B-Id
Status & Plans

S. Greder,

Institut Pluridisciplinaire Hubert Curien, Strasbourg

Outline

- TRFs, SFs How-to
  - Why preferably using SFs ?
- Certification status
- Combined multivariate heavy-flavour tagger
- On-going plans
- Conclusions

TRFs, Scale Factors and al.

Detailed talk by T. Gadfort:

http://www-d0.hef.kun.nl//askArchive.php?base=agenda&categ=a09171&id=a09171s1t27/transparenc

- Many people want to do a **direct tagging** analysis instead of the TRF approach
  - More accurate sensitivity to detector performances, regions, sample dependences, … w.r.t only (Et,eta) based averages
- At first glance, everything seems straight-forward. If you tag a jet, then you apply the scale factor, right ? *But what happens if you don't tag a jet ?* What do you apply ?
- In the TRF approach *you don't know which jet is tagged*, just a probability that could be tagged for each flavor.
- This has been discussed in details by **F. Filthaut**, D0 Note 5481, e.g:

$$P_{\text{event}}^{\text{tag}}(0\ \text{tag}) = \prod_{j-1}^{N_{\text{jets}}}[1 - \mathcal{P}_{\alpha_j}(p_{T_j}, \eta_j)].$$

$$P_{\text{event}}^{\text{tag}}(1\ \text{tag}) = \sum_{j-1}^{N_{\text{jets}}} \mathcal{P}_{\alpha_j}(p_{T_j}, \eta_j) \prod_{i \neq j}[1 - \mathcal{P}_{\alpha_i}(p_{T_i}, \eta_i)],$$

$$P_{\text{event}}^{\text{tag}}(\geq 2\ \text{tag}) = P_{\text{event}}^{\text{tag}}(\geq 1\ \text{tag}) - P_{\text{event}}^{\text{tag}}(1\ \text{tag}),$$

Sébastien Greder

IPHC Institut Pluridisciplinaire Hubert Curien Strasbourg

TRFs, Scale Factors and al.

Scale factors

If you have two jets and both are tagged then the event weight (ignoring taggability SF) is given by:

$$w = SF_b^{1}(pT,\eta) \times SF_b^{2}(pT,\eta)$$

If you are doing a double tag analysis with exactly 2 jets, then you are done.

- What about single tagged events? You can NOT just apply the scale factor to the tagged jet and nothing to the *un-tagged* jet. You need to apply a new scale factor: $!SF_b(pT,\eta)$

- This scale factor is not directly provided by our group, but you can calculate it from the scale factor and the MC TRFs shown below. We can measure this directly in the future

$$!SF_j(p_T,\eta) = \frac{1 - TRF_j^{MC}(p_T,\eta) \times SF_j(p_T,\eta)}{1 - TRF_j^{MC}(p_T,\eta)} = \frac{1 - TRF_j^{Data}(p_T,\eta)}{1 - \frac{TRF_j^{Data}(p_T,\eta)}{SF_j(p_T,\eta)}}$$

Sébastien Greder

IPHC Institut Pluridisciplinaire Hubert Curien STRASBOURG

TRFs, Scale Factors and al.

Conclusions

- **But ...**
  - Situation is more complicated in reality:
    --> same discussion applies to taggability & vertex confirmation !
- **Code**
  - Eventually this will be handled within a dedicated Processor
- **More discussion:**
  - J. BackusMayes @ b-Id, R. Schwienhorst @ Top group

Certification status

Efficiency: method

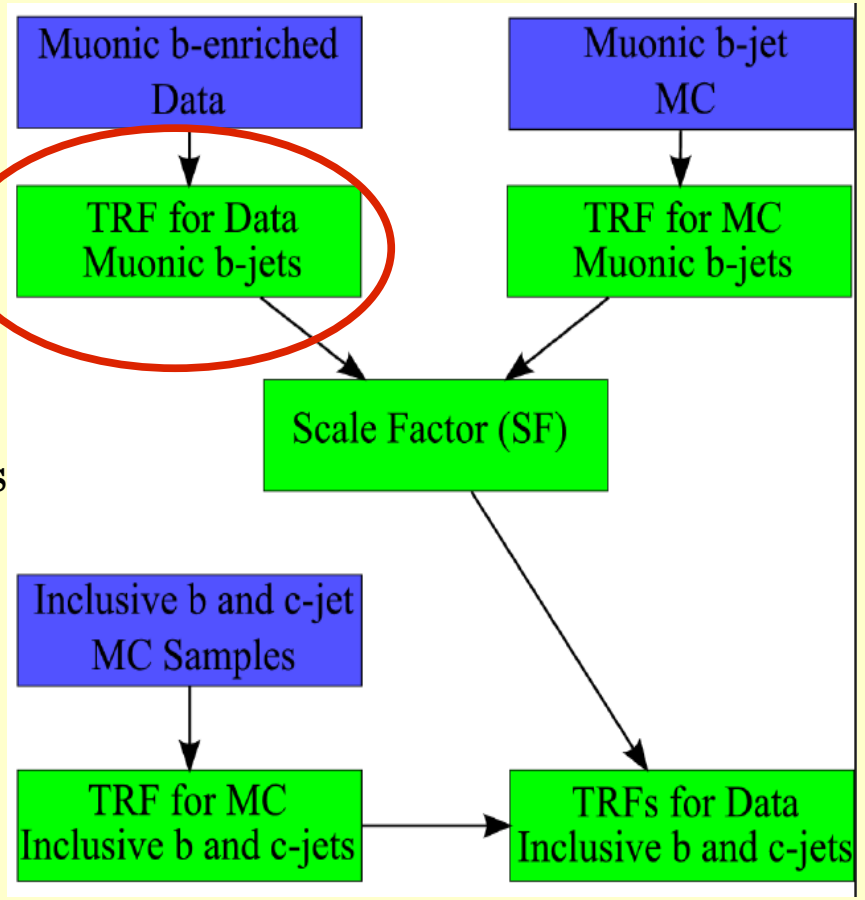
Goal

- Estimate of b & c efficiencies

Measured in *data*

- MUInclusive skim
- System8 with back-2-back jets / muonic-jets
- *Muonic b*-TRFs in data/MC in f(pT,eta)
- *Muonic* data/MC **b-Scale Factor (SF_b)**:

$$\varepsilon_b^{\text{data}} = \frac{\varepsilon_{b\to\mu X}^{\text{data}} \cdot \varepsilon_b^{\text{MC}}}{\varepsilon_{b\to\mu X}^{\text{MC}}} = \text{SF}_b \cdot \varepsilon_b^{\text{MC}}$$



- Apply SF to **inclusive** b & c TRFs to get data efficiencies
- Detailed talk by Tim Scanlon:
  http://www-d0.hef.kun.nl//askArchive.php?base=agenda&categ=a09171&id=a09171s1t26/transparencies

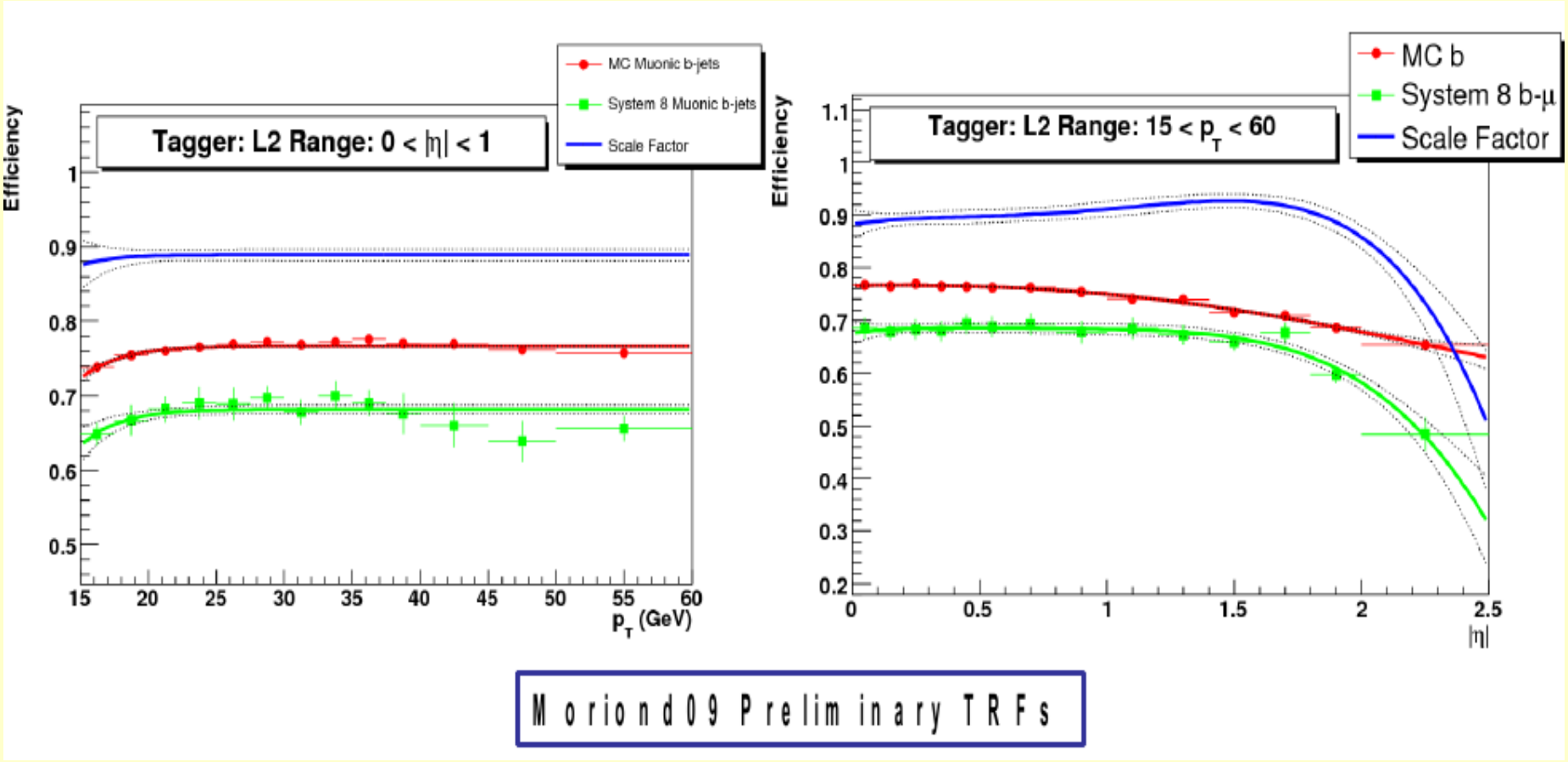
Sébastien Greder

IPHC Institut Pluridisciplinaire Hubert Curien STRASBOURG

System 8 issues (I)

Efficiencies

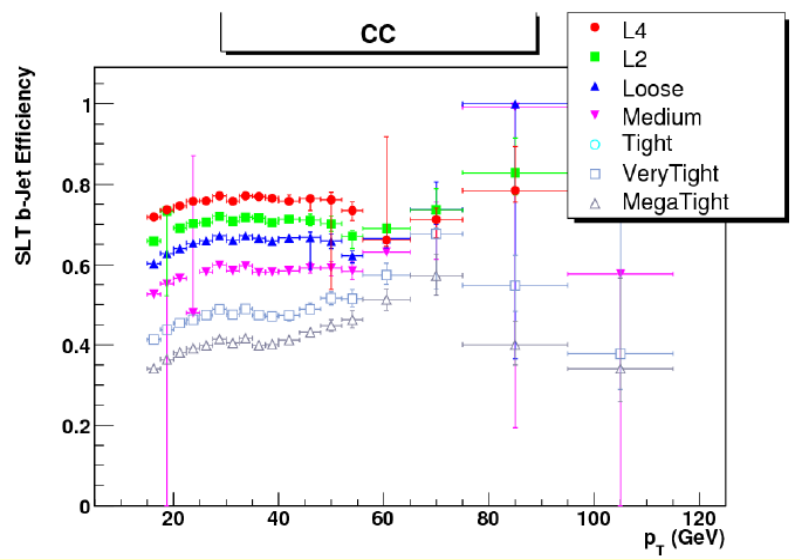
- Suspicious behavior of System8 observed @ hight jet pT (i.e > ~45 GeV)
- Well behaved in eta :)

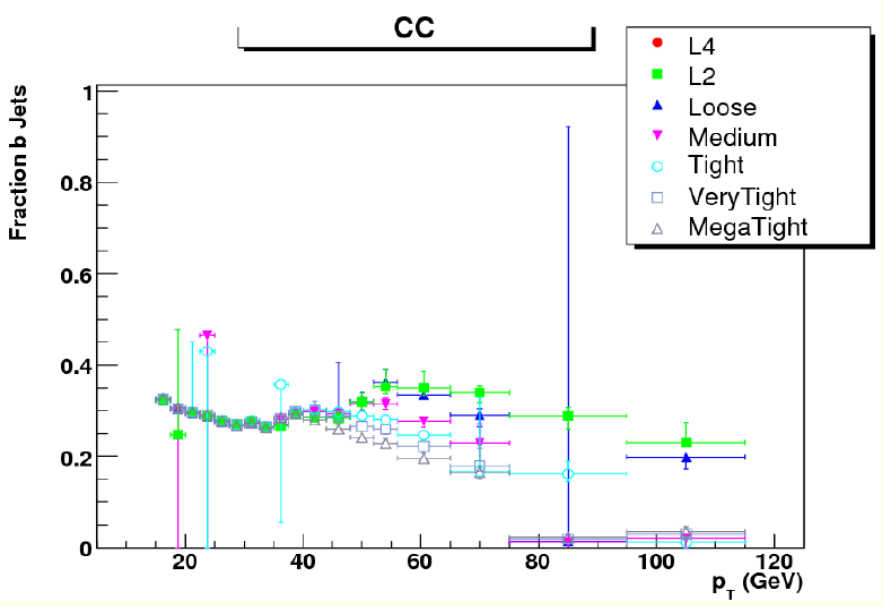


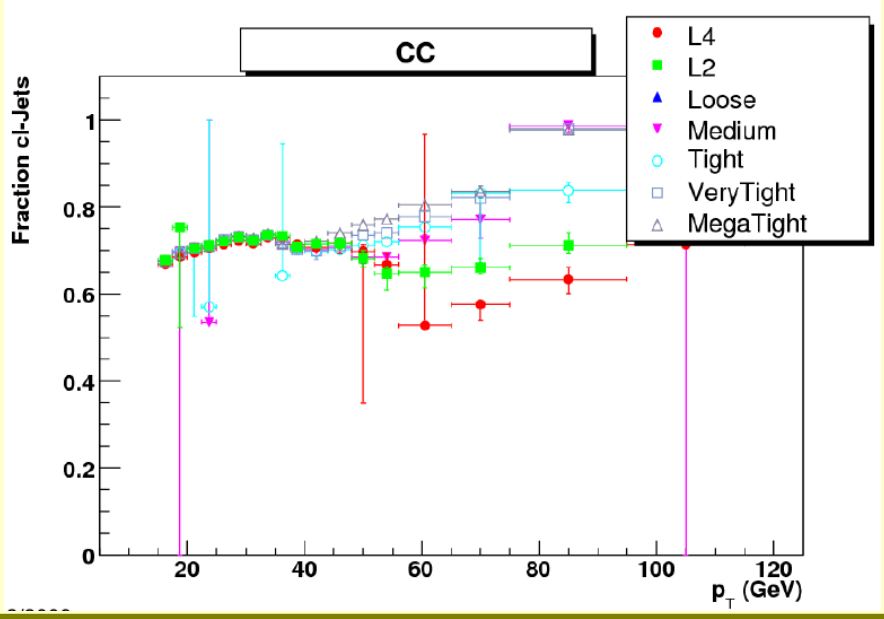
System 8 issues (II)

Flavour fractions

- suspicious behaviour @ hight jet pT
- b and non-b (*i.e cl*) fractions expected to be similar w.r.t OP



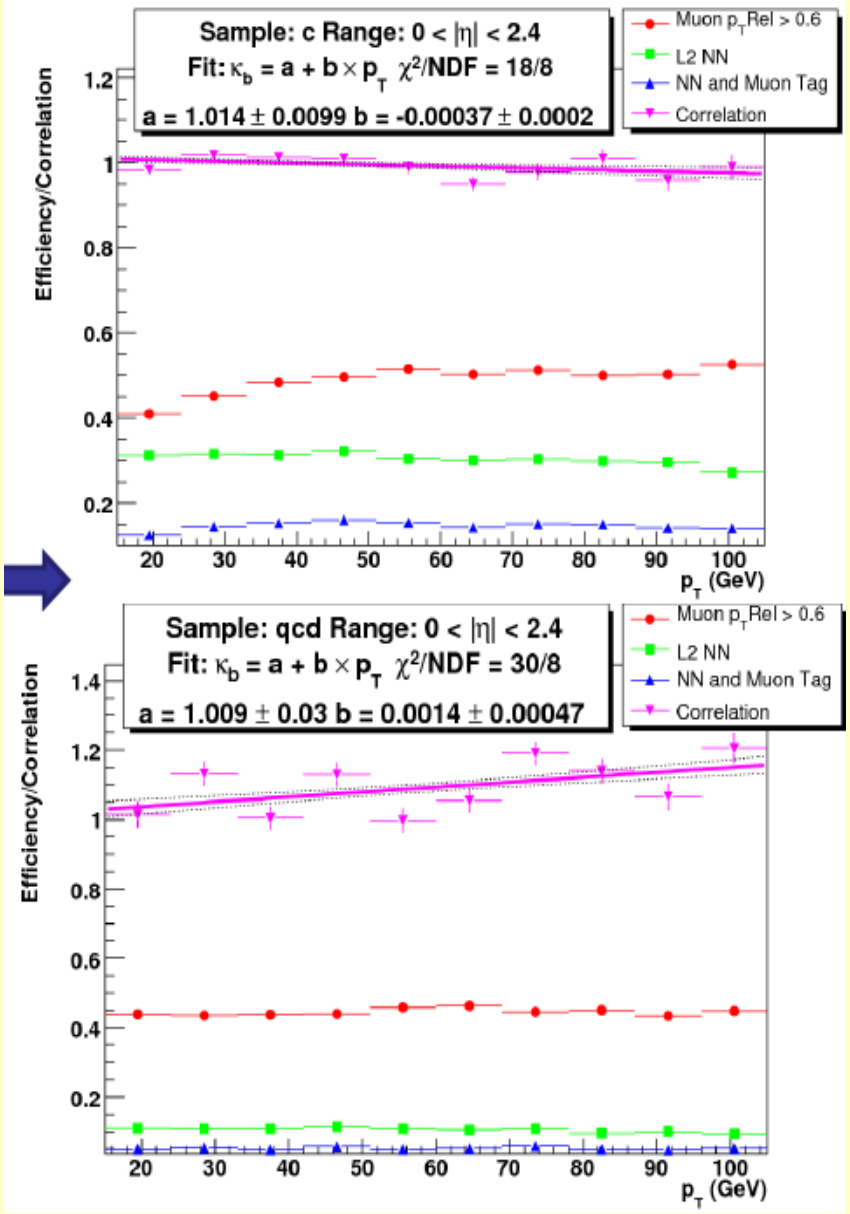




S8 investigations

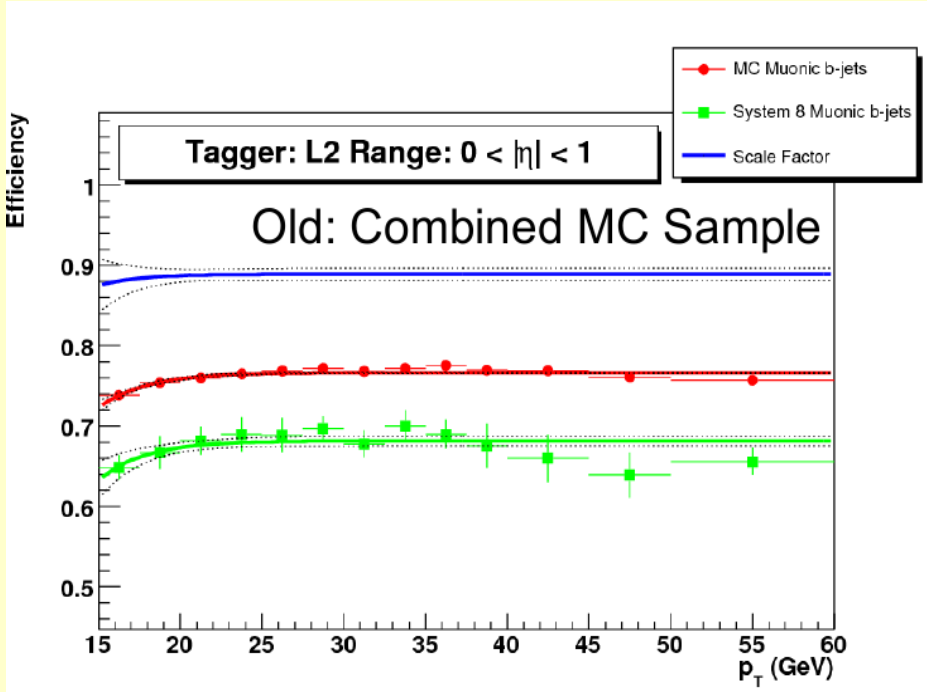
Investigations

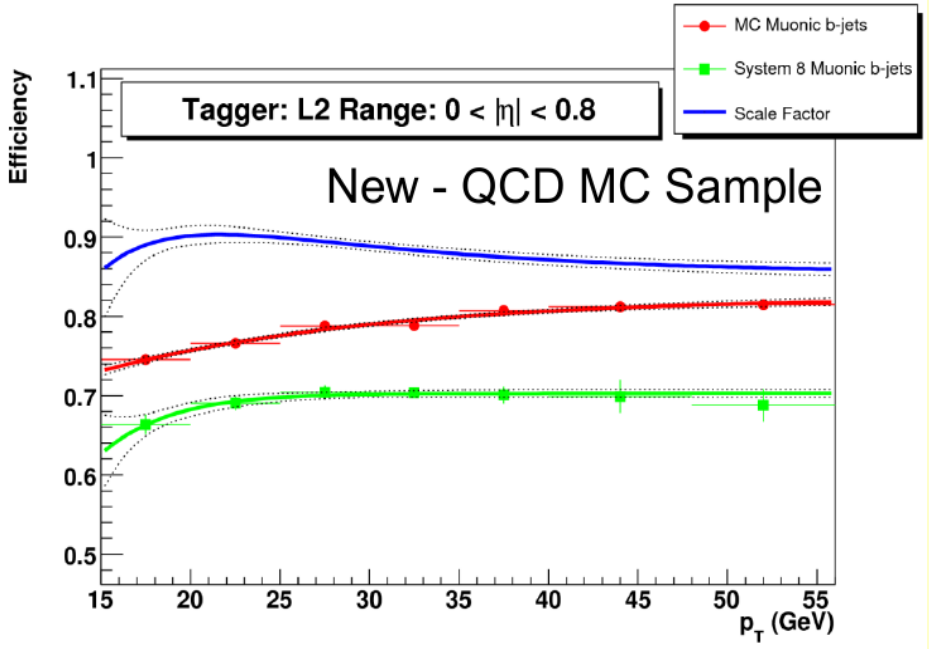
- **Data**
  - Tightened quality cuts
  - Huge QCD inc, bb & cc MC request
- **S8 correction factors**
  - Measure **only** on QCD samples
- **MC muonic b-jet TRF**
  - Measure on QCD sample
- **Away tag and SLT pTRel cuts**
  - Re-optimise based on S8
  - Systematic error
  - Away-tag: Tight
  - SLT pTRel > 0.6



S8 Results (I)

- New results with all improvements now ready
  - smaller errors and more stable result
- Within systematic and statistical errors **no change in scale factors**
- **But:** didn't solve higher pT problem :(

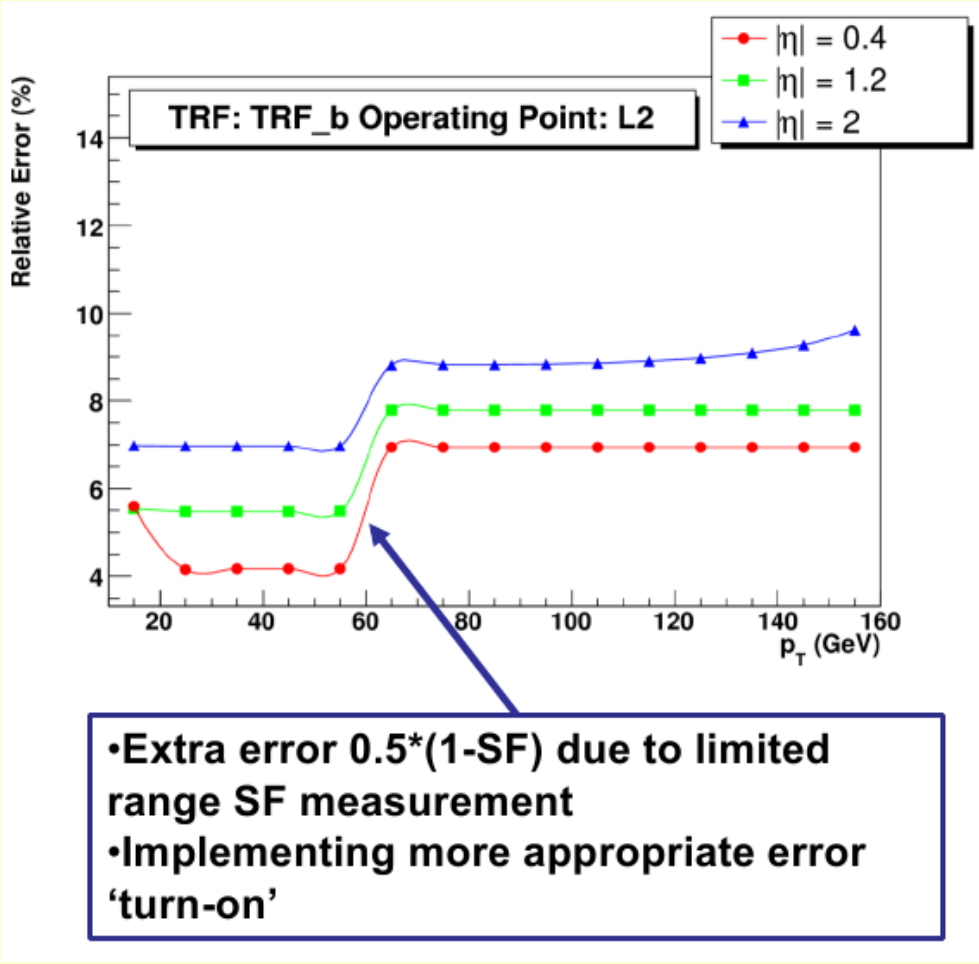




S8 Results (III)

- Updates **halved** the S8 systematic error :)

  !! Good news for top & Higgs physics program !!



Conclusion

- Specific pb @ hight pT still unclear
- CMS has implemented S8

  --> they faced the same issue @ high pT

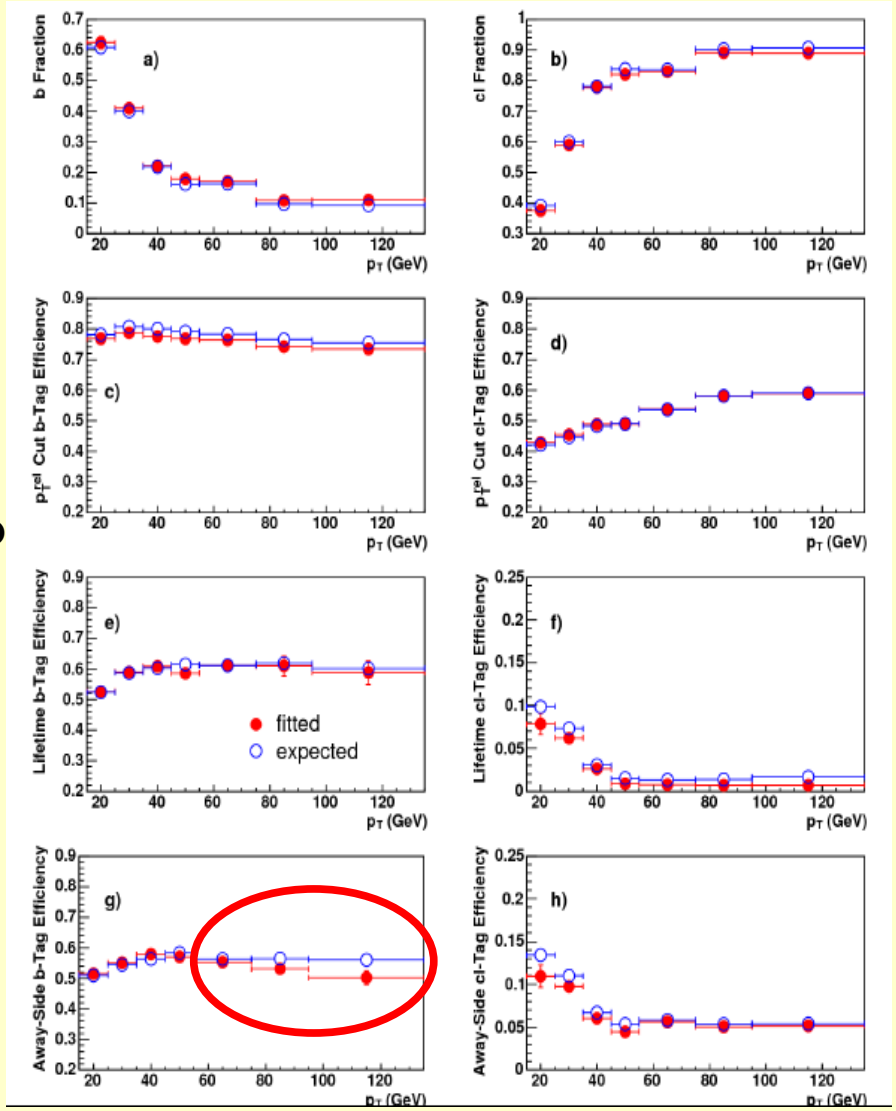
  --> extended S8 system of equations
  with 2 additional correction factors

  --> incorporated a iterative procedure to
  improve convergence of the method

- **Plans:**

  - re-do a p14 study on MC only

  - implement CMS S8 upgrades



Sébastien Greder

Combined multivariate tagger

Sébastien Greder

IPHC
Institut Pluridisciplinaire Hubert Curien
STRASBOURG

Introduction

Method

- Exploit full objects' properties: CSIP, JLIP,SVT and Track-jet
  - Specialize learning technic **for each type** of object
- Test / Combine different multivariate techniques
- Specialize learning for 2 cases: b vs. light **(BL)** and b vs. c **(BC)** jets

Datasets

- Re-cafed MC:
  - directbb/cc (20-320 GeV)
  - Inclusive pythia QCD (20-320 GeV)
    - New dataset has >= 14M taggable jets !

Package

- TMVA package, latest release (3.9.6, dec. 06, 2008)

Sébastien Greder

IPHC Institut Pluridisciplinaire Hubert Curien STRASBOURG

SVT

Method

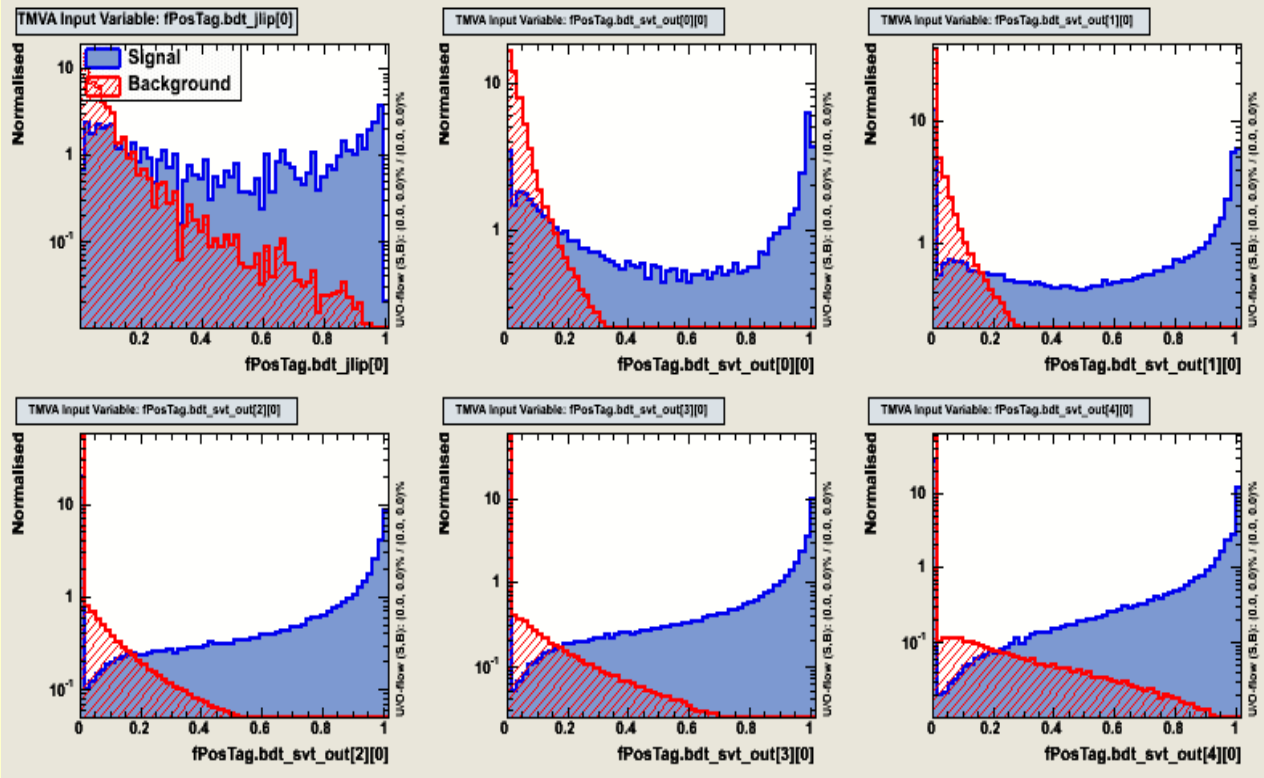
- Different SVTs algorithm settings lead to different objects:
  *Superloose **(SL)**, MediumLoose **(ML)**, Loose_xtracks **(LX)**, Loose **(L)** & Tight **(T)***
  - see: http://www-d0.fnal.gov/D0Code/source/btags_cert/cuts/p14.05.01/
    - **SL:** track Pt cut: 0.5, track DCA cut: 0.0, track Max Chi2: 15.0
    - **LM:** track Pt cut: 0.5, track DCA cut: 1.5, track Max Chi2: 15.0, **NEW**
    - **LX:** track Pt cut: 0.5, track DCA cut: 3.0, track Max Chi2: 10.0
    - **L:** track Pt cut: 1.0, track DCA cut: 3.0, track Max Chi2: 10.0
    - **T:** track Pt cut: 1.0, track DCA cut: 3.5, track Max Chi2: 3.0
- **Differences in properties, e.g:** $sdls_{xy}$, mass, #tracks, pTrel, eta-phi size, ...
  - Random Forest (+ bagging): best performance when # variables > ~15-20 *(see backup slides for settings)*
  - Very little sensitivity to non-relevant variables
  - Very fast training rate (e.g: ~10min for 500k signal & 7M background events !)

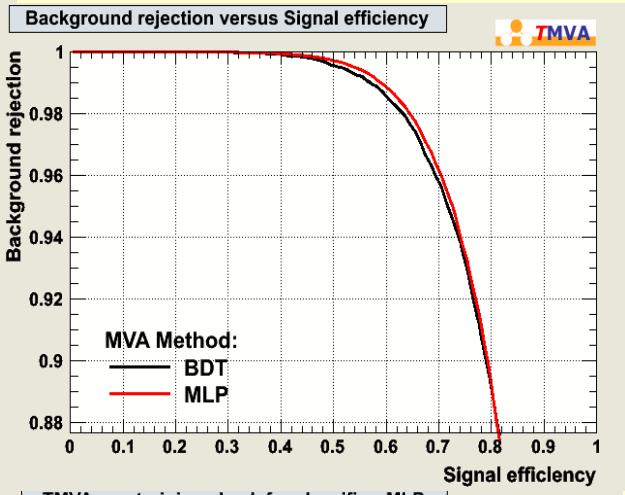
  → dedicated learning for each SVT type
- Lifetime, kinematic, shape, opening angle, missing transverse momentum, #tracks, charged pT fractions, track-jet kinematic, ... a complete *(see back-up slides for complete properties)*

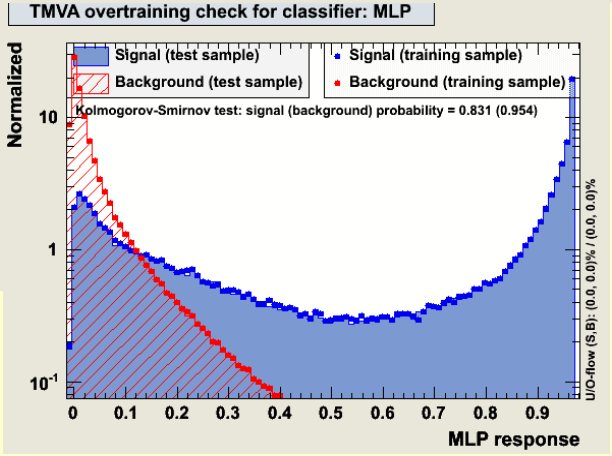
Combined SVT & JLIP BDT

Combine BDTs in a neural network (MLP)

- MLP was found to be more powerful than a Boosted/Bagged Decision Tree
- Train an MLP NN with JLIP-CSIP / SVT BDTs + muon variables (pT, pTrel,chi2,deltaR)





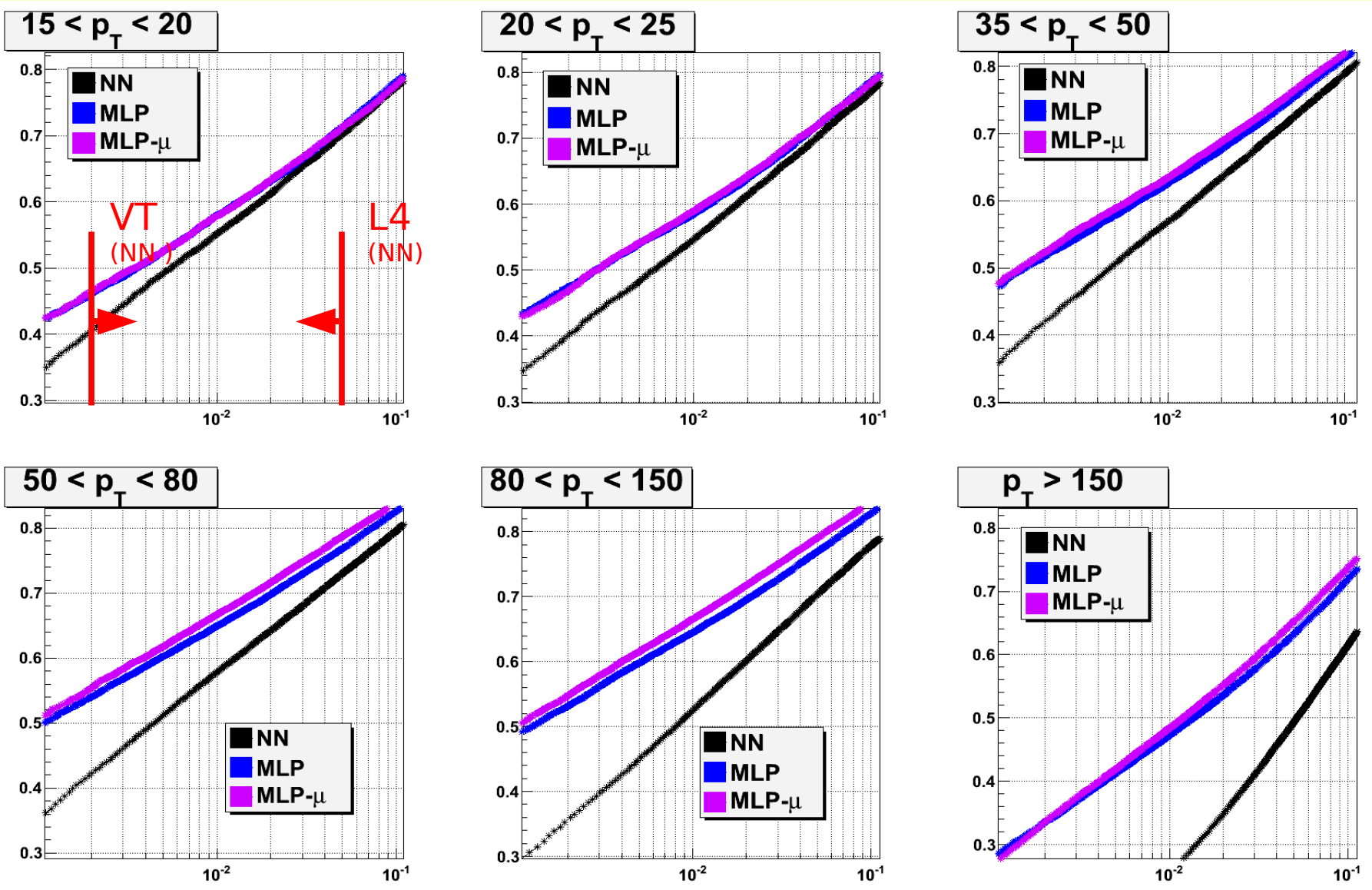


Sébastien Greder

IPHC
Institut Pluridisciplinaire Hubert Curien
STRASBOURG

Combined SVT & JLIP BDT

MLP performances vs. pT

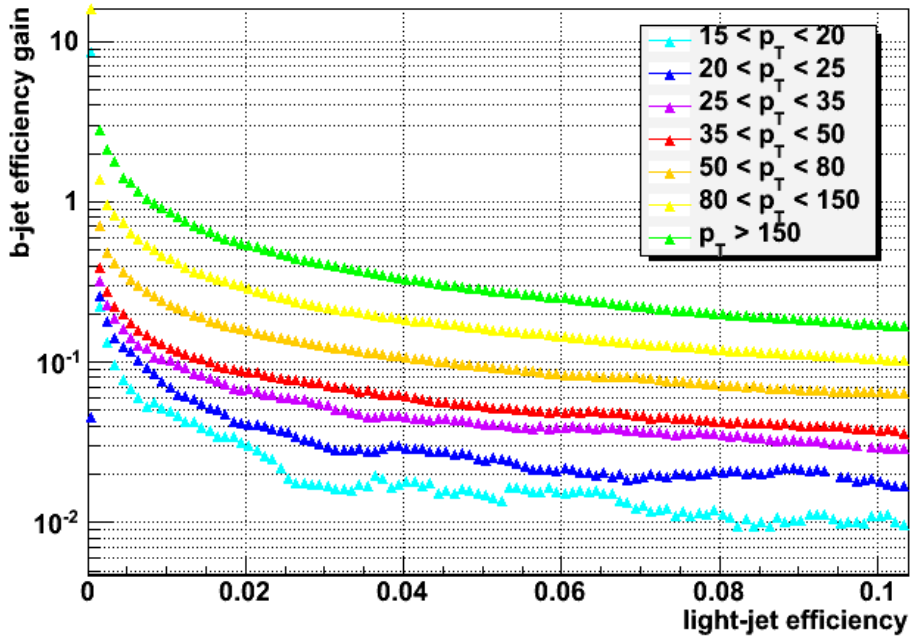


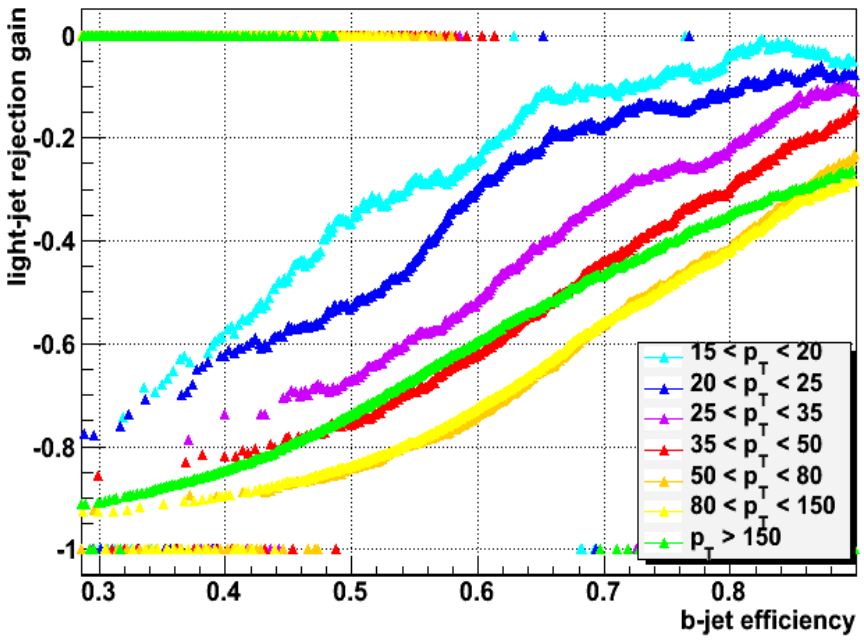
Sébastien Greder

Performance gain w.r.t NN BL

Performances gains

- **Gain** = eff(BL) – eff(NN)/ eff(NN)
- **@ fixed** light-jet efficiency / b-jet efficiency



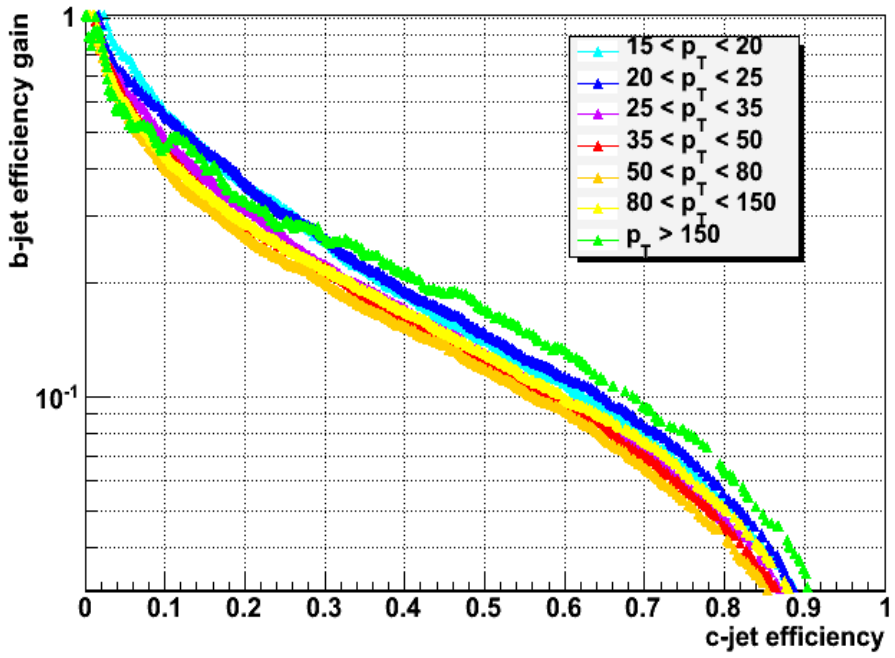


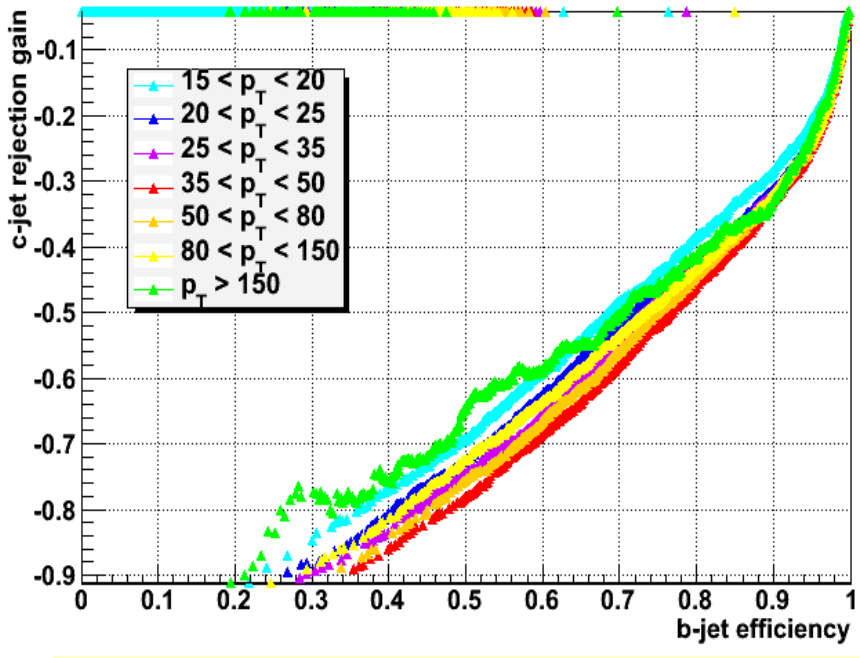
Sébastien Greder

Performance gain w.r.t NN BC

Performances gains

- **Gain** = eff(BC) – eff(NN)/ eff(NN)
- **@ fixed** c-jet efficiency / b-jet efficiency



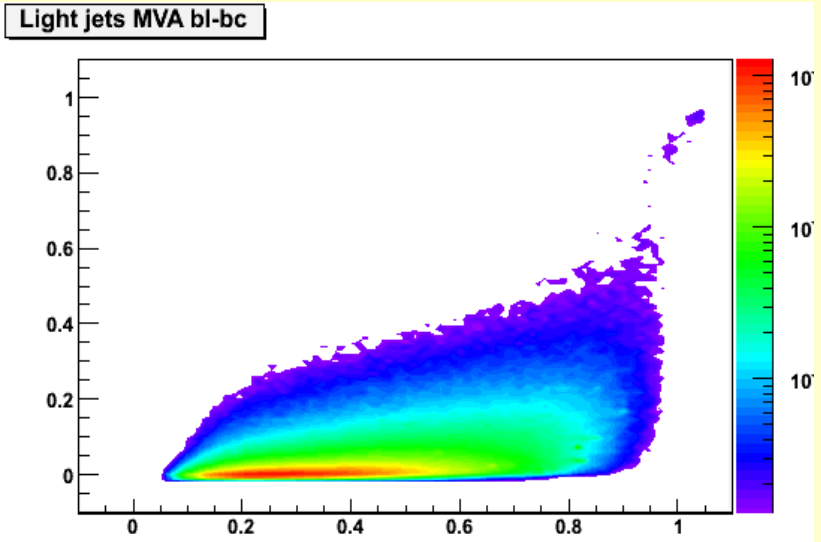


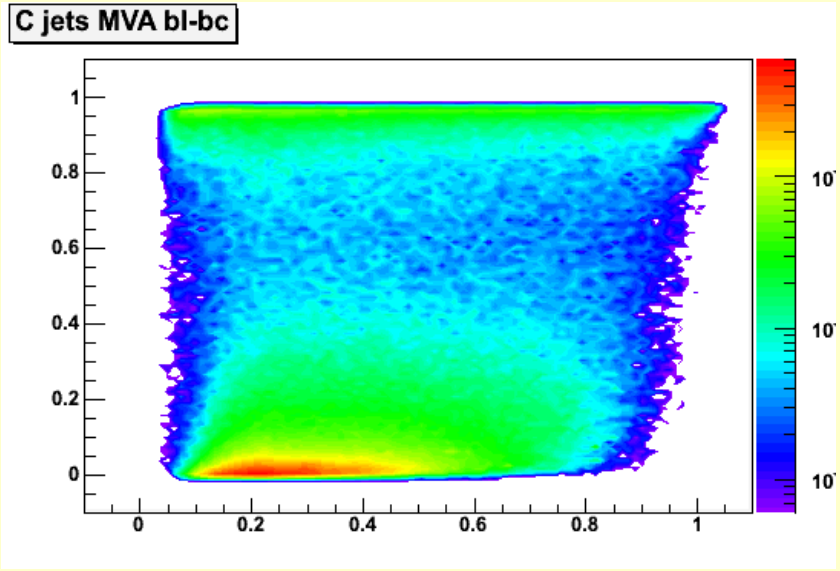
Sébastien Greder

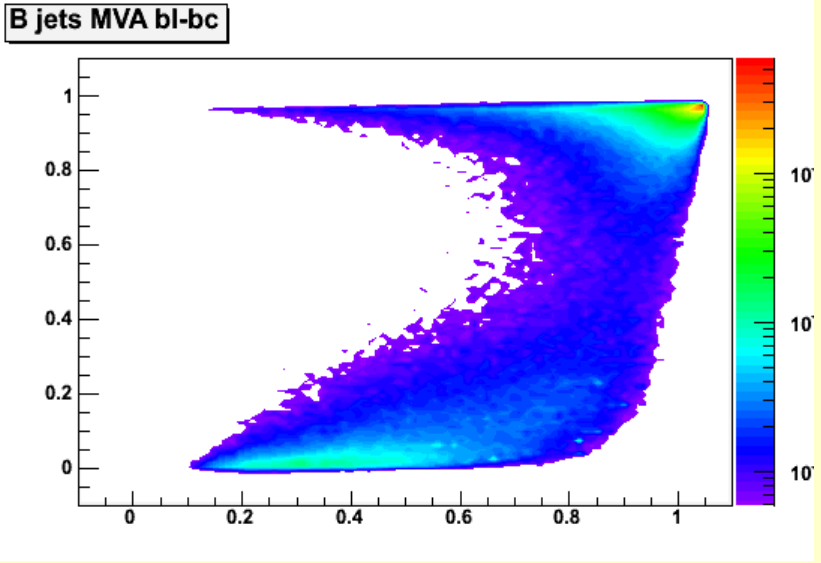
B,C,L 2D BL-BC outputs

MVA BL-BC outputs

(integrated over pT,eta, normalised to unit area)







Sébastien Greder

On-going tasks

Several other topics are being tackled

- **Taggability** (Y. Peters, ~2weeks)
  - Flavour dependancies
  - Tau, electrons
- Using non-prompt J/Psi ->mumu for b-tagging efficiency measurement.
- Ongoing effort with JER group to improve vertex-confirmation efficiency
- Dataset-dependent fake rate measurement a. la. system8 developed for hbb analysis can be used by others as well
- Fake-track killer test
- Vbb/Hbb discriminant: *see Florian Miconi talk*
- **B fragmentation systematic**
  - Already studied by Top group, see D0 Note 5325, dedicated Processor already exists
  - Will include in syst. errors

Conclusions

- Converge to proper use-case of SFs / Event weight
- Improved System8
  - More stable
  - Reduced systematic errors by ~50% !
  - Still need studies to tackle high pT issues
- Revisited b vs. light discrimination
  - Huge improvements relative to NN, esp. @ hight pT
  - 15-50% signal / background gain / rejection !
- Several on-going tasks to refine performances and errors measurements
- Full release after Moriond: test asap (not in June :) !)
- Expect major impacts in many analysis : testbed : WH

Back-up

TMVA Settings

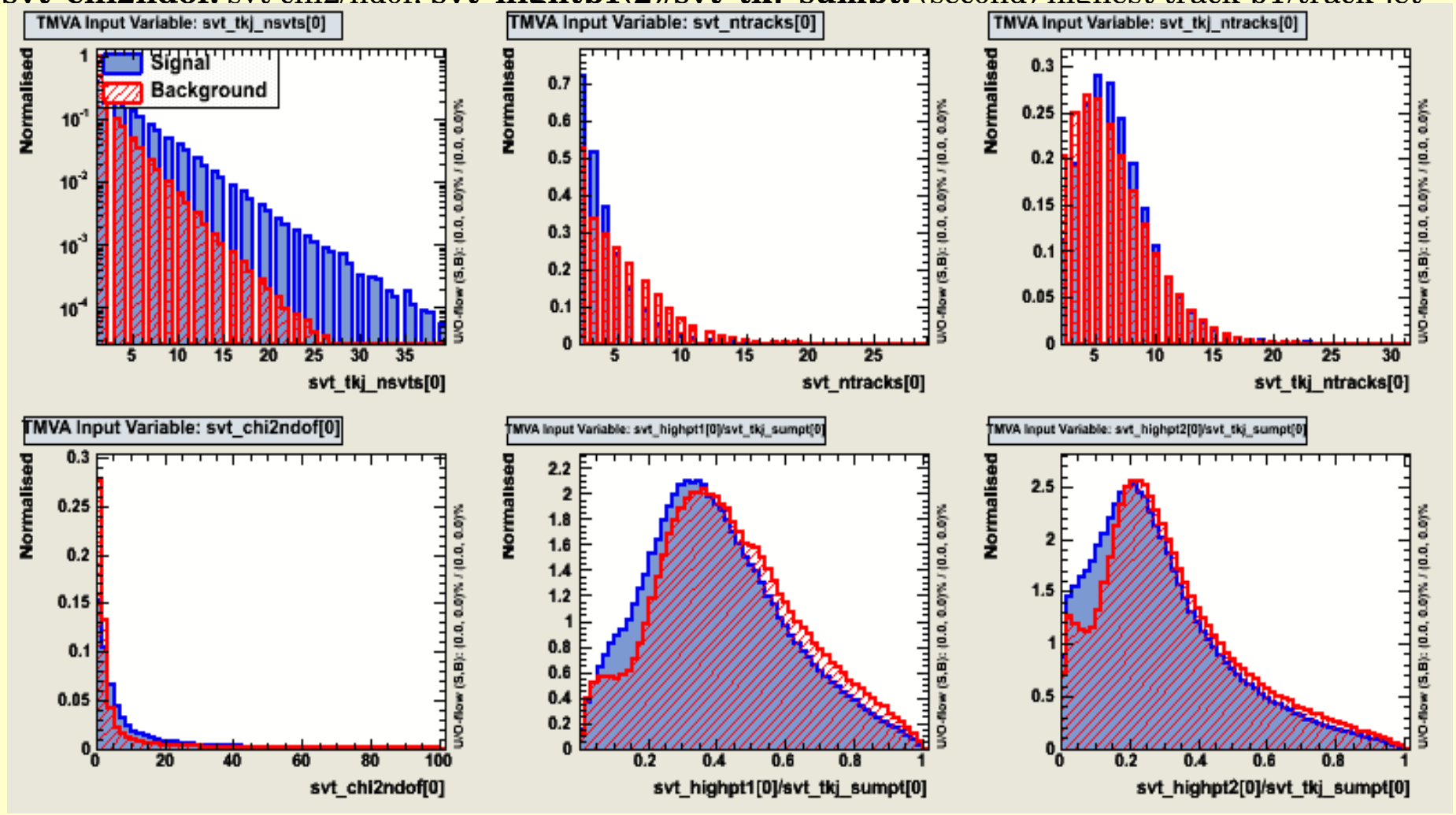
SVT SL,LX,LT BDT

```
Cut: # svt sl/lx/l/t ntracks >=2

Method: TMVA::Types::kBDT

Parameters:
   "H:V:NTrees=5:UseRandomisedTrees=true:UseNvars=all:SeparationType=GiniIndex:nCuts=20:
   UseYeNoLeaf=False:NoNegWeightsInTraining=True"
```

SVT SL Properties (I)

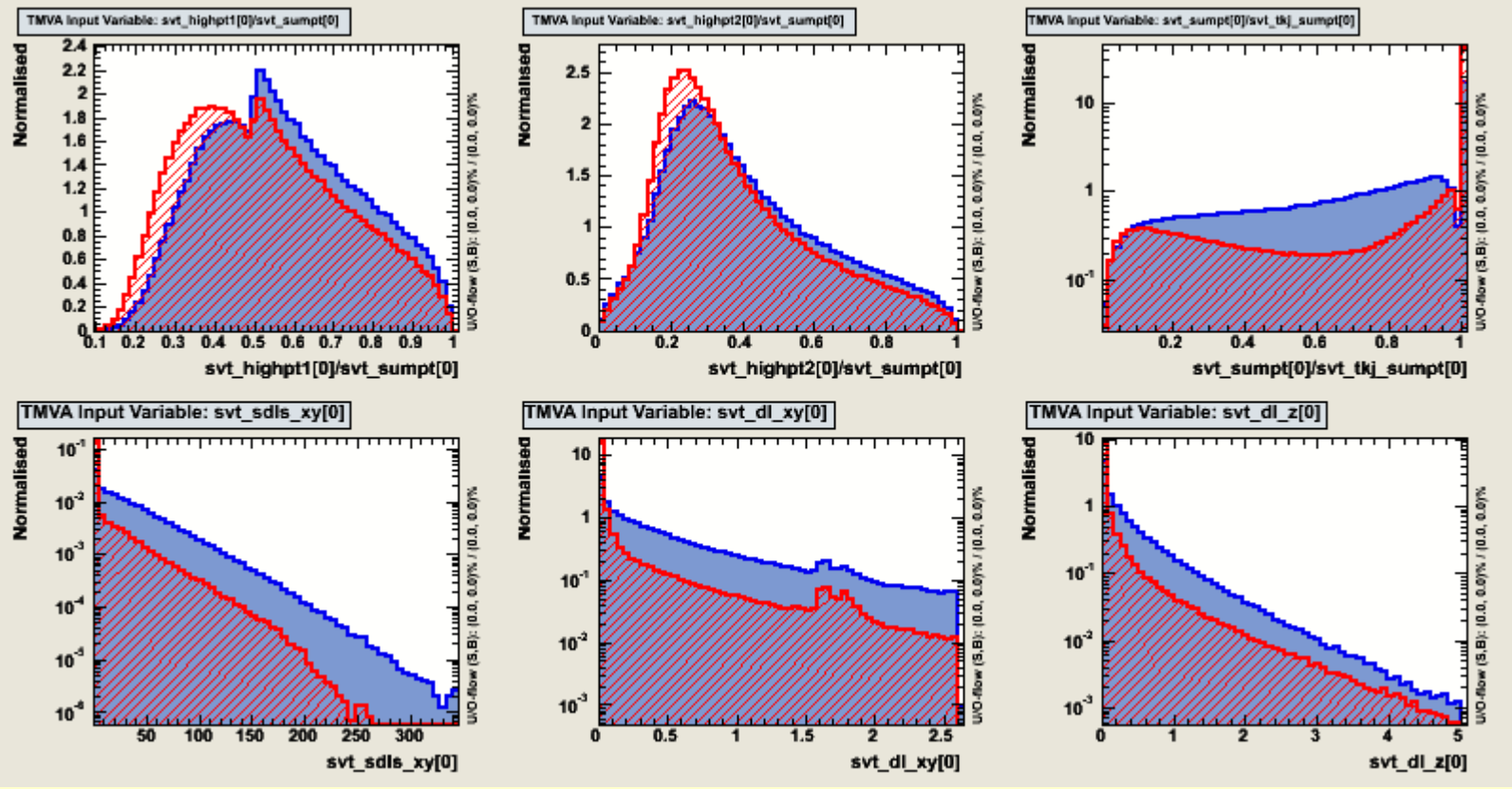
svt_tkj_nsvts: # of svts in track-jet, **svt_ntracks:** # tracks in svt, **svt_tkj_ntracks:** # tracks in track-jet,

svt_chi2ndof: svt chi2/ndof, **svt_highpt1(2)/svt_tkj_sumpt:** (second) highest track pT/track-jet



Sébastien Greder

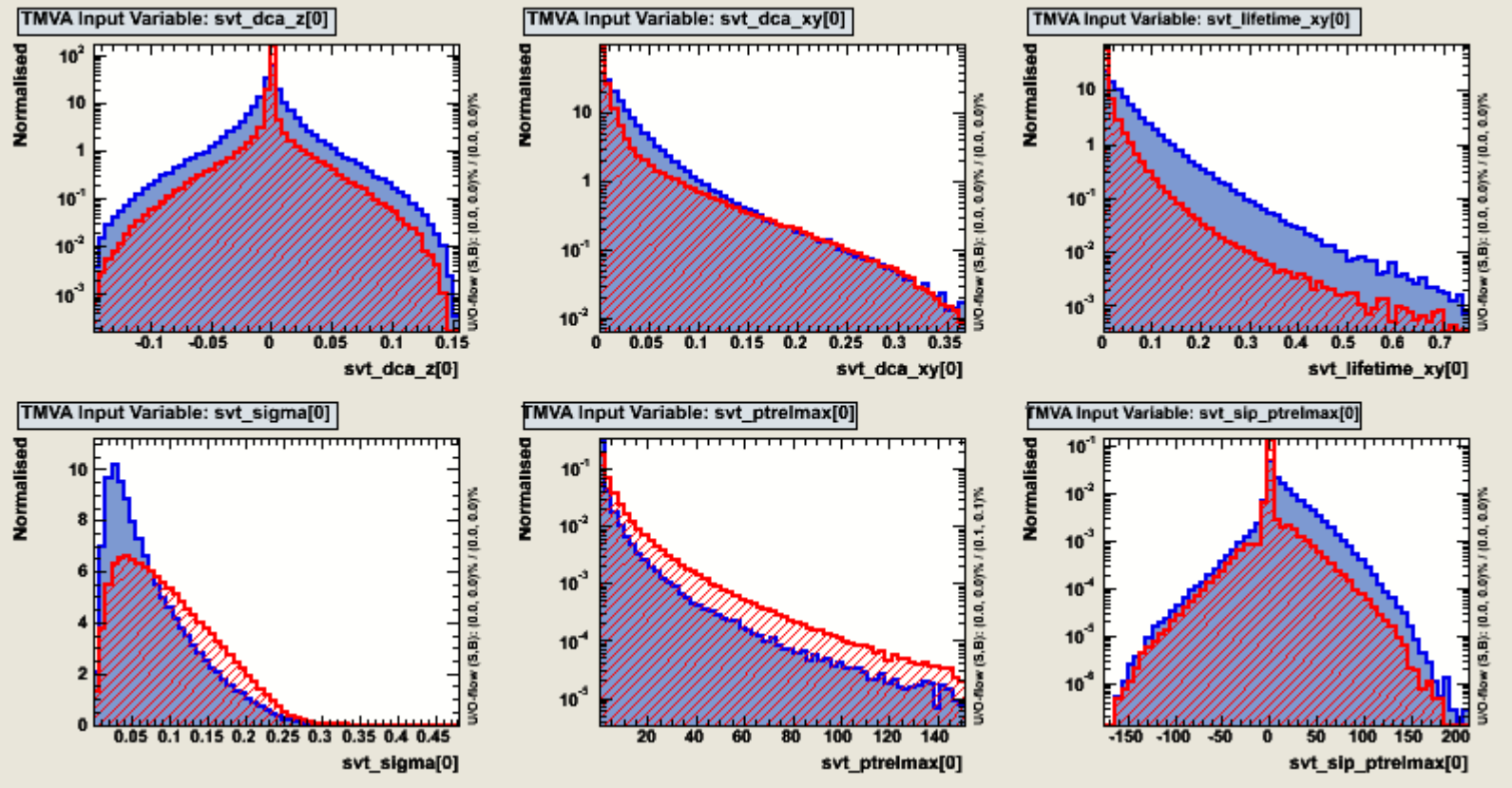
SVT SL Properties (II)



Sébastien Greder

IPHC
Institut Pluridisciplinaire Hubert Curien
STRASBOURG

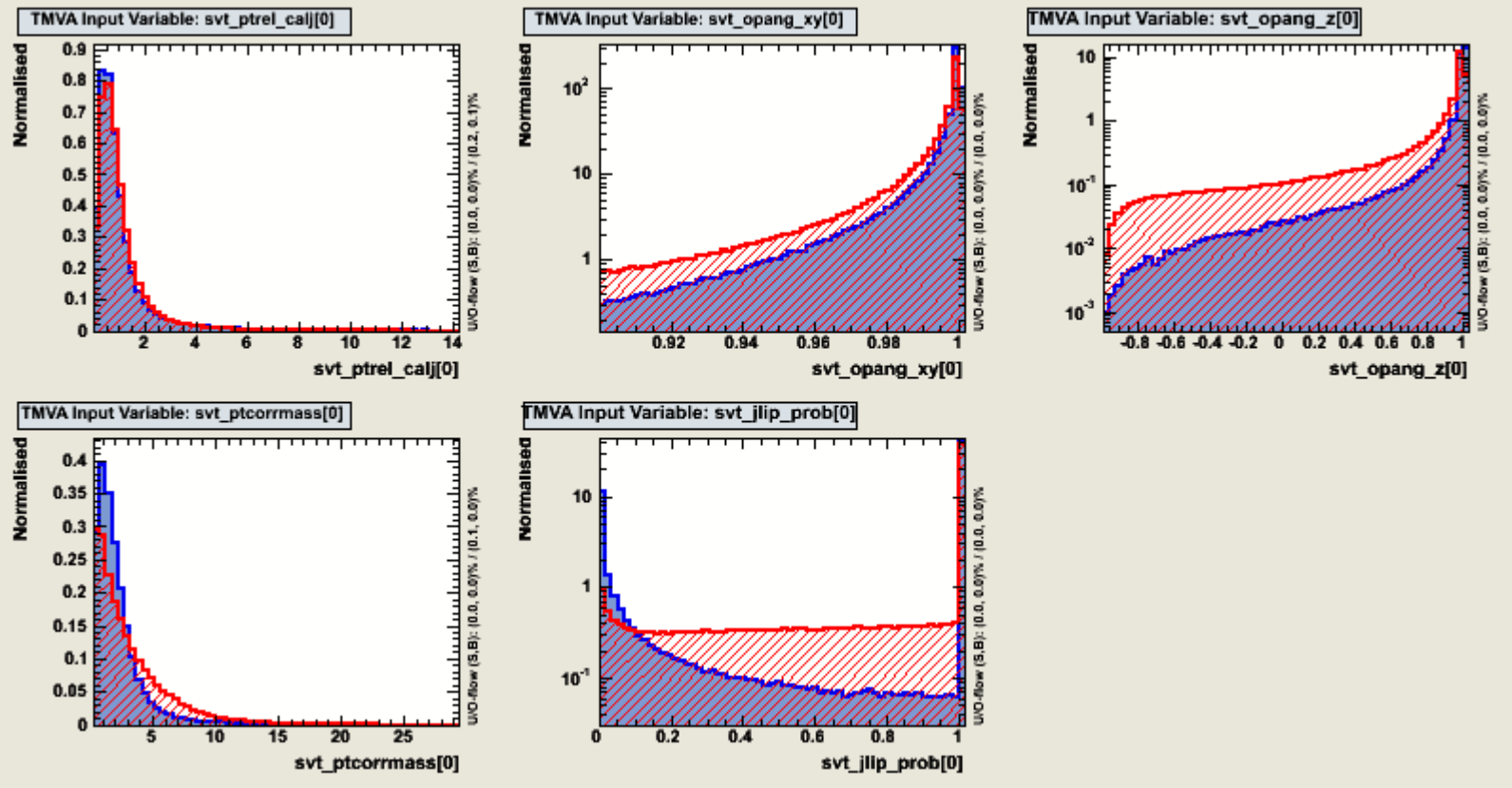
SVT SL Properties (III)



Sébastien Greder

IPHC
Institut Pluridisciplinaire Hubert Curien
STRASBOURG

SVT SL Properties (IV)



Sébastien Greder

IPHC Institut Pluridisciplinaire Hubert Curien STRASBOURG