

B-id status

S. Greder

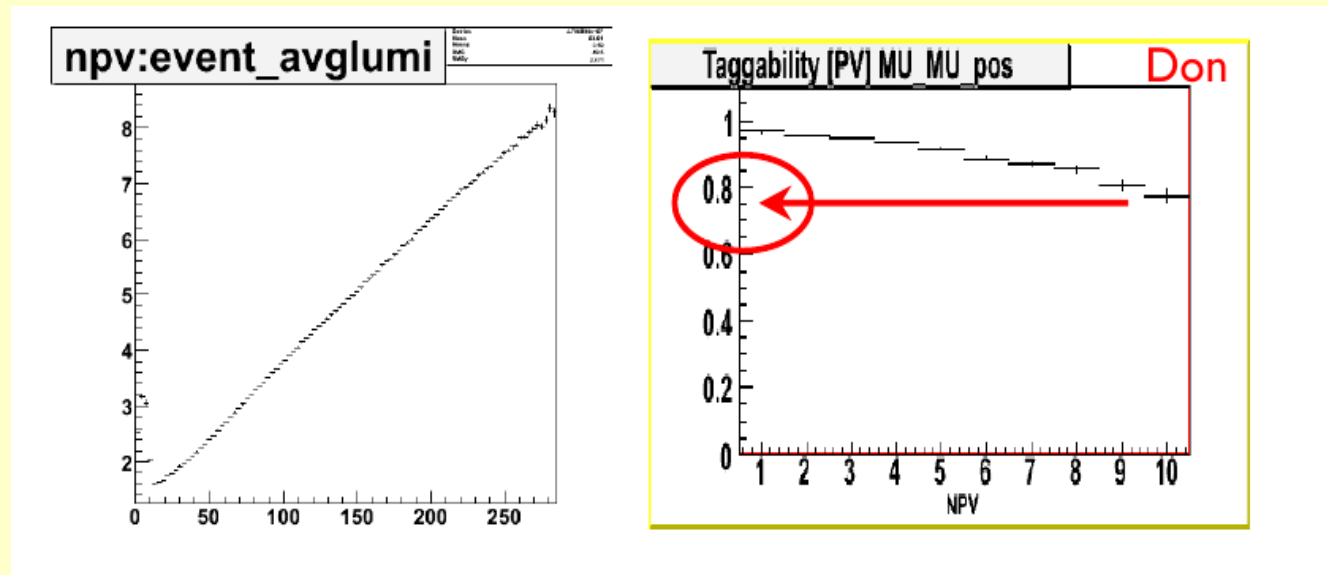
Institut Pluridisciplinaire Hubert Curien, Strasbourg

Outline

- B-id status
 - Taggabilité
 - Framework de certification
 - Analyse des données récentes (p12.12.00/01/02)
- SLT Neural Net
- Discriminant multivarie b/c

Taggabilité

- Dépend du lot de données analysées
- Dépend de la **luminosité** (= nombre de vertex d'interaction):



- Dépendance avec **cpf0** ? (vertex min. bias)

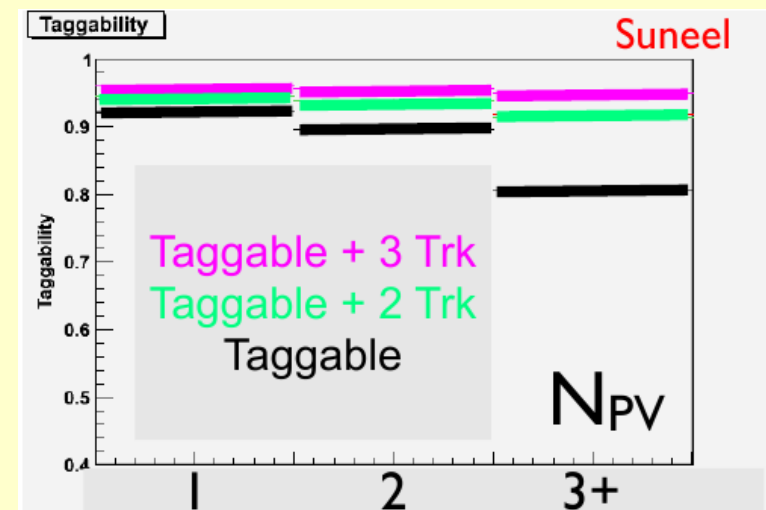
Taggabilité

- Nouvelle définition de la taggabilité:

$$\text{Taggability} = \frac{\text{Taggable} \oplus \text{Good}}{\text{Good}}$$

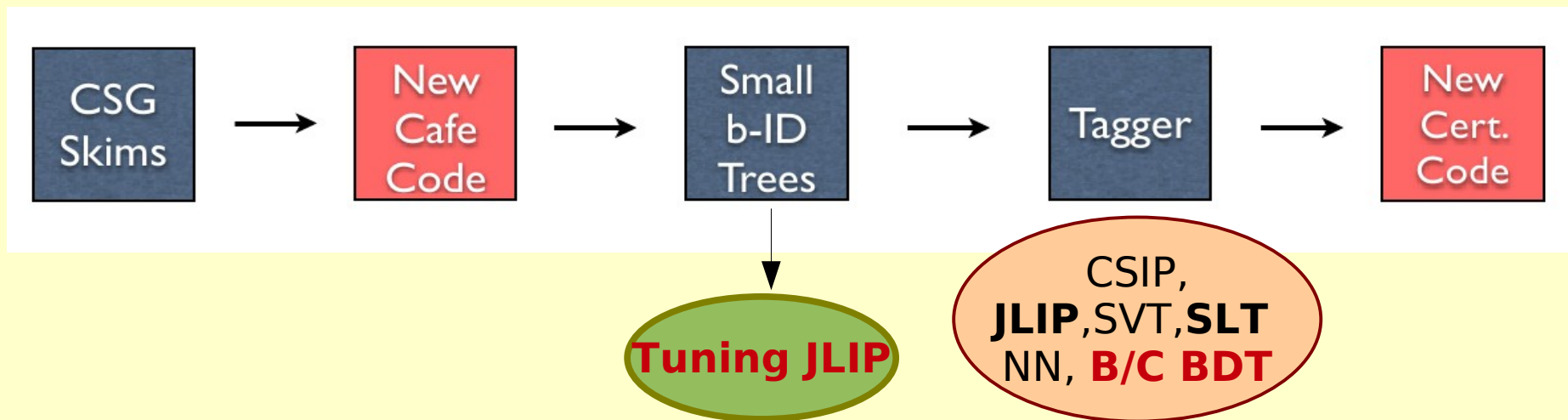
$$\text{Taggability} = \frac{\text{Taggable} \oplus \text{Good} \oplus \text{Ntrack}}{\text{Good} \oplus \text{Ntrack}}$$

- Ou Ntrack représente le nombre de trace utilisé pour le calcul de **cpf0** (*TMBJet::ntrkMultiplicity0()*)
- La dépendance en luminosité est réduite de **13%** (défaut) a **3(1)%** pour $\text{Ntrack} \geq 2(3)$



Framework de certification

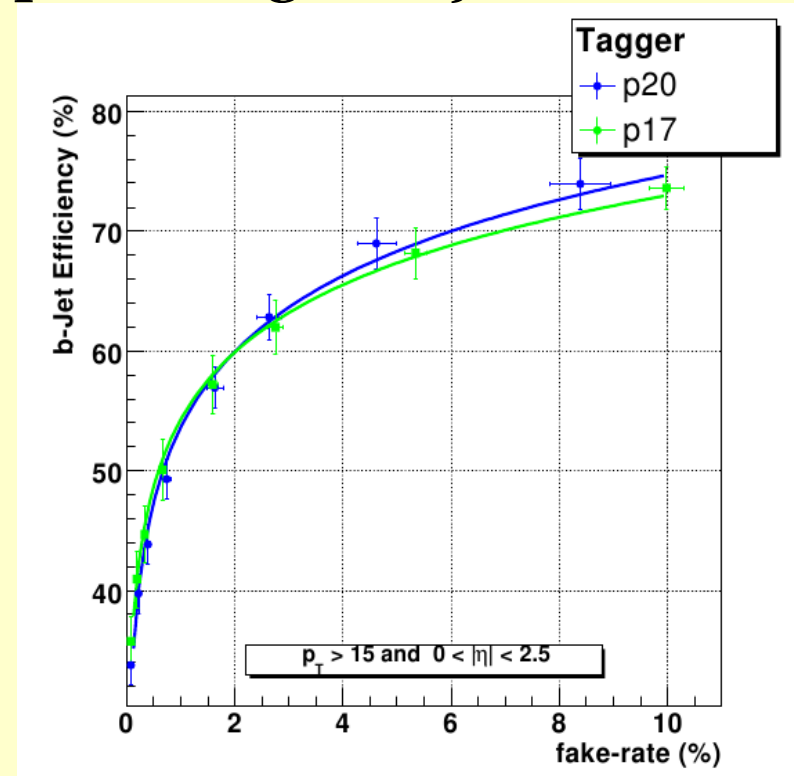
- Un nouveau :p framework de certification a été mis en place (TG) afin de centraliser l'étude, la certification et la maintenance du code b-id :



- Instructions Moriond08/ICHEP08 documentées ici:
 - <https://plone4.fnal.gov/P1/D0Wiki/object-id/bid/p20ICHEPInstructions>
 - <https://plone4.fnal.gov/P1/D0Wiki/object-id/bid/p20MORIOND08Instructions>

Dataset summer'08

- Les performances NN p20 == p17 malgré l'ajout layer0 ... :(
- ... bug sur l'occupance ?

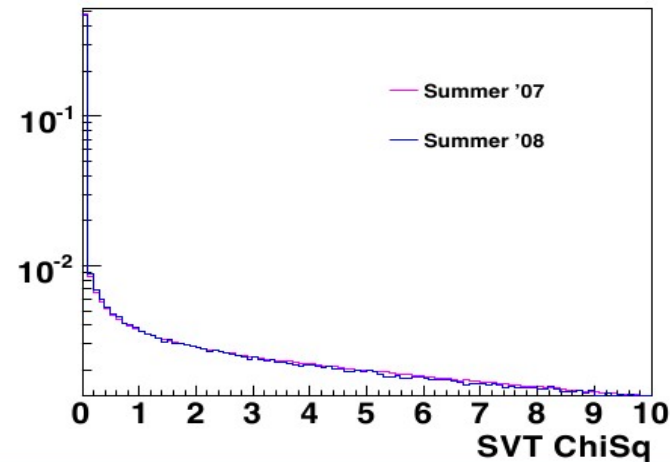
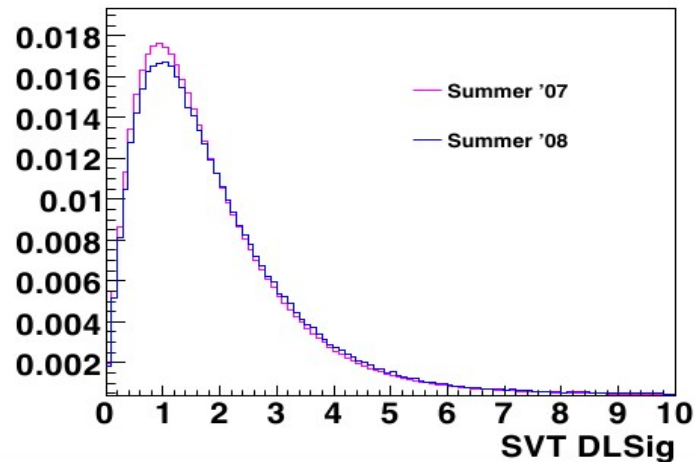
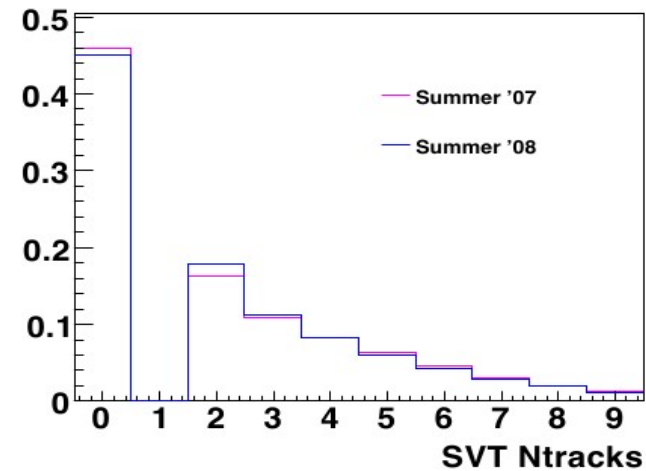
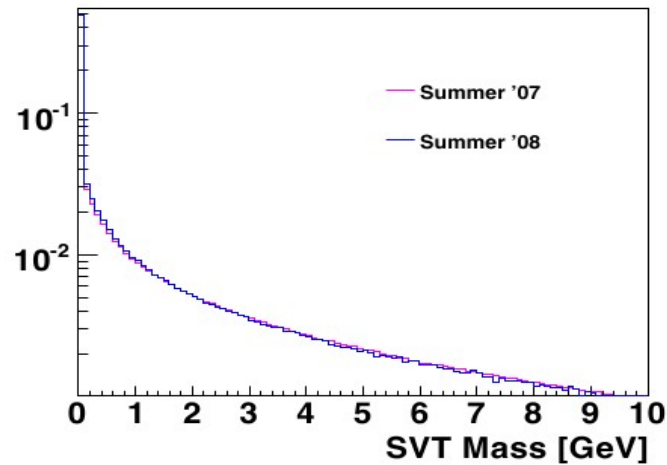


- **Datasets:**

- CSG_CAF_<skim>PASS4_p21.10.00_p20.12.00
- CSG_CAF_<skim>PASS4_p21.10.00_p20.12.01
- CSG_CAF_<skim>PASS4_p21.10.00_p20.12.02_summer08

Dataset summer'08

SVT b-tagging Variables



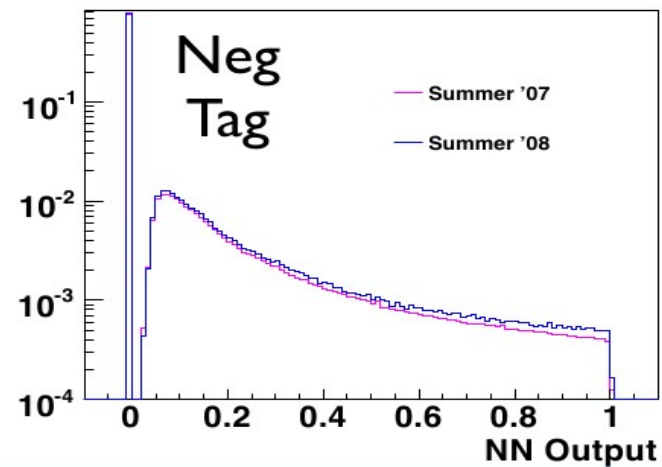
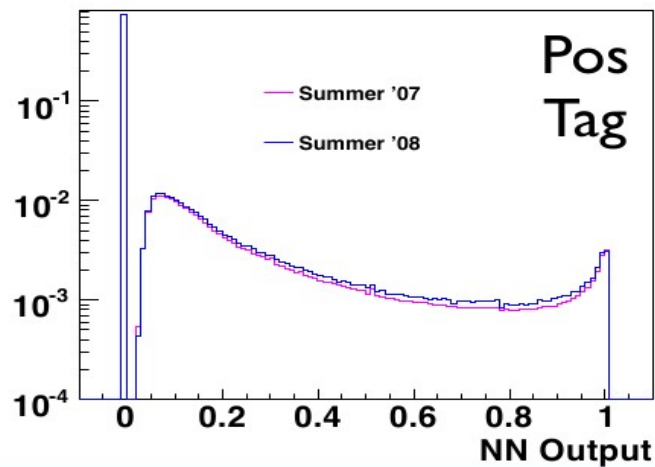
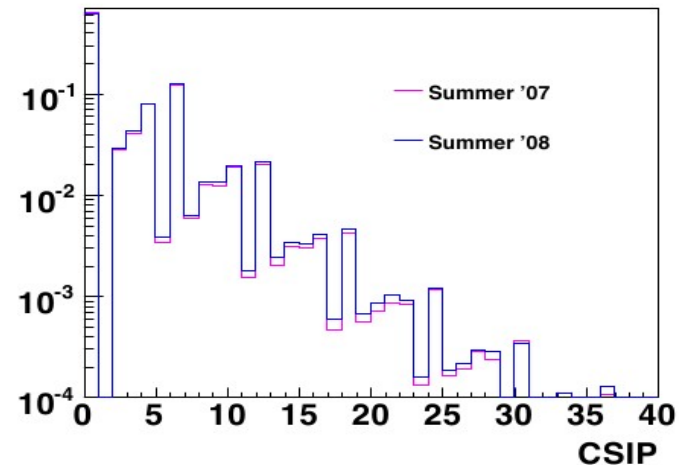
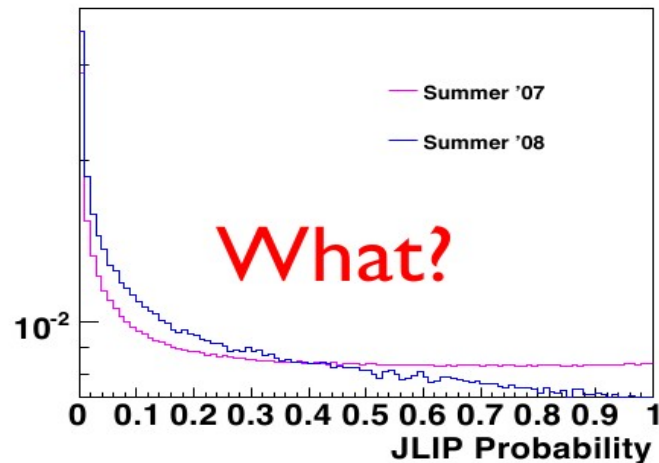
Thomas Gadfort

Convener's Meeting - June 6, 2008

10

Dataset summer'08

NN b-tagging Variables



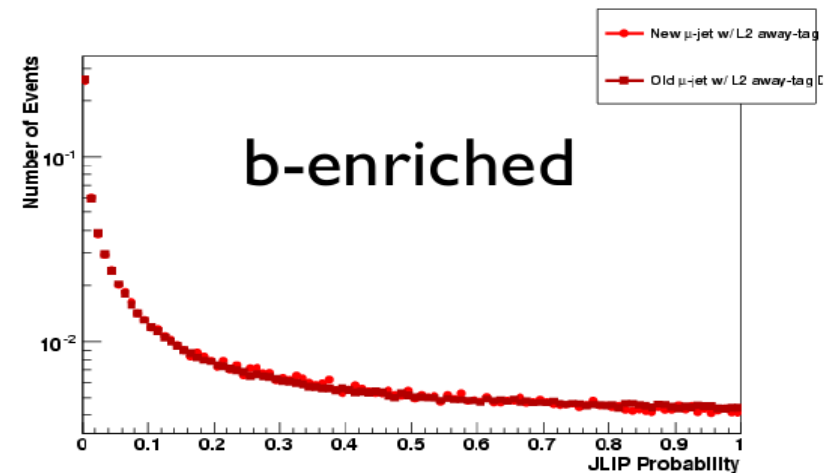
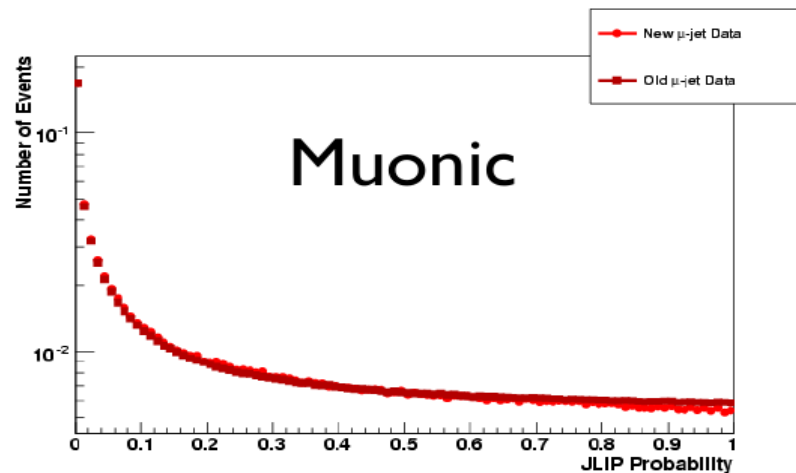
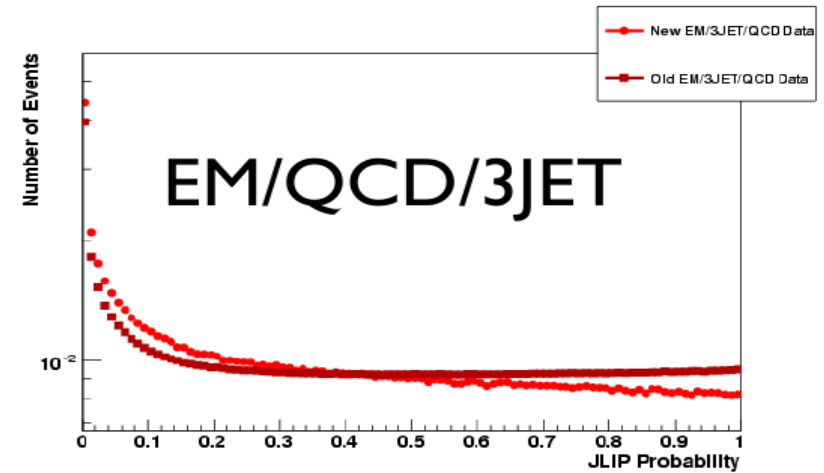
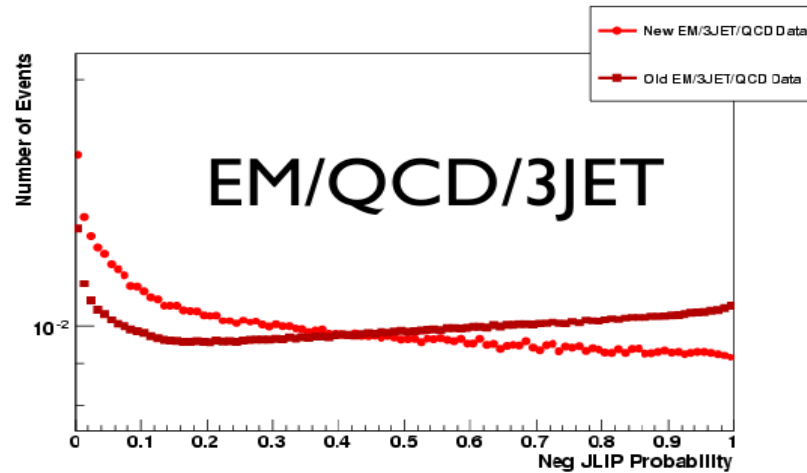
Thomas Gadfort

Convener's Meeting - June 6, 2008

11

Dataset summer'08

JLIP in 3 samples



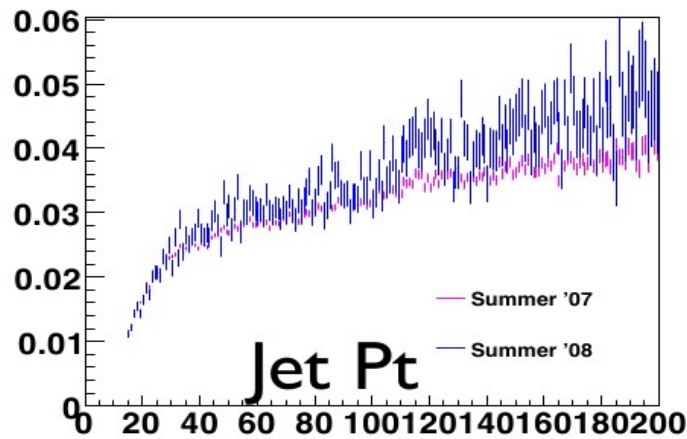
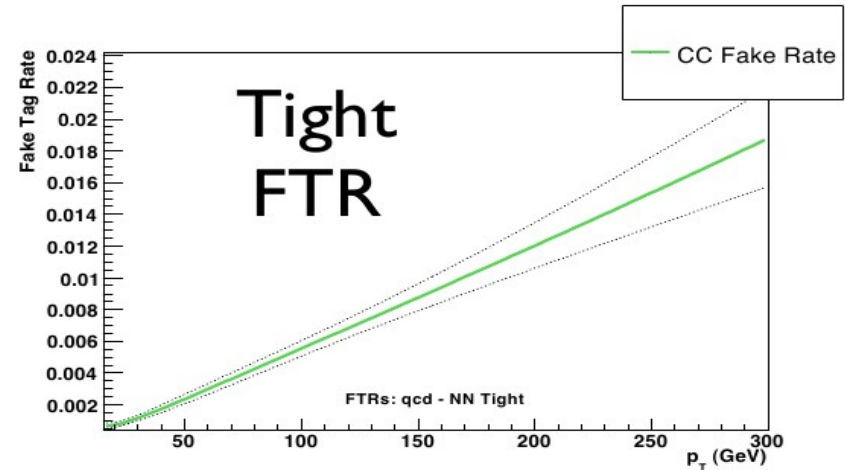
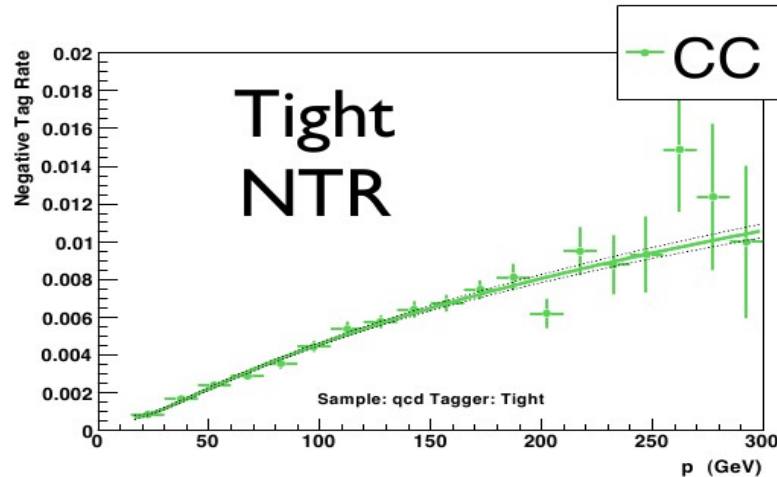
Thomas Gadfort

b-ID Meeting - June 20, 2008

18

Dataset summer'08

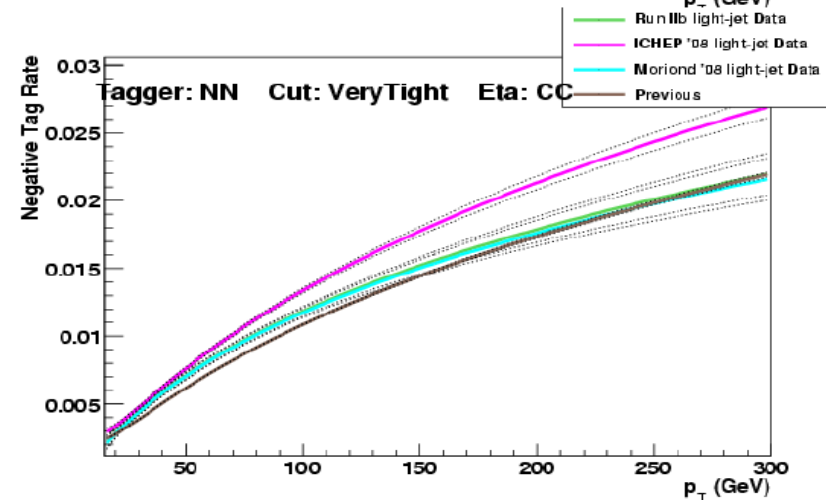
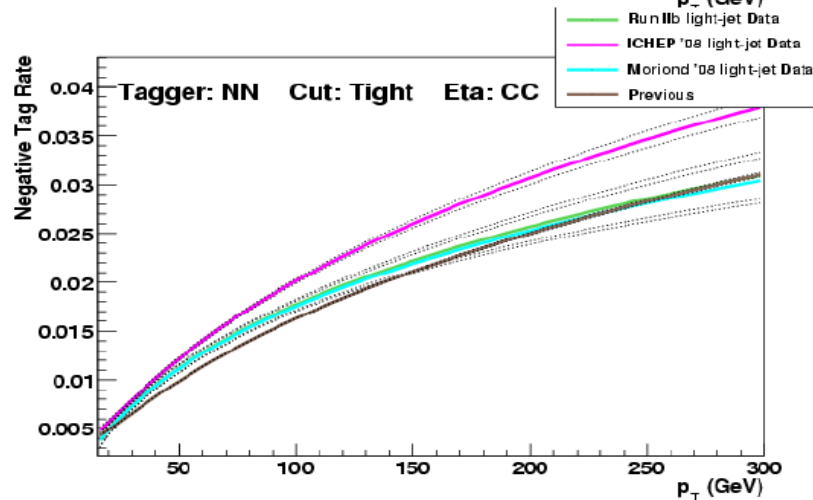
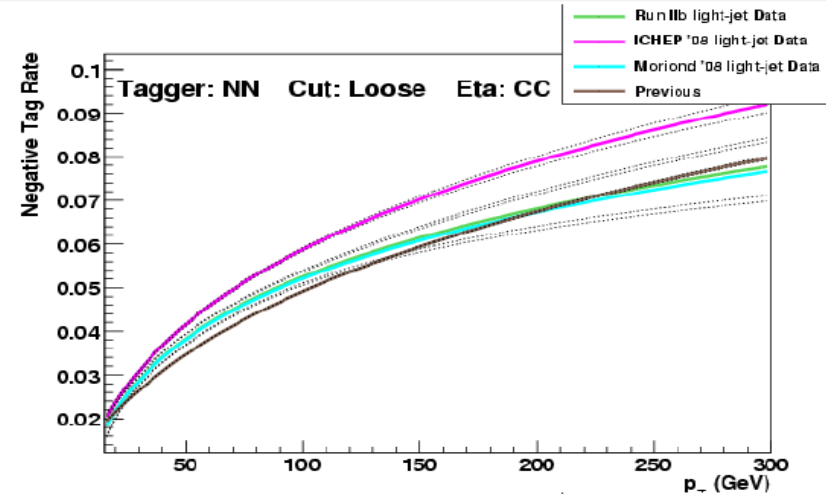
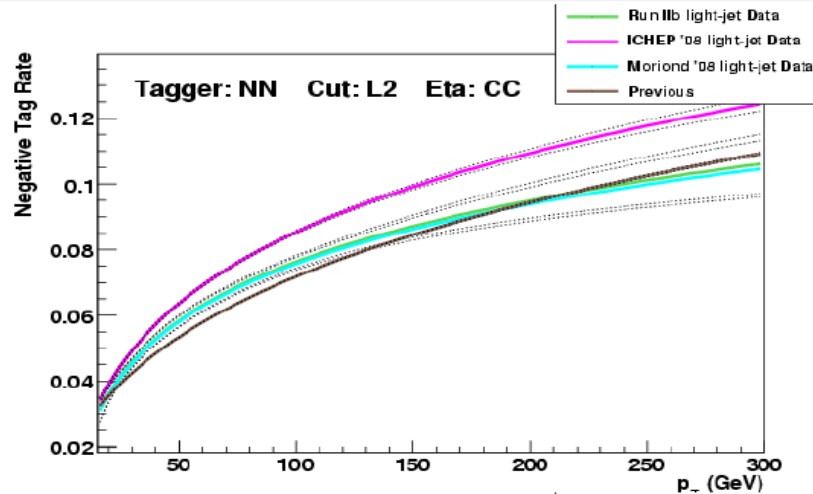
Fake Rates For TIGHT NN



- Fake rate has increased slightly, but this must be due to the odd JLIP behavior. The authors have been contacted and we hope to have a fix soon.
- Once we have a new JLIP, we'll release new TRFs/FTRs if the central value/errors have changed significantly.

Dataset summer'08

Negative Tag Rate Comp.



Thomas Gadfort

b-ID Meeting - June 20, 2008

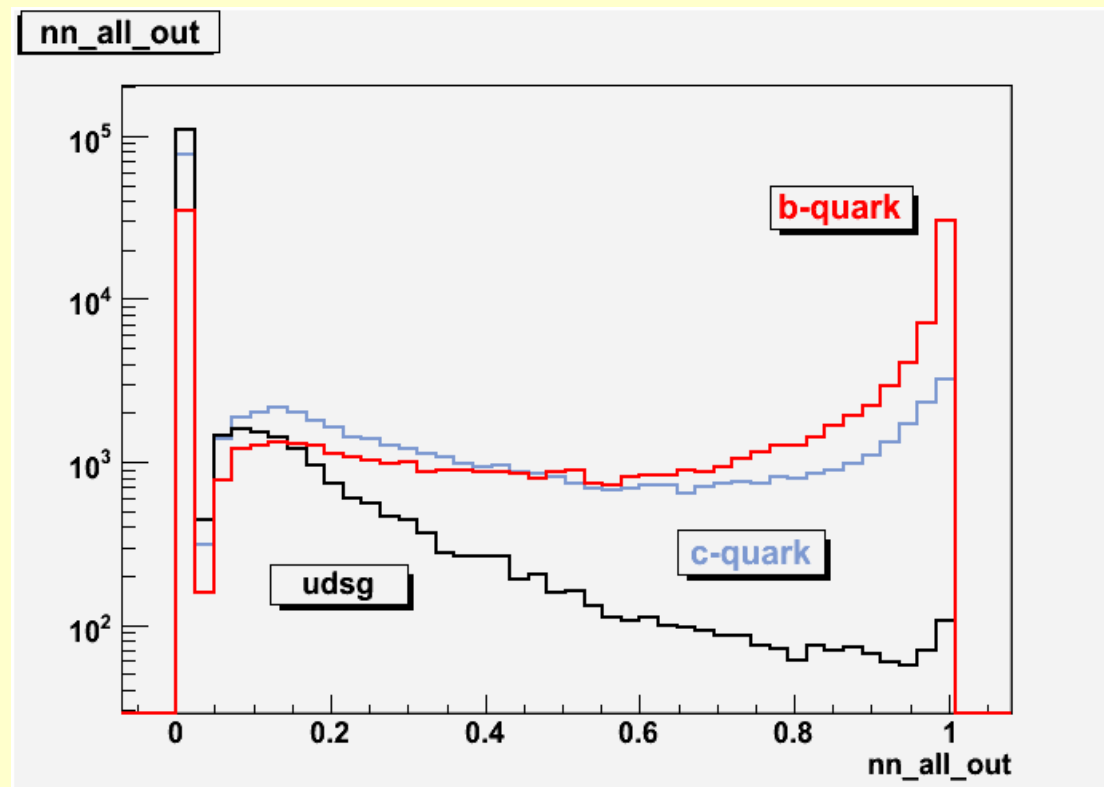
13

Discrimination b/c

Motivations

Current NN b-tagging

- Discrimination jets b / jets légers (u,d,s and g)
- Efficacité de sélection des jets c $\sim 10\text{-}15\%$



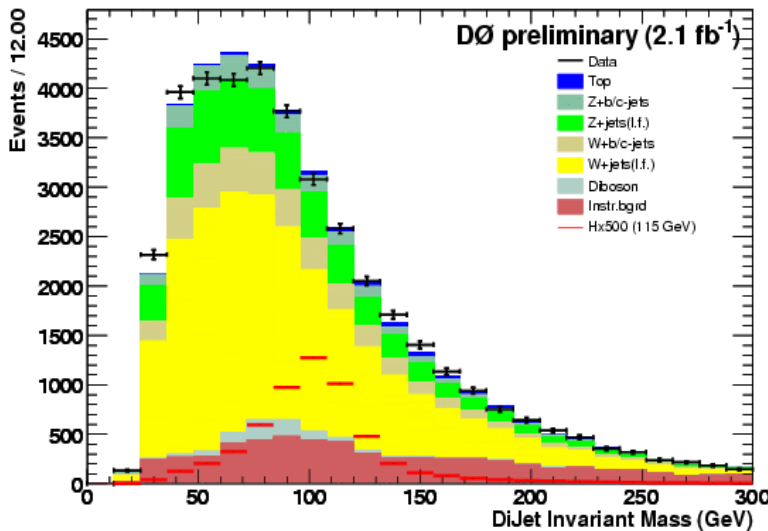
Motivations

Work-flow des analyses

1) Réduire immense bruit de fond léger (multijets, V+jets)

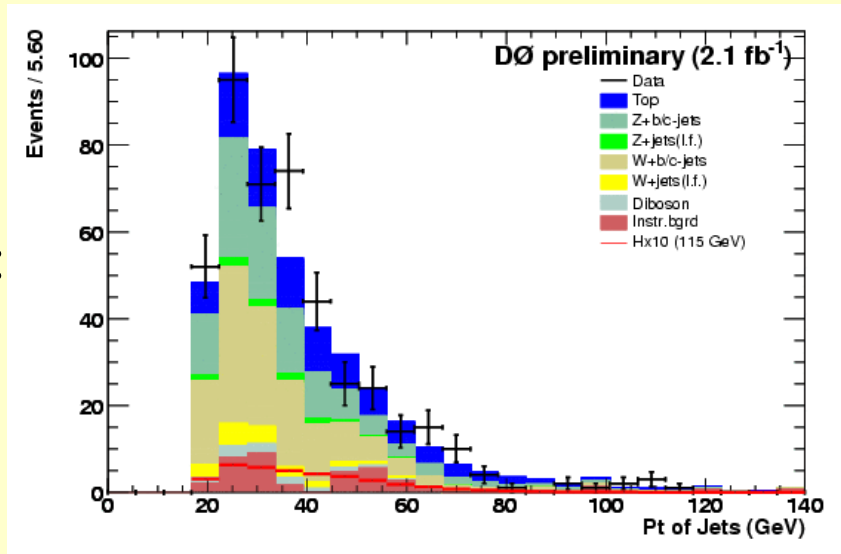
Ex: $ZH \rightarrow \nu\nu b\bar{b}$ (K. Peters, Physics Workshop)

--> pre-tag



--> post-tag:

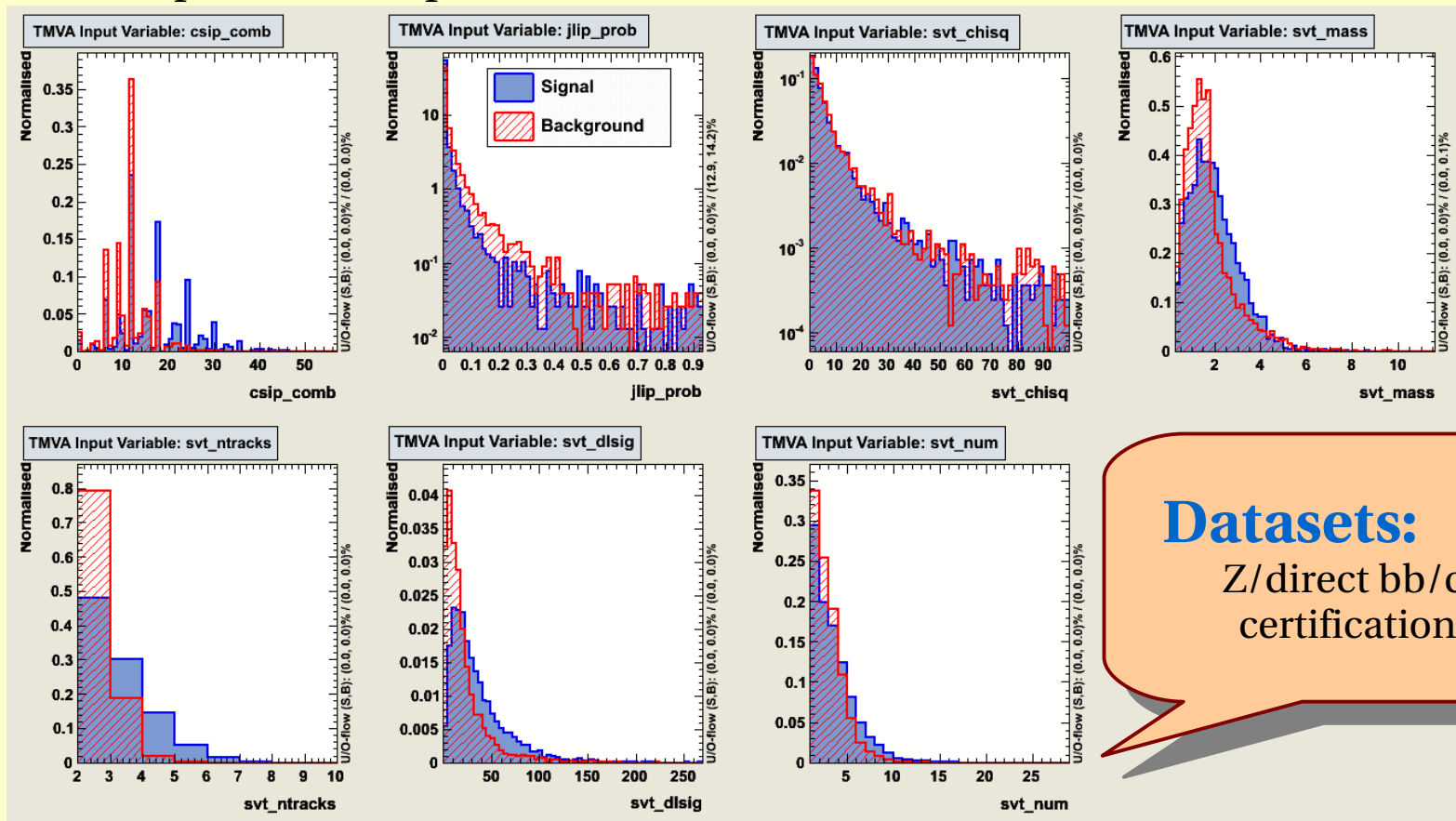
b/c ~ 2-3 (analyses simple/double-tag)



Discriminant b/c

Premiere etape:

- Meilleures performances observées: Multi-Layer Perceptron NN (MLP) et Boosted Decision Tree (BDT) du package **TMVA**
- Dans un premier temps entrainement avec les memes variables du NN b/l :

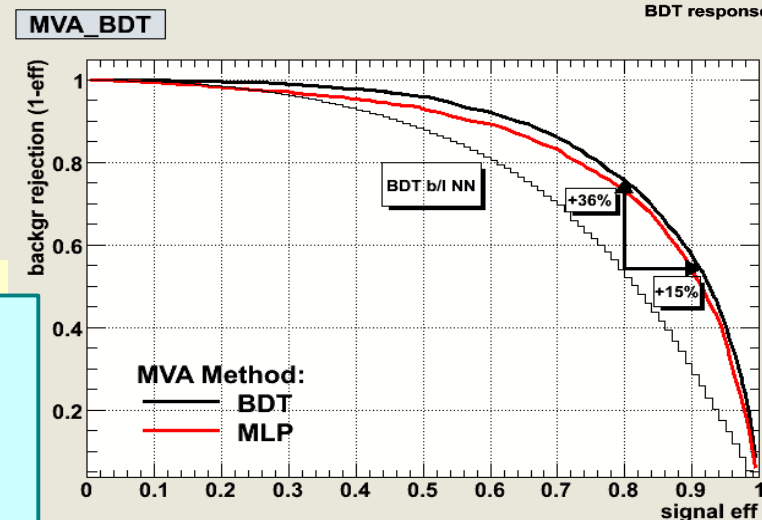
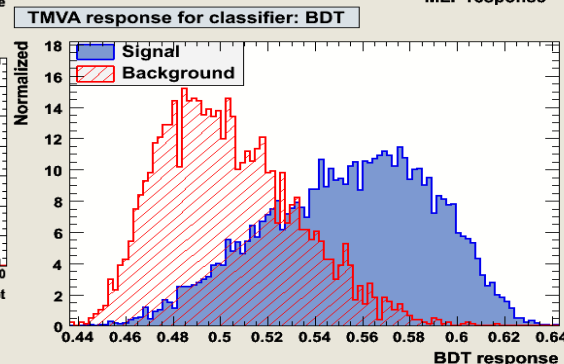
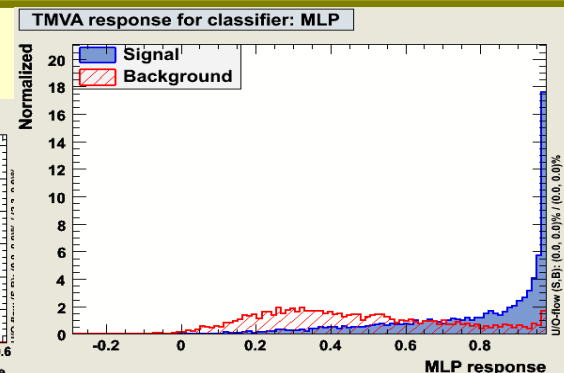
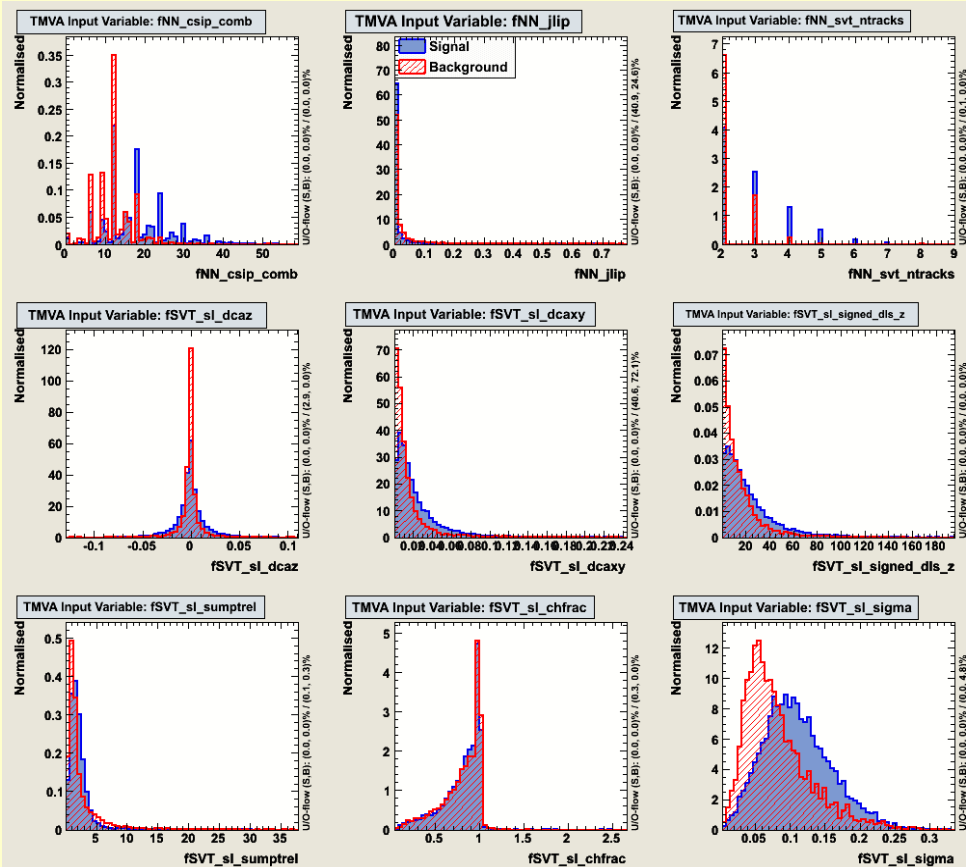


Datasets:

Z/direct bb/cc
certification p20

Discriminant b/c

Augmenter le nombre de variables



Datasets:

Z/direct bb/cc

@ 80% b eff., 75% c rej.
@ 90% b eff., 55% c rej.

For L6 b/l jets !

Data/work-flow

Comment évaluer les performances dans les données réelles ?

- **Rappel:** NN , $Sf_c = Sf_b \dots$
- **Approche *a la CDF***
 - Origine(s) des différences données réelles/simulation ?
 - Etalonnage NN , BDT ? But: prédiction directe des fractions de saveurs

Datasets:

- **Simulation:** directbb/cc, qcd(15-...), zbb/cc/qq et zbb/cc/qq $\rightarrow \mu$
- **Data:** post-shutdown 2007 MUinclusive et/ou QCD skims

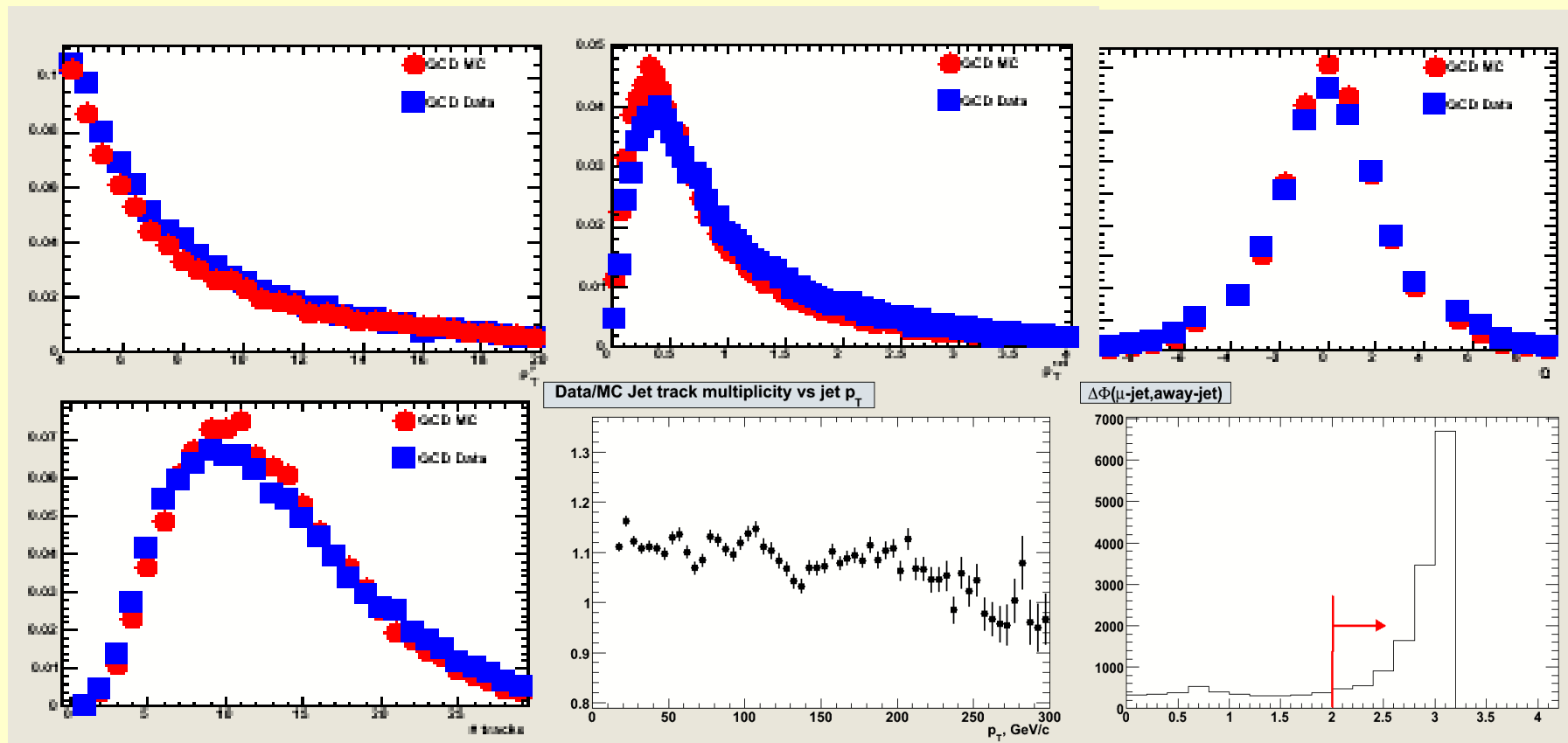
Data/work-flow

- **MuQCD:** 1 jet avec un medium-muon (track-match, nseg 3) a $dR(\text{jet}) < 0.5$ (muon-jet)
- **HFQCD:** de MuQCD, un muon-jet taggable, avec un jet taggable, $d\Phi > 2$, (away-jet)
- Comparer toutes (~35) les propriétés/variables d'identification

MuQCD

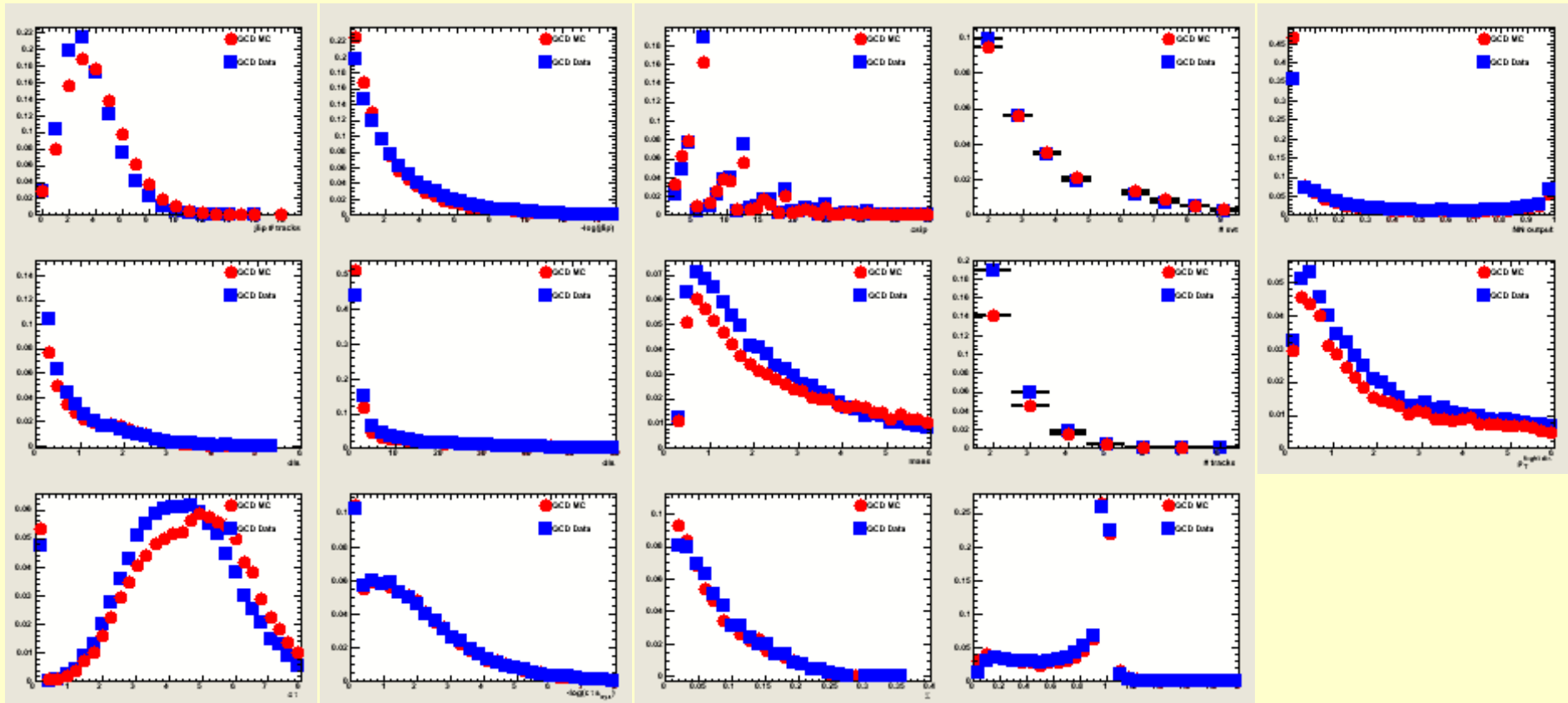
Propriétés

- Comparer les propriétés/variables d'identification entre données réelles et simulées sélectionnées à l'identique



HFQCD

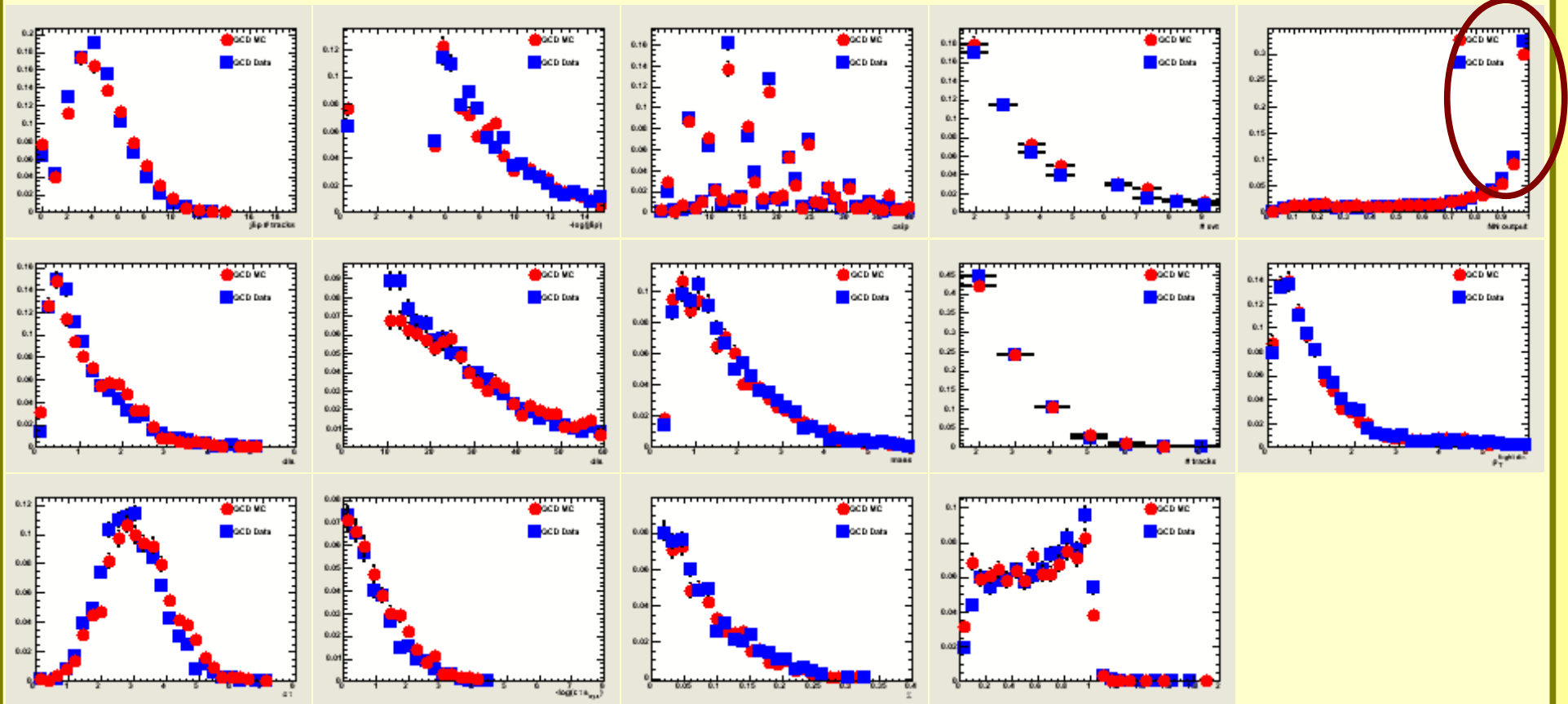
MuQCD + vertex super-loose ($\sim 90\%$ b, $\sim 50\%$ legers)



HFQCD signal

La simulation decrit-elle bien des jets de b ?

On impose $j_{\text{lip}} < 0.005$ ($\sim 0.5\%$ background)



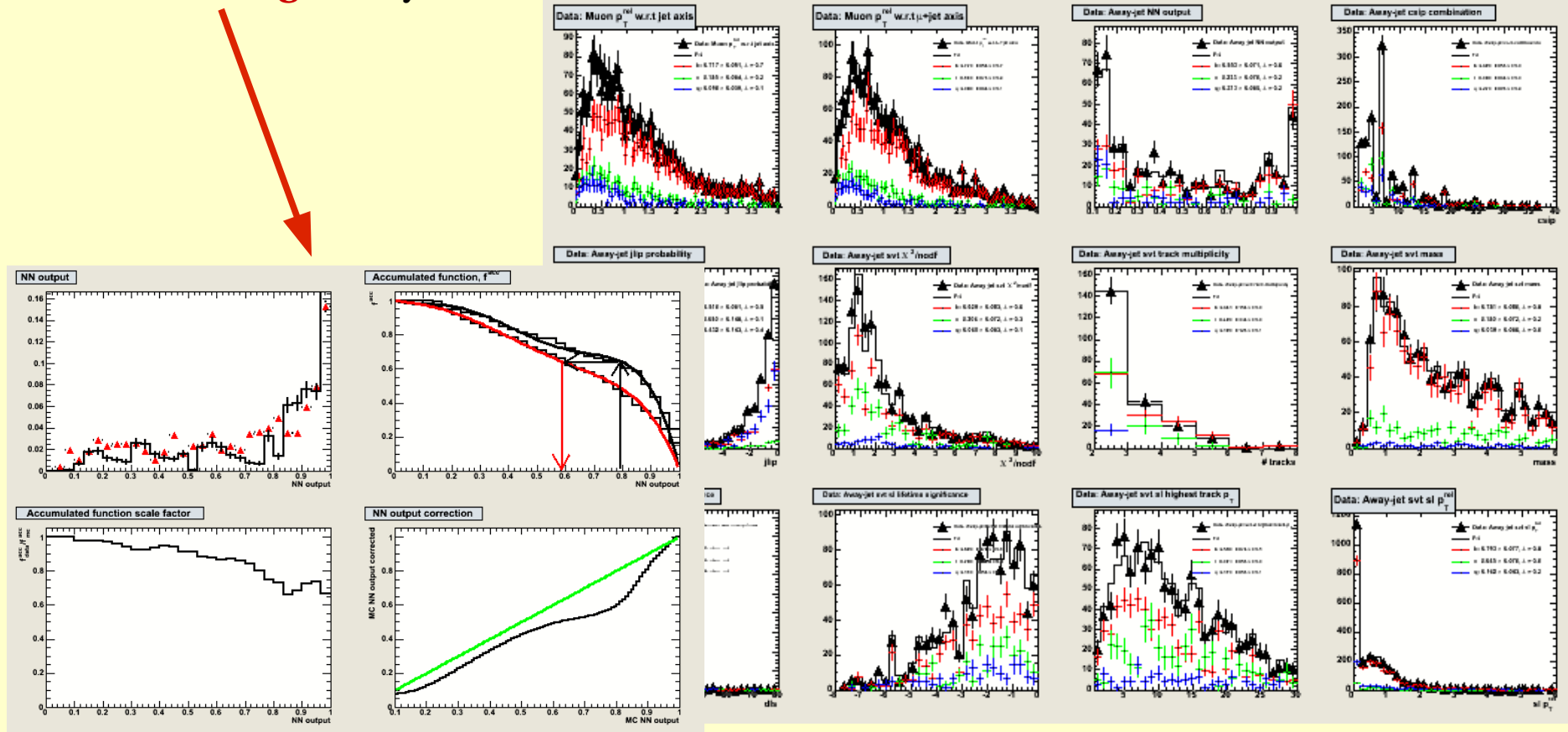
Bonne description des proprietés d'etiquetage !

(le muon-jet et away sont controlés)

B-tagging *a la* CDF

These **Svenja Richter** (chap.6, "Search for single-top Quark Production with the CDFII experiment")

- Si proprietes ok: comparer fractions data/simu
- Differences: fractions de saveurs mal décrites dans simulation --> **courbes d'etallonnage**--> systématiques



- Identification et sélection des **"bonnes"** variables de tagging

Conclusion/Plans

- **Augmentation de la lumi.:** taggabilité, performances, ...
- **Layer0:** quel(s) gain(s) ?
- **SLT NN:** nouveau (p17 & p20), gain absolu en performance
- **Tagger B/C/L:** nécessite une tres bonne comprehension des fractions de saveurs
 - Construction d'un discriminant: facile avec les outils actuels (e.g: TMVA), efficacité de sélection de jet-b de 80% pour 80% de jet-c rejetes (MC)
 - Description/comprehension data/simulation est moins triviale
- Code est *presque* pres pour test dans les analyses (post-ICHEP)
- Etudes/optimisations nécessaires pour BDT B/C
- Single-top intéressé par disc. B/C; WH, ZH ??
- “Bug” JLIP a corriger

Back-up

TMVA BDT input variables

