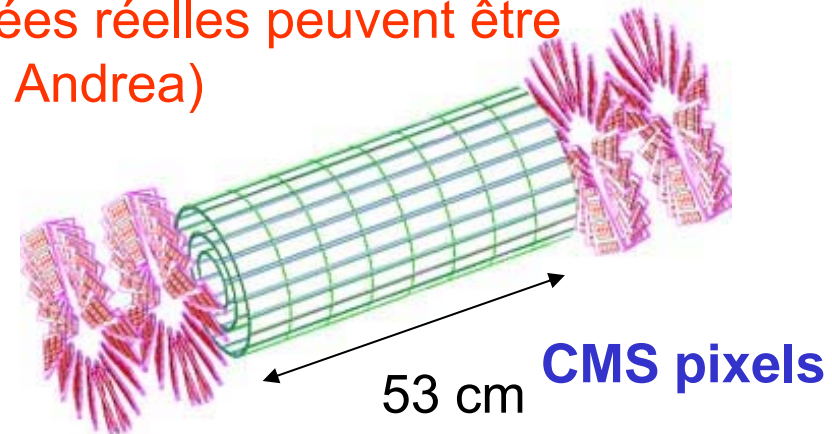
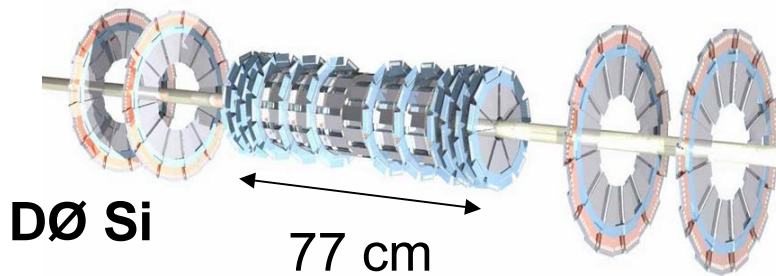


# Identification des jets b : CMS versus DØ

Daniel Bloch (IPHC Strasbourg)

- Objectif** : vérifier si les méthodes utilisées au Tevatron pour mesurer les efficacités de b-tagging à partir des données réelles peuvent être appliquées au LHC (stage M2 de Jérémie Andrea)



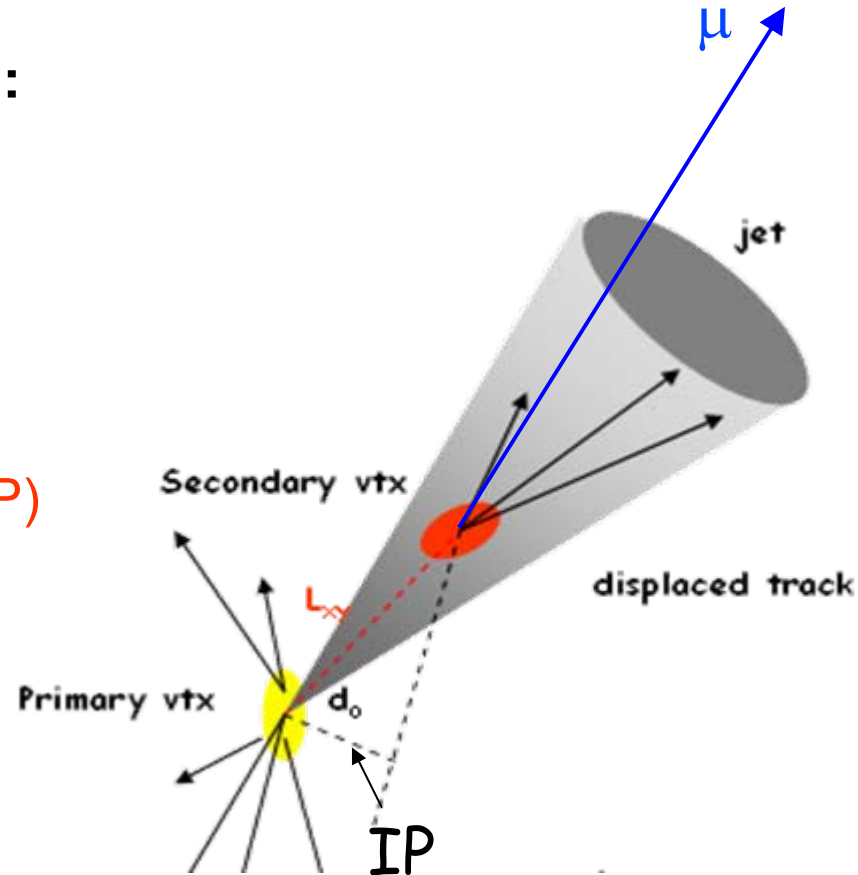
- DØ** : 4 couches centrales dét. Si (simple&double faces,  $R = 2.7-10.5$  cm, pas de  $50 \mu\text{m}$ ,  $\sigma_{R\phi} < 15 \mu\text{m}$ ) +  $2 \times 8$  disques avant ( $3 \text{ m}^2$  Si en tout) + 8 couches fibres scintillantes (stereo,  $R = 15-51$  cm) :  $|\eta| < 2.0$   
→ exploité depuis 2001 (Remarque : layer 0 à  $R=1.7$  cm installé en avril)
- CMS** : 3 couches pixels ( $R = 4-10$  cm,  $100 \times 150 \mu\text{m}^2$ ,  $\sigma_{R\phi} = 10 \mu\text{m}$ ,  $\sigma_z = 20 \mu\text{m}$ ) + 8 couches centrales dét. Si ( $R = 20-110$  cm, pas de  $80-180 \mu\text{m}$ ) +  $2 \times (2 \text{ pixels} + 9 \text{ Si})$  disques avant ( $1 \text{ m}^2$  pixels +  $200 \text{ m}^2$  Si en tout) :  $|\eta| < 2.4$   
→ pixels installées en 2008

# Identification des jets b

Important pour étudier le **top**, découvrir un **Higgs** léger ( $H \rightarrow b\bar{b}$ )  
et / où la **SUSY** (sbottom, stop, ...)

## Caractéristiques des hadrons beaux :

- Grande durée de vie ( $\sim 1.6$  ps)
- Fragmentation dure ( $E_B / E_b \sim 0.7$ )  
 $\Rightarrow$  grande distance de vol  
(vertex secondaire)
- $\Rightarrow$  traces à grand paramètre d'impact (IP)
- Masse élevée ( $\sim 5$  GeV)
- Désintégrations semi-leptoniques  
 $BR(b \rightarrow e \text{ ou } \mu X) \sim 21\%$ ,  
 $BR(b \rightarrow c \rightarrow e \text{ ou } \mu X) \sim 19\%$ , ) :  
muon ou électron dans les jets



# Des algorithmes de tagging similaires

- **Comptage de traces à grand paramètre d'impact (IP)**
- **Méthode probabiliste** par jet, basée sur les IP des traces (à 2D dans DØ (JLIP), 3D dans CMS)
- **Vertex secondaire** : distance de vol  $\Delta L / \sigma$  (3D, DØ et CMS)
- **Identification combinée** :
  - **dans DØ** : combine les 3 algorithmes précédents + des info sur les vertex secondaires ( $\chi^2$ , masse, nb. de traces, nb. de vertex) dans un réseau de neurones (NN)
  - **dans CMS** : fonction de vraisemblance basée sur les vertex secondaires ( $\Delta L / \sigma$ , masse, nb. de traces,  $E_{\text{traces}} / E_{\text{jet}}$ ), rapidité des traces / au jet, paramètre d'impact
- **Identification des leptons** : muon (DØ et CMS) , électron (CMS)

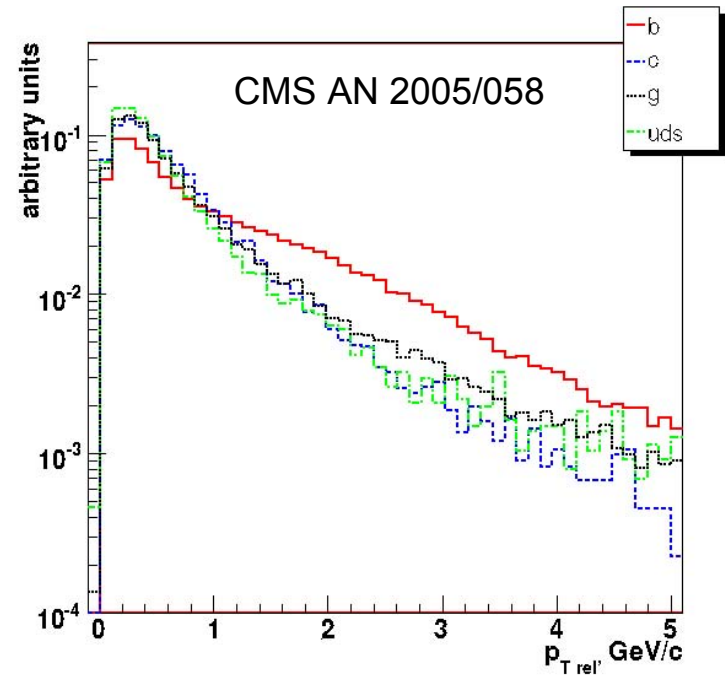
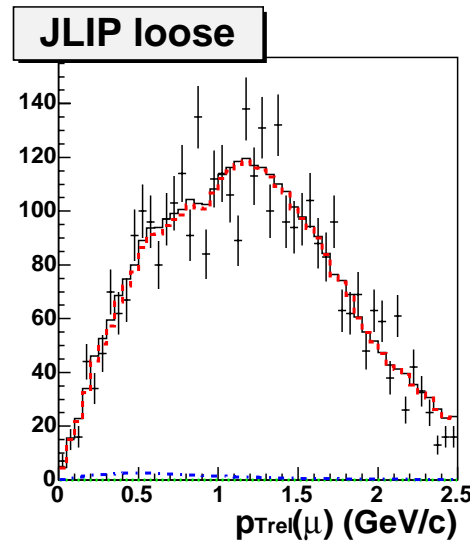
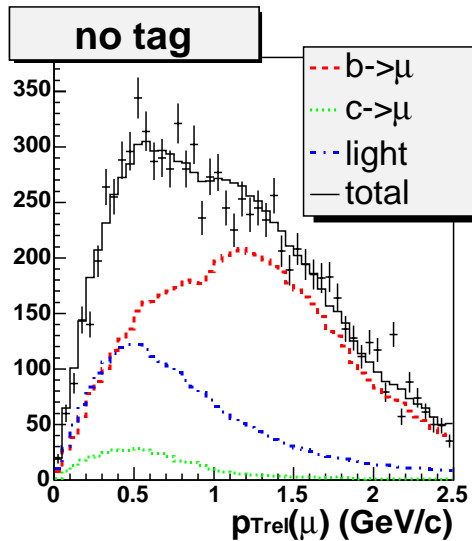
# Exemple 1 : soft muon

exemple de variable discriminante :

moment transverse relatif,  $p_{Trel}$ , du muon / à l'axe du jet :

• Dans  $D\bar{D}$  : comparaison Data/MC

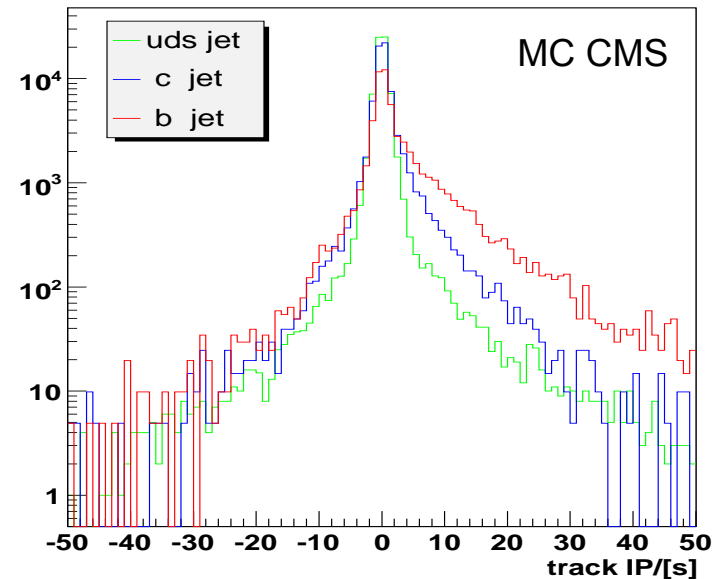
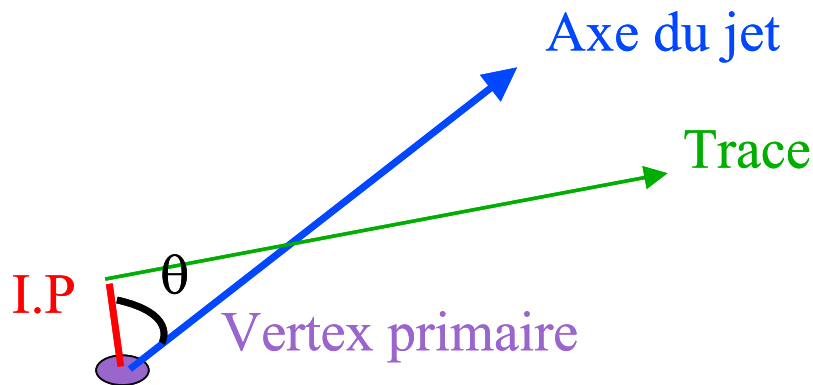
Dans CMS : MC



- $p_{Trel}$  plus élevé pour les  $b$  que pour les autres saveurs
- moindre différence entre  $c/uds$

# Exemple 2 : paramètre d'impact

- On calcule pour chaque trace son paramètre d'impact (IP) et son erreur ( $\sigma$ ) :  $\text{significance} = \text{IP} / \sigma$
- La signficance est signée positive si la trace est en amont ( $\theta < \pi/2$ ) du vertex primaire suivant l'axe du jet, ou bien négative si elle est en aval ( $\theta > \pi/2$ )

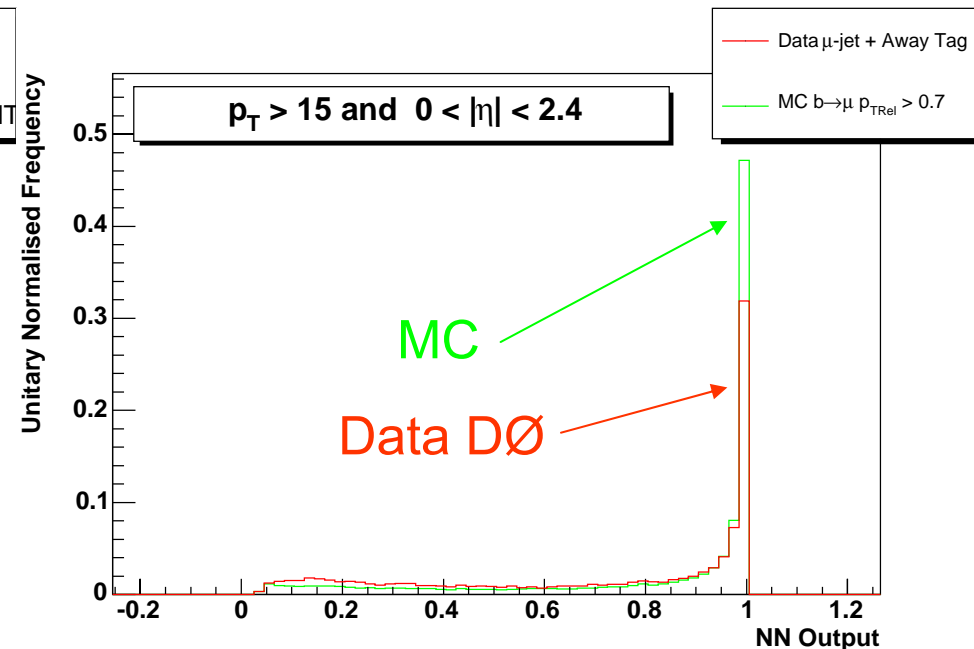
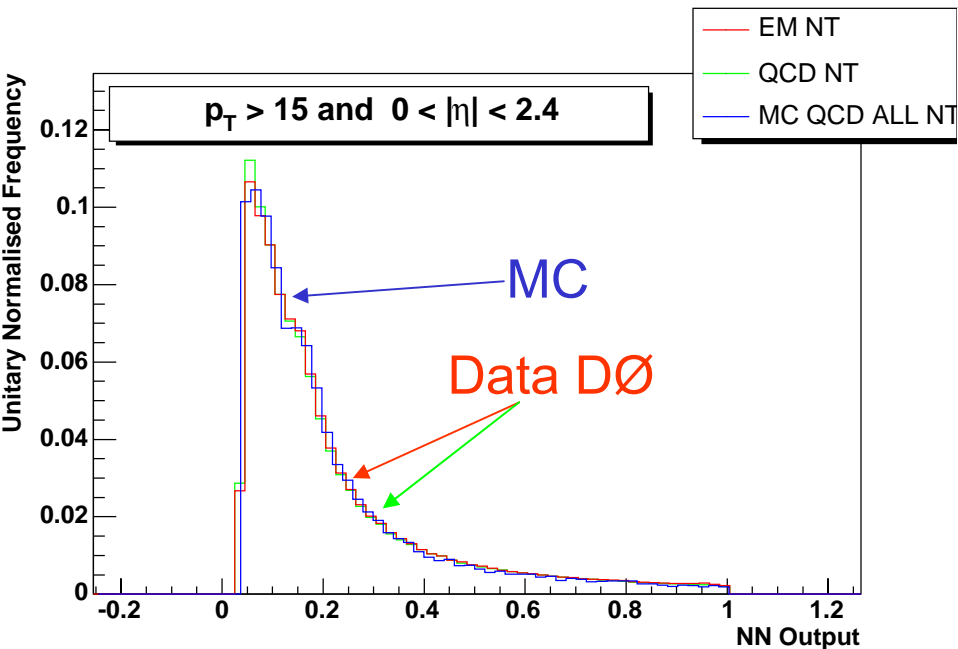


- Les traces à paramètre d'impact négatif sont utilisées pour définir une fonction de résolution. On en déduit une loi de probabilité pour que toutes les traces d'un même jet soient issues du vertex primaire

# Exemple 3 : tagging combiné

Ici dans  $D\emptyset$  avec le NN : comparaison Data / MC

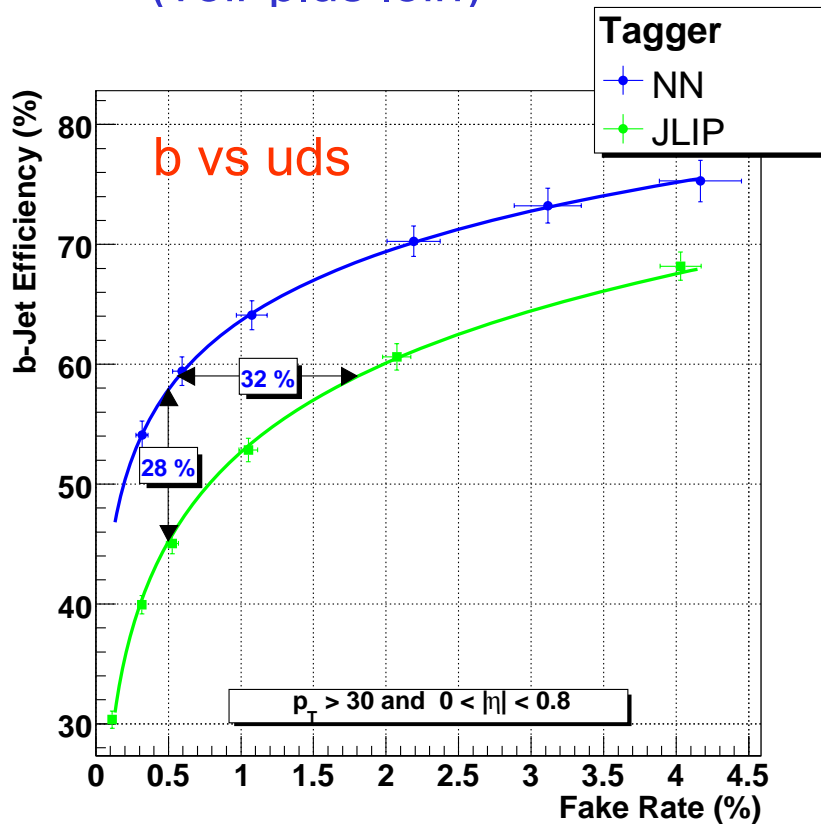
- bon accord pour les jets légers
- désaccord pour les jets-b



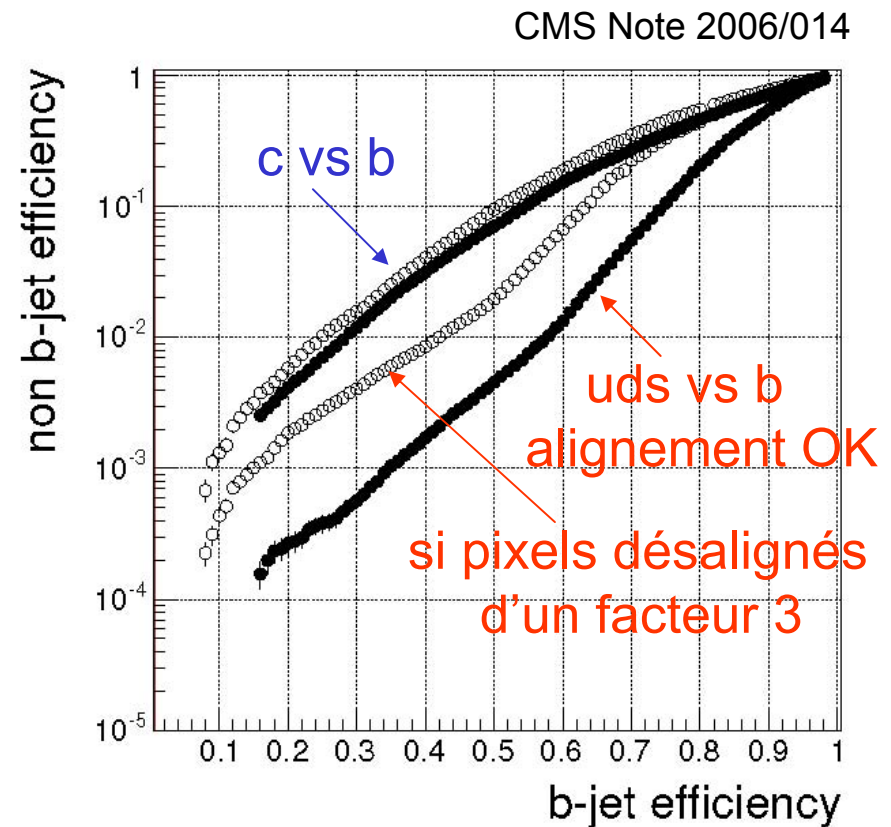
# Tagging combiné (suite)

Performances calculées

- DØ : à partir des Data (voir plus loin)



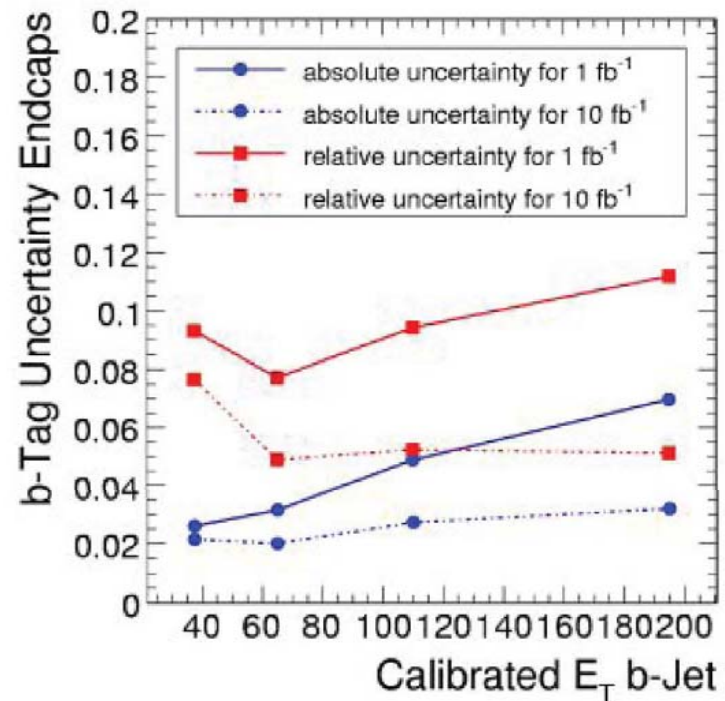
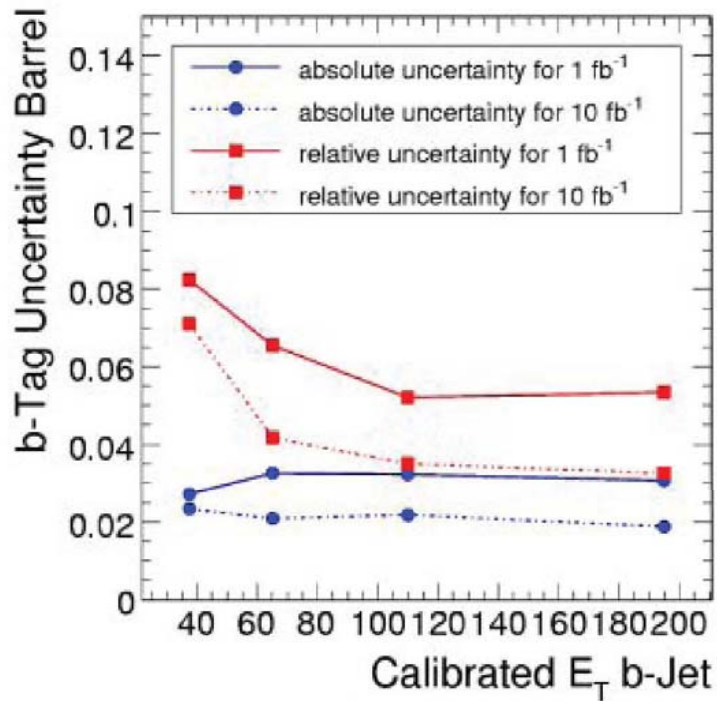
- CMS MC :



# Evaluation des efficacités

- Dans CMS : basée sur du pur MC (bien sûr),
- une étude a été menée pour utiliser des événements  $t\bar{t}$  doublement taggés afin de mesurer l'efficacité de b-tag sur les données : prometteur, mais nécessite une bonne statistique de  $t\bar{t}$

CMS Note 2006/013



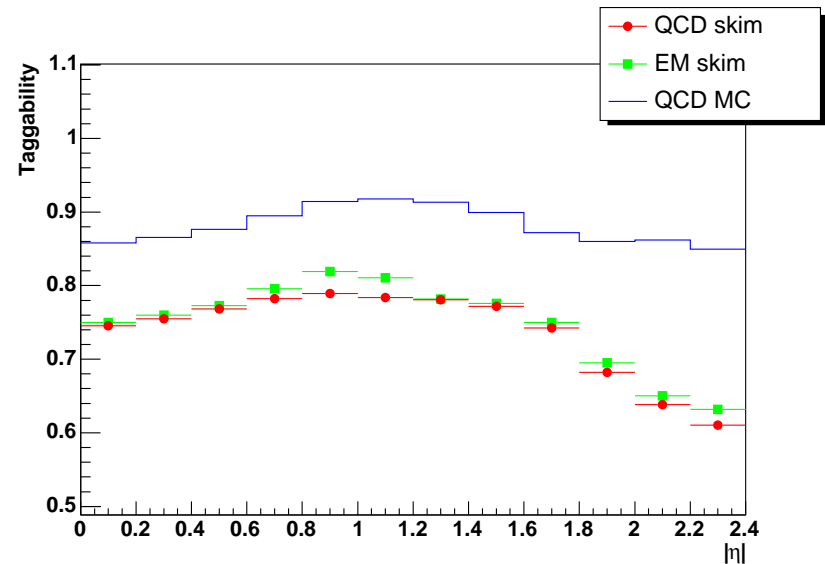
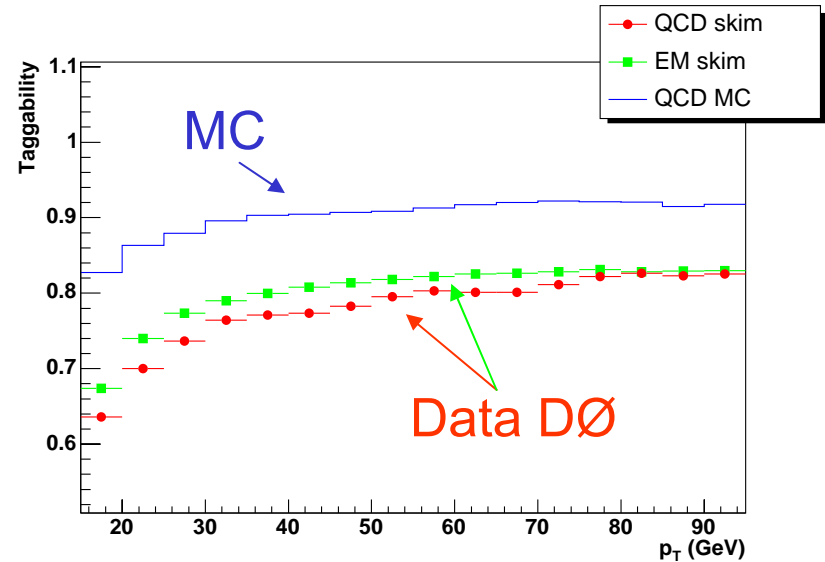


# Evaluation des efficacités (suite)

- Dans  $D\bar{O}$  : basée sur les données et des corrections MC (car le MC ne décrit pas les données à mieux que 20%).
- On distingue 3 étapes :
  1. **“Taggabilité”** : efficacité pour l’on associe au moins 2 traces à un jet ( $pt > 0.5-1$  GeV, IP proche du vertex primaire), paramétrisée en fonction de  $pt(\text{jet})$ ,  $\eta(\text{jet})$
  2. pour les jets “taggables” : calcul d’une **efficacité de tagging**, paramétrisée en fonction de  $pt(\text{jet})$ ,  $\eta(\text{jet})$  : **“Tag Rate Function”** définie pour chaque saveur  $u, d, s, c, b$  :  $TRF(pt, \eta, \text{saveur})$ , ou bien calcul d’un **“Scale Factor”**  $SF(pt, \eta, \text{saveur}) = \text{eff. Data} / \text{eff. MC}$ .  
On calcule TRF et SF pour des **points de fonctionnement** pré-définis (6 points correspondant à des efficacités moyennes de 0.1% à 4% pour les jets de quarks légers)
  3. pour les données réelles (Data) : on applique l’algorithme de tagging à chaque jet taggable  
pour le MC : on **pondère** chaque jet par sa taggabilité et son TRF (ou son SF, suivant les analyses)

# Taggabilité dans DØ

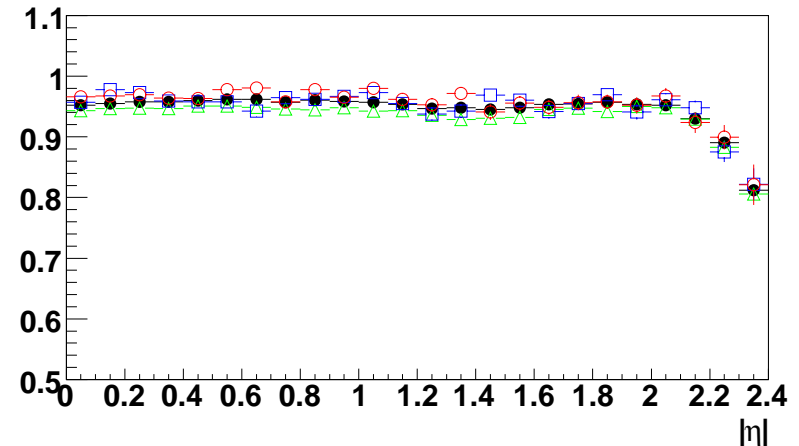
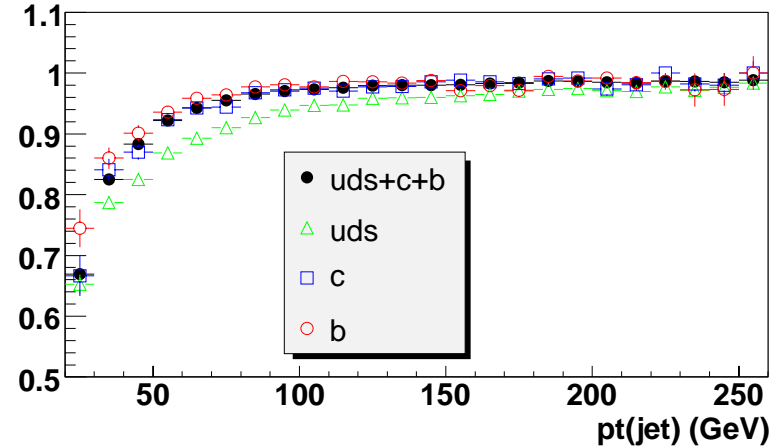
- Évaluée sur des triggers multi-jets, ou sur des triggers leptons (en relâchant les critères d'identification des leptons)
- rapport des taggabilités  
**Data / MC**  $\sim 0.85$



# Taggabilité dans CMS ?

- On peut répéter la même procédure pour du MC multi-jets dans CMS : par ex. en demandant  $\geq 2$  traces de  $pt > 1$  GeV et  $|IP| < 2$  mm associées à chaque jet :
- trop bon MC ?

Taggability in qcd 30-230 MC



# Mauvaise identification des jets légers (“mistags”) dans DØ

- On utilise un lot d'événements réels multi-jets contenant peu de quarks c et b :  
le tagger est appliqué aux traces à **significance négative** (traces à  $IP/\sigma < 0$ , ou vertex secondaire à  $\Delta L / \sigma < 0$ )

- Le MC qcd est utilisé pour appliquer une correction **SF<sub>1</sub>** :

$$\epsilon_{\text{sig}>0}^{\text{Data}}(\text{uds}) = \epsilon_{\text{sig}<0}^{\text{Data}}(\text{total}) \times \text{SF}_1$$

où

$$\text{SF}_1 = \epsilon_{\text{sig}>0}^{\text{MCqcd}}(\text{uds}) / \epsilon_{\text{sig}<0}^{\text{MCqcd}}(\text{total} = \text{uds} + \text{c} + \text{b})$$

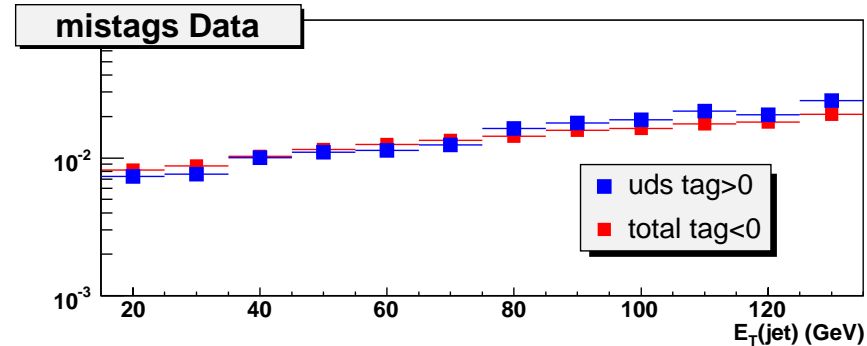
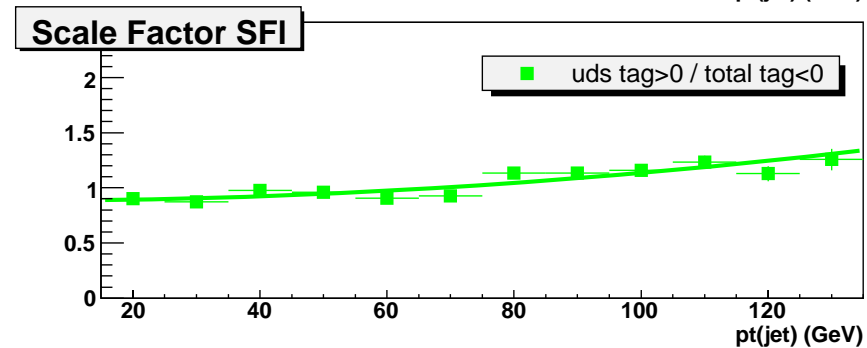
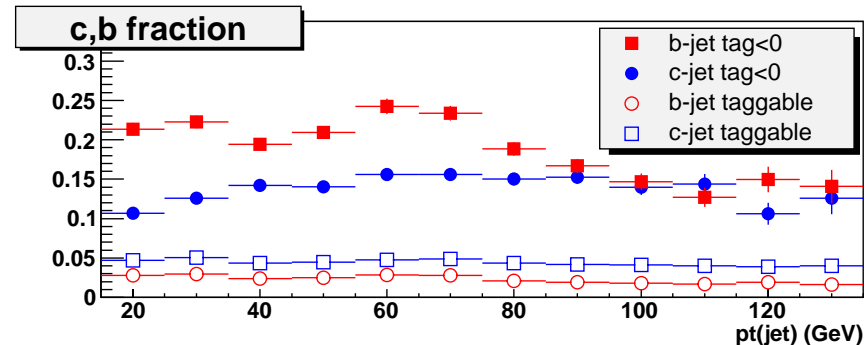
- Les systématiques  $\sim \pm 6\%$  à  $\pm 12\%$  dépendent de la composition en saveurs des Data multi-jets (évaluée d'après le MC qcd) et du taux de tags négatifs observés sur différents lots de données réelles

# “mistags” dans DØ (suite)

- MC qcd :  
contamination  
de **c** et de **b**  
avant et après  
tagging négatif

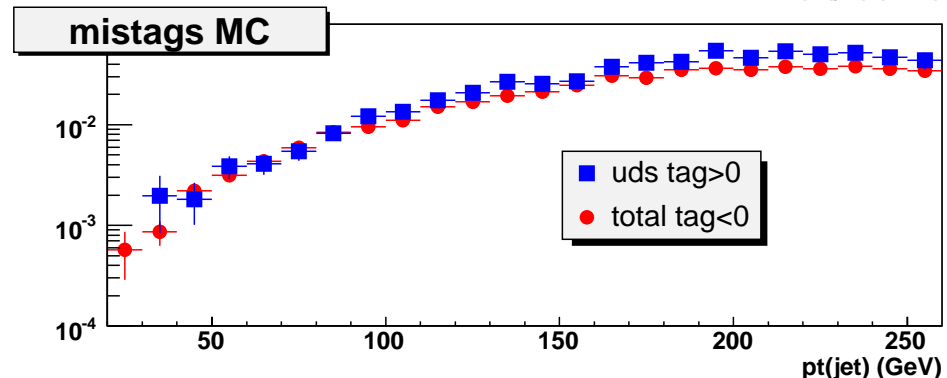
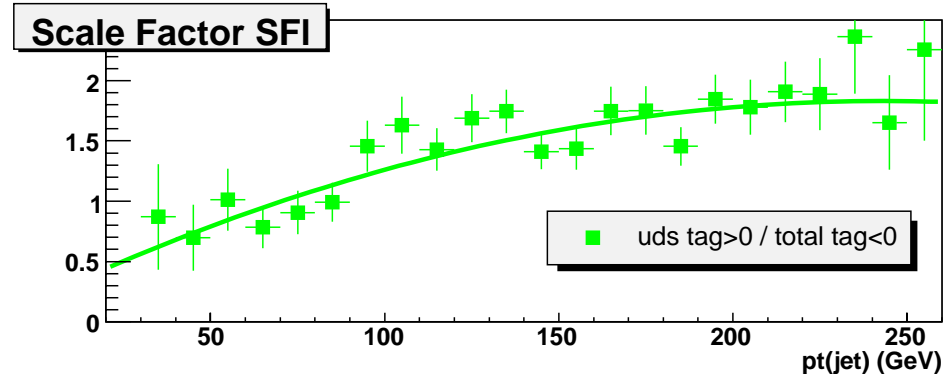
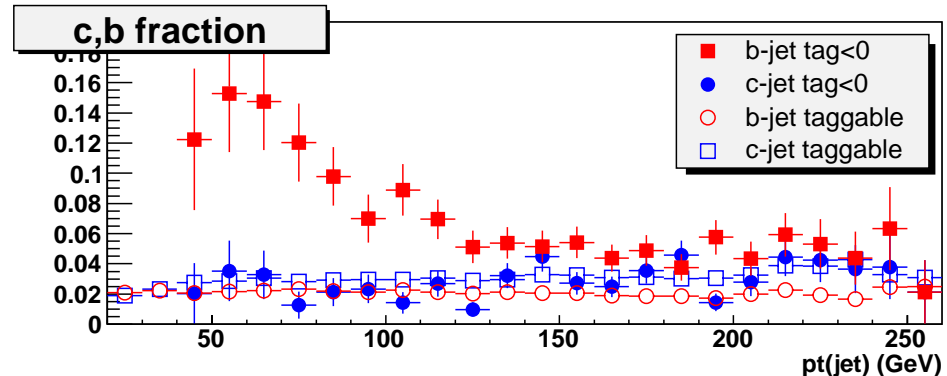
- Facteur  
correctif MC

- Data tag  
négatifs et  
efficacité aux  
jets légers



# mistags dans CMS ?

- Ici : méthode de comptage des traces à grand IP ( $\geq 3$  traces avec  $IP/\sigma > 2$ )
- Si tag négatif :  
~3-4% de jets-c,  
~4-10% de jets b
- Il semble donc possible d'utiliser les Data multi-jets dans CMS pour mesurer l'efficacité aux jets légers

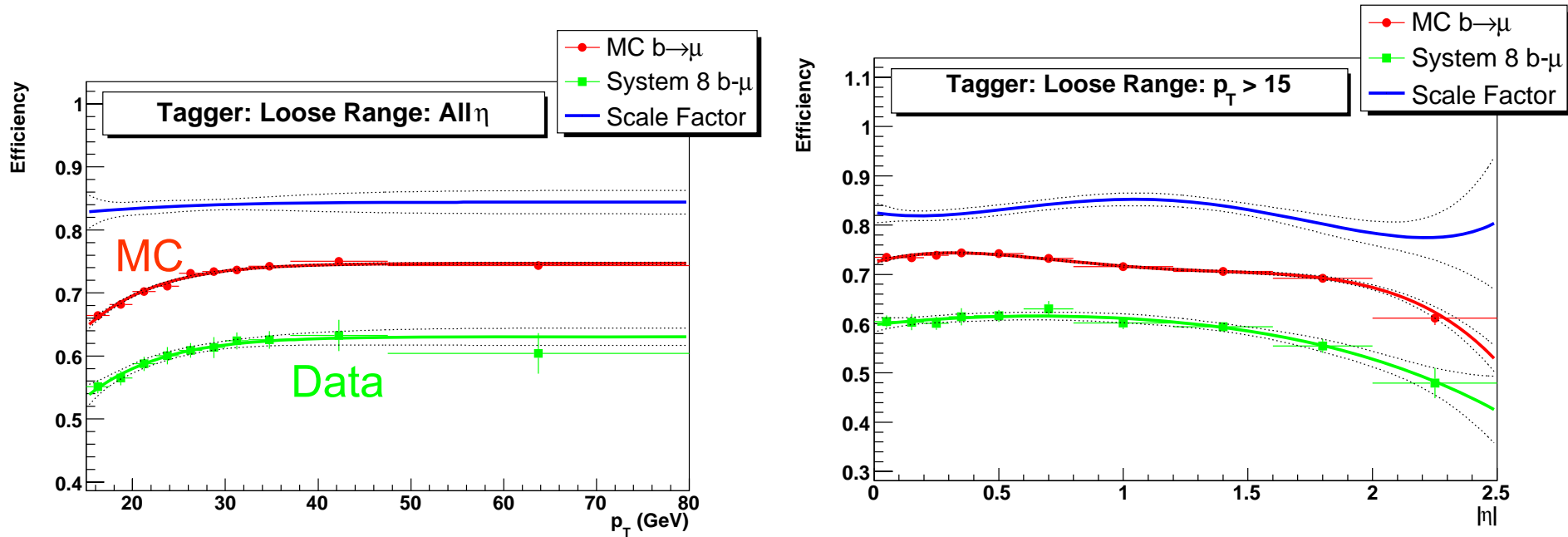


# Evaluation de l'efficacité de b-tagging dans $D\bar{0}$

- On utilise les **triggers “soft-muons”** pour avoir des lots d'événements contenant des jets avec un muon : candidats  $b \rightarrow \mu X$ , mais avec une contamination de jets  $c \rightarrow \mu X$  et de jets légers  $\rightarrow \pi/K \rightarrow \mu X$
- **A)** on peut utiliser **2 conditions de tagging** quasi-indépendantes :
  - une coupure sur le  $p_{Trel}$  du muon / à l'axe du jet
  - le tagging basé sur la durée de vie
- **B)** on peut aussi demander la présence d'un **second jet taggé** (par la durée de vie) dans le même événement
- On peut alors résoudre un système de 8 équations (non-linéaires) à 8 inconnues, dont l'efficacité de b-tagging qui nous intéresse.
- Le MC n'est utilisé que pour évaluer les facteurs de corrélations ( $\sim 1$ ) dues aux conditions A) et B), et leurs erreurs ( $\sim \pm 2\%$ ).

# Efficacités b-tag dans DØ

- avec le NN :



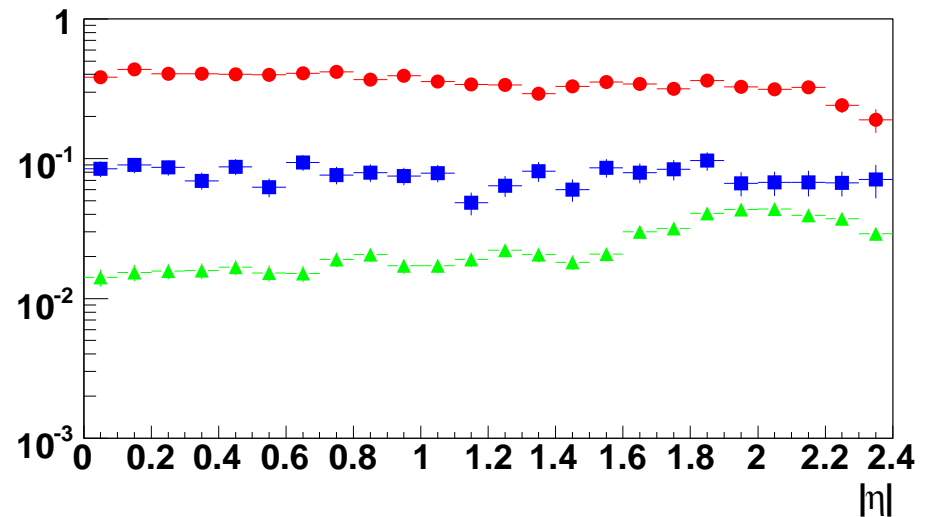
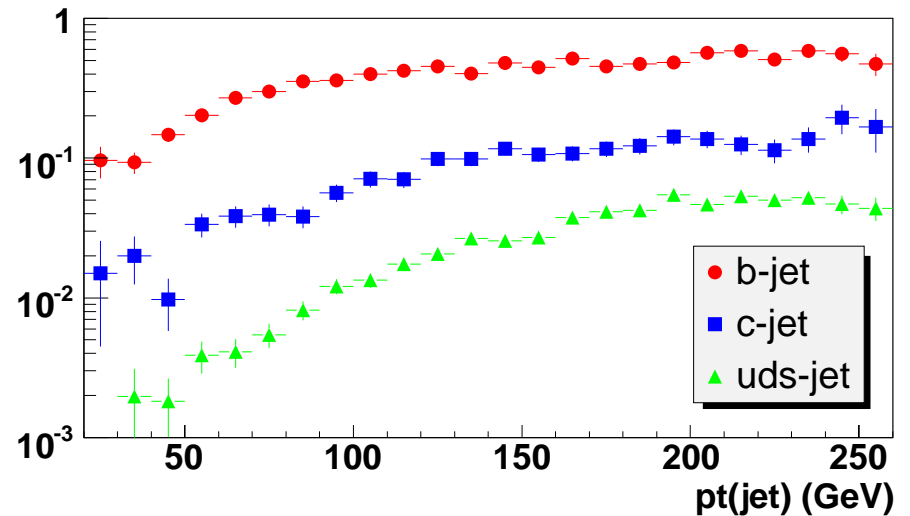
Rapport des efficacité de b-tag **Data** / **MC**  $\sim 0.83$



# dans CMS :

- Étude à venir sur l'utilisation de soft-leptons
- En attendant, efficacités vs pt et  $\eta$  dans le MC qcd : ici avec la méthode de comptage des traces à grand  $IP/\sigma$  : assez bon accord avec CMS Note 2006/019

Tagging efficiency in qcd 30-230 MC



# Conclusion

- Poursuivre ces études pour évaluer les efficacités de tagging à partir des “futures” données réelles dans CMS
- Demandra sans doute une modification des algorithmes de tagging : tags négatifs nécessaires (pas prévu pour le moment)
- Trigger soft-lepton dans CMS ?
- Objectif suivant : préparer la physique du top,  $\sigma(tt\bar{b})$ , sans b-tagging, puis avec b-tagging.