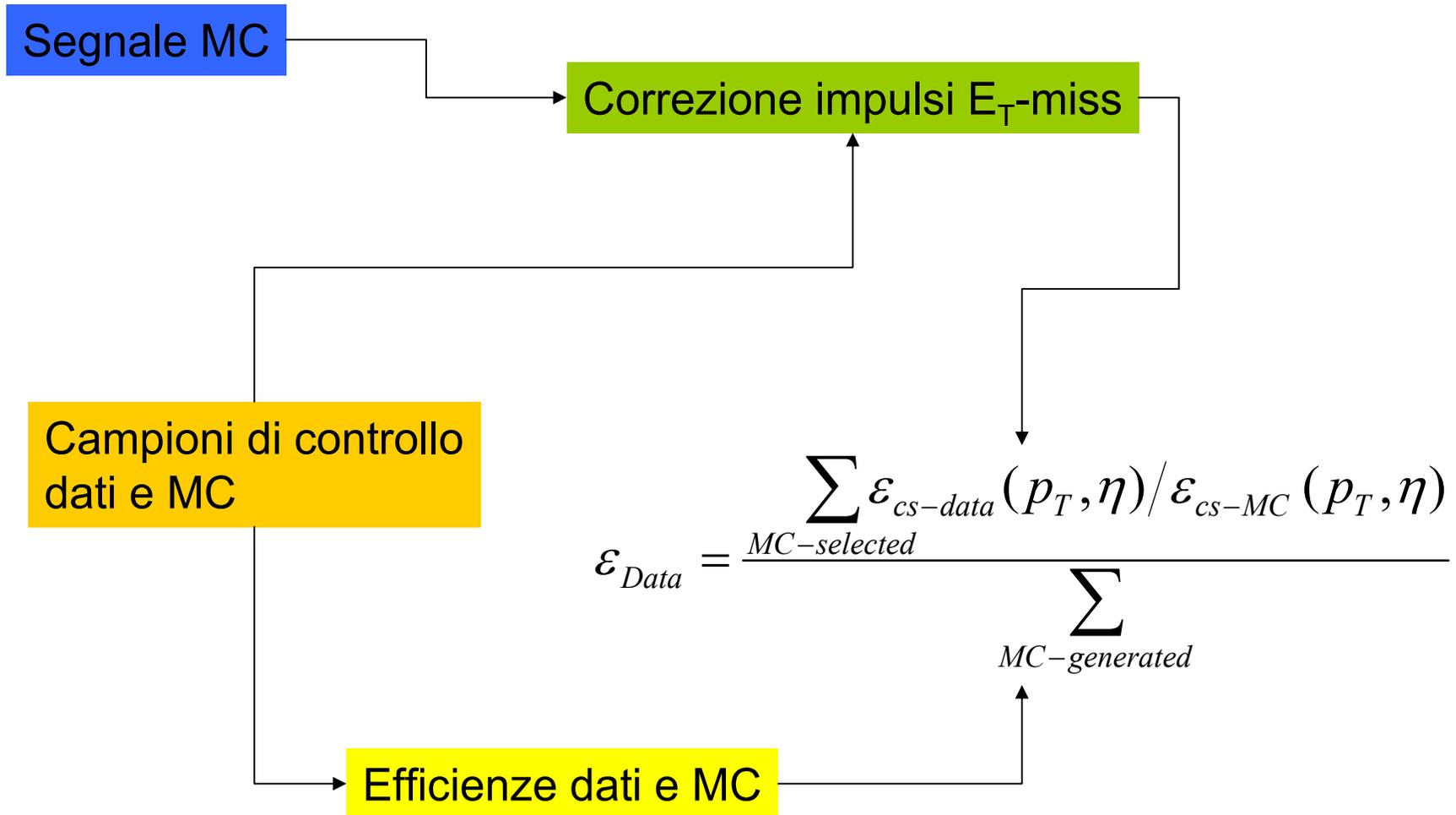
## Risoluzioni

Vanno fornite funzioni di correzione con relativi parametri per correggere gli impulsi (e missing Et) nel MC per riprodurre i dati.

## PDF

Sarebbe interessante avere le distribuzioni ( $p_T$  e eta di W e Z) di eventi MC generati con diverse PDF, o generatori, per ripesare quelle del MC ricostruito e vedere l'effetto che ha sulle efficienze.

## Accettazione da MC corretto



## Considerazioni finali

L'altro giorno durante la riunione di ATLAS Italia si è discusso del fatto che le analisi debbano girare su AOD e non su entupla.

Personalmente ritengo che sia importante comunque sviluppare l'analisi, anche su entupla, e poi in un secondo momento, con l'aiuto di alcuni "volenterosi", portare l'analisi su Athena.

Credo inoltre che se si vogliono processare non  $100 \text{ pb}^{-1}$  ma  $10\text{-}100 \text{ fb}^{-1}$  il modello "Ntupla" non regga, e che si debba quindi prima o poi automatizzare gran parte dell'analisi su Athena (vedi discussione fatta a Milano).

## MC samples: update

Files mc08 tipo e323\_s400\_d99

Parlo di digit, i recon (AOD) sono molto pochi

### Segnale

Channel	Dataset name	Task ID	Events
$Z \rightarrow \mu\mu$	mc08.005145.PythiaZmumu.digit.e323_s400_d99	21821	$5 \times 10^5$
$W \rightarrow \mu\nu$	mc08.005105.PythiaWmunu.digit.e323_s400_d99	21971 21819	$5 \times 10^5$ $5 \times 10^5$
LowMass $\mu\mu$	mc08.005155.PythiaDYanLowMmumu.digit.e323_s400_d99	21839 21838	$5 \times 10^5$ $5 \times 10^5$

## Fondi: EW+ttbar

Channel	Dataset name	Task ID	Events
$Z \rightarrow \tau\tau$	mc08.006903.PythiaZtautau.digit.e323_s400_d99	21822	$5 \times 10^5$
$W \rightarrow \tau\nu$	NONE		
LowMass $\tau\tau$	mc08.005156.PythiaDYanLowMtautau.digit.e323_s400_d99	21841 21840	$5 \times 10^5$ $5 \times 10^5$
ttbar	mc08.005568.ttbar_Pythia.digit.e323_s400_d99	21835	$5 \times 10^5$

## Fondi: heavy Q

Channel	Dataset name	Task ID	Events
bb $\mu$ 9X	mc08.017509.PythiaB_bbm9x.digit.e323_s400_d99	21852	$5 \times 10^5$
		21851	$5 \times 10^5$
bb $\mu$ 18X	mc08.017510.PythiaB_bbm18x.digit.e323_s400_d99	21881	$5 \times 10^5$
cc $\mu$ 9X	mc08.017524.PythiaB_ccm9x.digit.e323_s400_d99	21854	$5 \times 10^5$
		21853	$5 \times 10^5$
cc $\mu$ 18X	mc08.017525.PythiaB_ccm9x.digit.e323_s400_d99	21882	$5 \times 10^5$

## Fondi: Jets

channel	Dataset name	Task ID	Events
J3 (70-140 GeV)	mc08.005012.j3_pythia_jetjet.digit.e323_s400_d99	21810	$5 \times 10^5$
		21809	$5 \times 10^5$
		21808	$5 \times 10^5$
		21807	$5 \times 10^5$
J4 (140-280 GeV)	mc08.005013.j4_pythia_jetjet.digit.e323_s400_d99	21812	$5 \times 10^5$
		21811	$5 \times 10^5$
J5 (280-560 GeV)	mc08.005014.j5_pythia_jetjet.digit.e323_s400_d99	21813	$5 \times 10^5$
J6 (560-1120 GeV)	mc08.005015.j6_pythia_jetjet.digit.e323_s400_d99	21814	$5 \times 10^5$
J7 (1120-2240 GeV)	mc08.005016.j7_pythia_jetjet.digit.e323_s400_d99	21815	$5 \times 10^5$

## Riepilogo:

Processo	Necessari (100 pb <sup>-1</sup> )	Simulati (mc08 s400_d99)	Da produrre
Z → μμ	150 k	0.5 M	0
W → μν	1.3 M	1 M	0.3 M
DY(μμ)	3 M	0.5 M	2.5 M
Z → ττ	1.5 k	0.5 M	0
W → τν	1.3 M × BR(mu)	0	1.3 M
DY(ττ)	3 M × BR(mu)	1 M	2 M
ttbar	15 k	0.5 M	0
bbmu9X	166 M	1 M	-----
bbmu18X	14 M	0.5 M	14 M
ccmu9X	~150 M (?)	1 M	-----
ccmu18X	~ 14 M (?)	0.5 M	14 M
J3	588 M	2 M	600 M
J4	31 M	1M	
J5	1.3 M	0.5 M	
J6	36 k	0.5 M	
J7	0.5 k	0.5 M	

## Conclusioni

Per  $100 \text{ pb}^{-1}$  occorrono circa 30 M events full simulation + 600 M events di jets con ATLfast.

**Sui Tier2 italiani richiede:**  $3 \text{ min/evt/CPU} \times 3 \cdot 10^7 \text{ events} = 60 \times 10^3 \text{ giorni/CPU}$   
assumendo circa  $10^3$  full-time sui tier2  $\sim 60$  giorni =  $2 \pm 1$  mesi

Altrettanto (x2) richiederebbero i 600 M eventi di jets con ATLfast: un fattore 10 in meno per simulare ma un fattore 20 in più di eventi.

Allo stato attuale si ha una statistica equivalente di pochi  $\text{pb}^{-1}$ .

Realisticamente in un mese potremmo avere circa  $20 \text{ pb}^{-1}$  equivalenti full simulation + magari  $jN(N>3)$  con ATLFast.

ps: dalla frammentazione mi aspetto  $E_{\mu\text{on}} = E_{\text{Jet}} \times 0.5(\text{leading fragment}) \times 0.5(\pi \text{ decay}) = E_{\text{Jet}}/4$ , per cominciare da J4 ( $E > 140 \text{ GeV}$ ) in su potrebbe bastare ( $E_{\mu} = 35 \text{ GeV}$ ). Allora circa 30 M eventi con ATLfast è poca roba.

pps: Ai tempi di simulazione va aggiunto un 10% di tempo per la ricostruzione.

ppps: Lo spazio disco stimato è per AOD. Per tenere RDO bisogna almeno moltiplicare per un fattore 10-20.