Estimating jet energy calibration factors

- Introduction
  - Jet energy corrections at CMS: factorized multi-level approach
- Method
  - Estimation of jet energy calibration factors with top quark events
- Results obtained with CMSSW_1_6_9 (TOP-PAS-07-004)
  - Short summary of previous results
- Present status in CMSSW_2_1_9
  - Estimation of the jet energy calibration (JEC) factors in (p_T,η)-bins
  - Combined estimation of the JEC factors and the top quark mass
- Outlook towards CMSSW_2_2_X

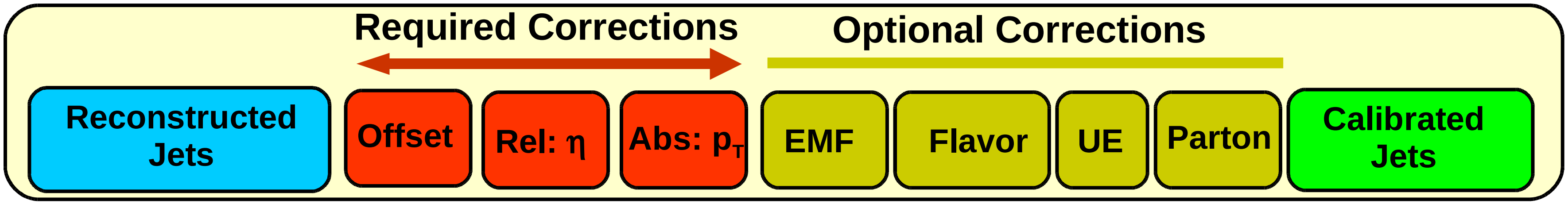
Petra Van Mulders

IIHE-Top Quark Meeting
November 24th 2008

Introduction: JEC at CMS

- **Goal of the jet energy correction is to relate the jet energy measured in the detector to the energy of the final state particle jet or parton jet.**

- **Plans for jet energy corrections at CMS: factorized approach** (PAS JME-07-002)



1. **Offset:** correction for pile-up and electronic noise
2. **Relative (η):** correction for variations in jet response with η relative to a control region
3. **Absolute (p_T):** correction to particle level versus jet p_T in the control region
4. **EMF:** correction for variations in jet response with electromagnetic energy fraction
5. **Flavor:** correction to particle level for different types of jets (light quark, c, b, gluon)
6. **UE:** correct ion for underlying event energy due to soft interactions involving spectator partons
7. **Parton:** correction to parton level

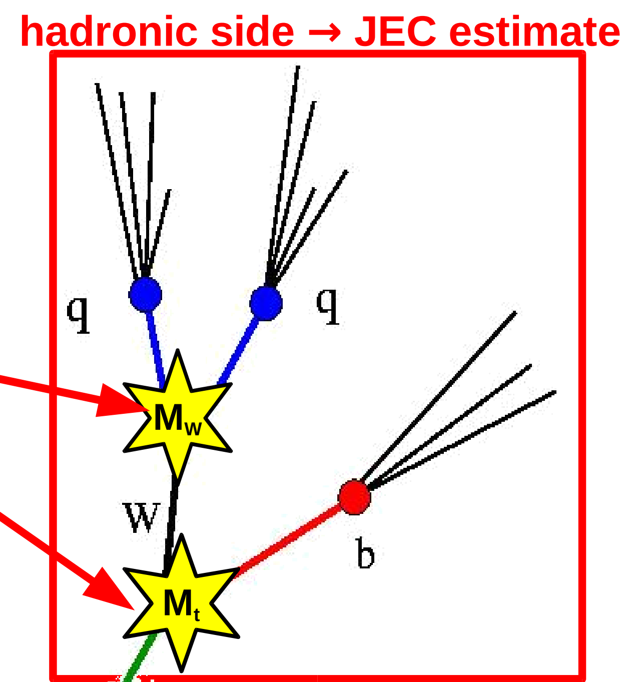
- **Top quark events:**
  - **Possible to provide a combined jet energy correction for: 2+3+5+7**
  - **One can also apply jet energy corrections obtained with other events and use top quark events for the validation of the applied corrections**

Introduction: JEC from top events

- **Semi-leptonic channel:**
  - **Leptonic side used to select the event**
  - **Hadronic side is used to estimate the jet energy calibration factors**

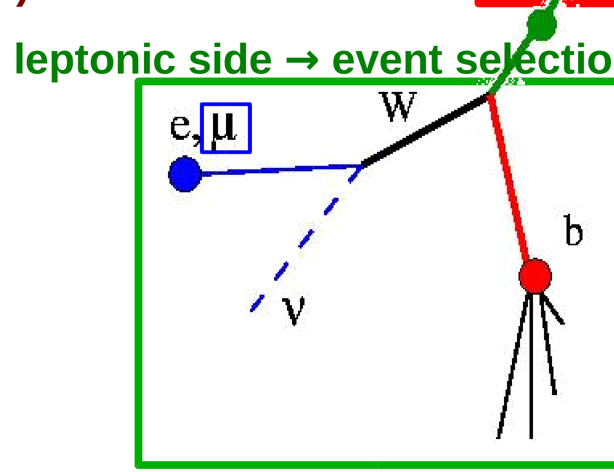
- **On the hadronic branch 2 mass constraints:**
  - **m_W = 80.399 ± 0.025 GeV/c^2 (precision: 0.03%)**
  - **m_{top} = 172.4 ± 1.2 GeV/c^2 (precision: 0.7%)**

➔ **apply the mass constraints on the event by means of a kinematic fit and estimate the jet energy scale**



Example of selection cuts in 1_6_9 (14 TeV):

- **p_T(jets)>40GeV,|η| <2.5**
- **p_T(μ)>30 GeV,|η|<2.1**
- **μ isolated:** **(tracker+calorimeter isolation)**
- **non-overlapping jets:** **ΔR(jet i,jet j) > 1.0**
- **μ separated from jets:** **ΔR(jets,μ) > 0.5** ⟷ **more info in backup (separate study)**



Kinematic fit in CMSSW

CMS Note 2006/023

- **Package: PhysicsTools/KinFitter (originally from Aleph and BaBar)**

- **Our knowledge of the observed event comes from measured objects in the final state (i = jets, lepton, 'neutrino').**
  - this can be summarized as $\mathbf{p}_i = \{ E_i , \theta_i , \varphi_i \}$ (for example)
  - together with the covariance matrix $\mathbf{V}_i$ for each object i
- **Extend this knowledge $\mathbf{p}_i$ and $\mathbf{V}_i$ by assuming some hypothesis for the event**
  - for example : $\mathbf{m_{jj} = m_W}$ & $\mathbf{m_{jjb} = m_t}$
- **Add Lagrange multipliers λ_k in the χ^2 equation to incorporate these hypothesed constraints in our knowledge of the event ($\Delta\mathbf{p} = \mathbf{p}^{fit} - \mathbf{p}^{measured}$)**

$$\chi^2(\mathbf{p}^{fit}) = \Delta\mathbf{p}^T \mathbf{V}^{-1} \Delta\mathbf{p} + 2 \sum \lambda_k f_k(\mathbf{p}^{fit},\mathbf{a})$$

  - where we have the *m* constraint functions f_k and unmeasured parameters $\mathbf{a}$
  - for the true measured and unmeasured parameters → $f_k(\mathbf{p}_{true},\mathbf{a}_{true}) = 0$
- **If the constraints are non-linear an iterative procedure is used to solve them**
  - the equation $f_k(\mathbf{p},\mathbf{a})=0$ are linearized in each iteration step (*Taylor expansion*)
  - the χ^2 equation is minimized ($\partial\chi^2/\partial\mathbf{p}=0$, $\partial\chi^2/\partial\mathbf{a}=0$, $\partial\chi^2/\partial\lambda_k=0$) and solved
  - the iteration stops when some pre-defined convergence criteria are fulfilled
- **A $P(\chi^2)$ is returned by the kinematic fit, reflecting the probability that the constraints are fulfilled**

Method

- According to the chosen jet-parton association, the 3 jets coming from the hadronic top decay are used in an **event-by-event kinematic fit**
- Jet resolutions(uncertainty on jet parameters) are parametrized versus p_T and η
- The constraints $m_W^{rec} = m_W^{world}$ and $m_t^{rec} = m_t^{world}$ are true at parton level

- Before the kinematic fit is applied, the reconstructed jet energies are altered by a factor ΔE_b (for the b-jet), ΔE_{j1} and ΔE_{j2} (for the 1st and the 2nd light jet)
  - **E/|p| is kept constant** when altering the jet energies
  - The $P(\chi^2)$ returned by the kinematic fit is translated in a $\chi^2(\Delta E_b, \Delta E_{j1}, \Delta E_{j2})$
- This step is repeated for correction factors between e.g. ± 50%, in this way a whole range of jet energy correction factors is scanned
- The best estimate of the jet energy correction factors is found by **minimizing the 3D-function** $\chi^2(\Delta E_b, \Delta E_{j1}, \Delta E_{j2})$

- To reduce the **process background** a tight event selection is applied
- A likelihood ratio or MVA discriminator is used to identify the correct jet combination
- A cut on the discriminator is made to reduce the **combinatorial background**
- To reduce contributions from **mis-reconstructed events** cuts are made on the probability returned by the kinematic fit

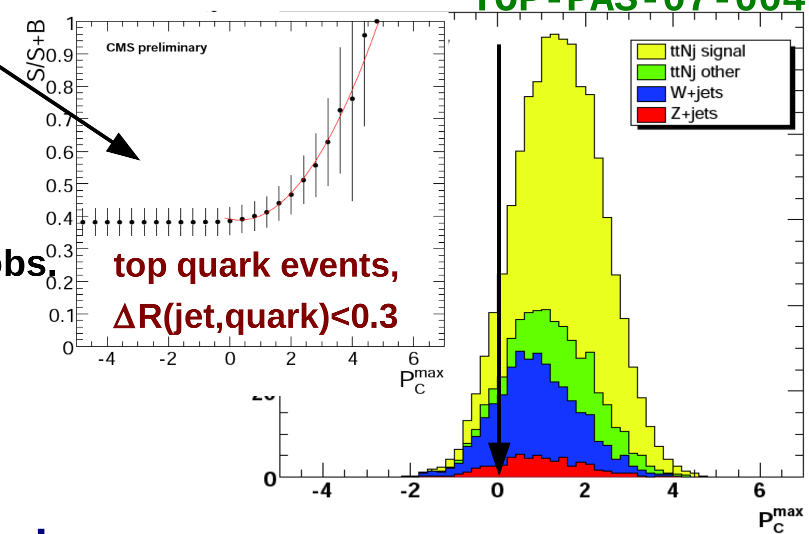
JEC factors with CMSSW_1_6_9

Remark: in CMSSW_1_6_9 both light jet energy correction factors were required to be equal

- **To identify the correct jet combination four observables are combined into a LR:**
  - $p_T^{had\ top} / \langle p_T^{had\ top} \rangle$
  - $(p_T^{b1}+p_T^{b2})/(p_T^{l1}+p_T^{l2})$
  - $\Delta R(l1,l2)$
  - b-value(b1)+b-value(b2)

S: correct jet comb.
B: wrong jet comb.

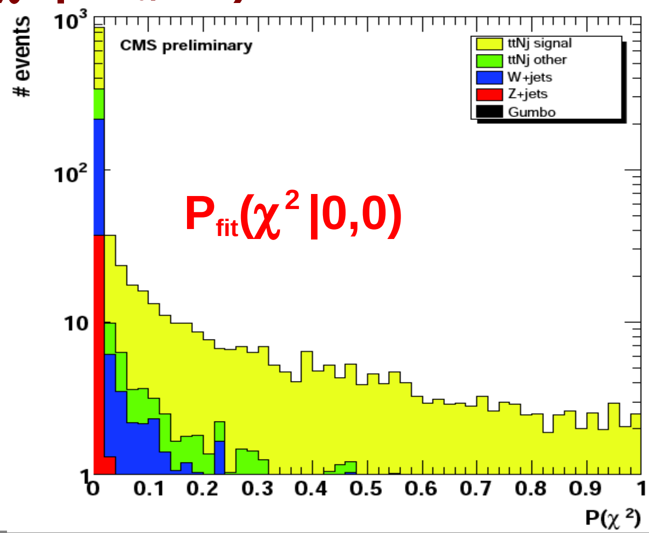
TOP-PAS-07-004

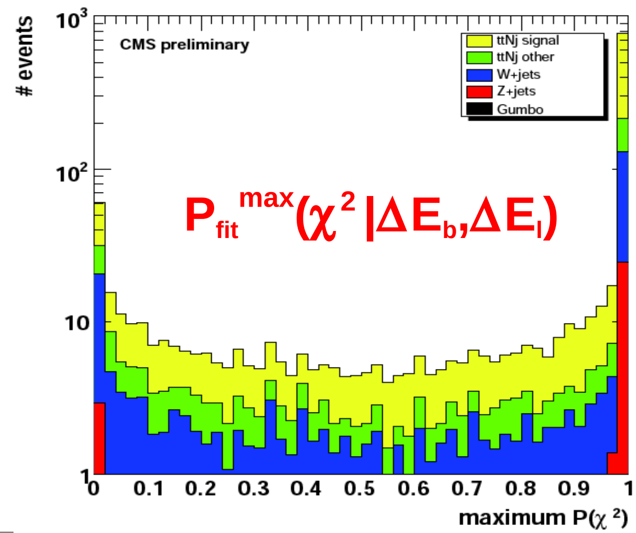


- $P_c^{max} = \max(\sum_i \log(L_i(x_i)))$ with $L_i(x_i)=S_i(x_i)/B_i(x_i)$, i=obs.
- **To purify event sample:** $P_C^{max} > 0$

- **For each event** $P_{fit}(\chi^2|0,0)$ **(no JES corrections)**
- **For each event and over whole scanned JES corrections range** $P_{fit}^{max}(\chi^2\ |\Delta E_b,\Delta E_l)$ **is calculated**

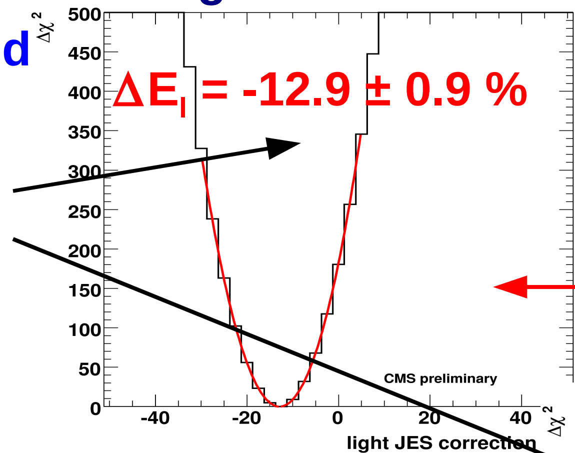
- **Removal of mis-reconstructed events:** $P_{fit}(\chi^2|0,0) > 0.01$
- **Requiring the JES corrections are found in the scanned range:** $P_{fit}^{max}(\chi^2\ |\Delta E_b,\Delta E_l) > 0.98$

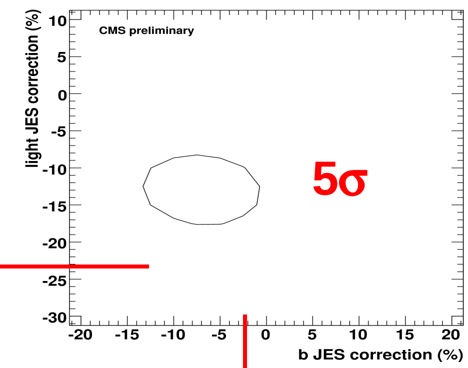


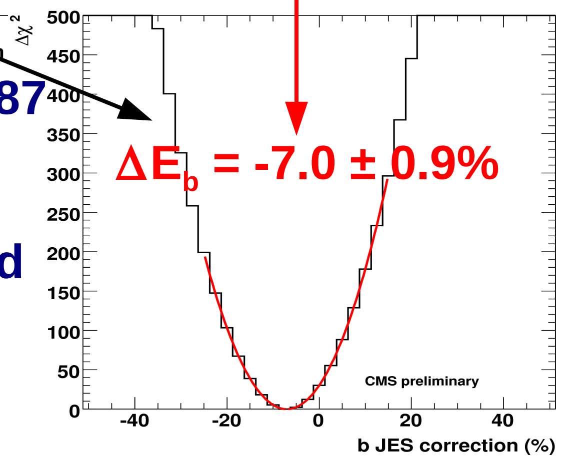


JEC factors with CMSSW_1_6_9

- For each event we have an estimate of the JES corrections, $\Delta E_{b,i}$ and $\Delta E_{l,i}$ (i=event)
- Events for which $\Delta E_{b,i}$ or $\Delta E_{l,i}$ > ±20% w.r.t first estimate are removed:
  - → The relative difference between the fitted expectation value of the m_W distribution and M_W^{world} is taken as a first estimate for light jets: $\Delta E_{l,incl.}$
  - → Difference between MC expectation values of light and b JES corrections (7%) is used to obtain the first estimate for b jets $\Delta E_{b,incl.}$ from $\Delta E_{l,incl.}$
- The $P^{fit}(\chi^2|\Delta E_b,\Delta E_l)$-values of the remaining events are translated into χ^2-values
  - The χ^2-values are combined and the minimum is searched for
  - Results are corrected for the width of pull distributions
  - The uncertainty reflects the uncertainty for 100 pb^{-1}



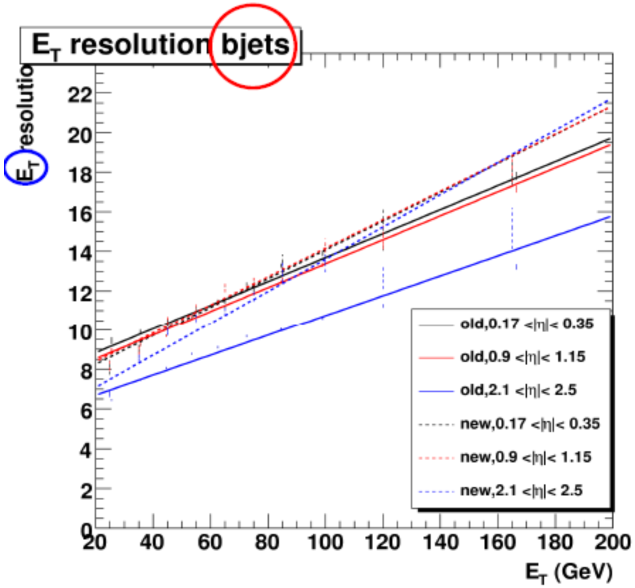


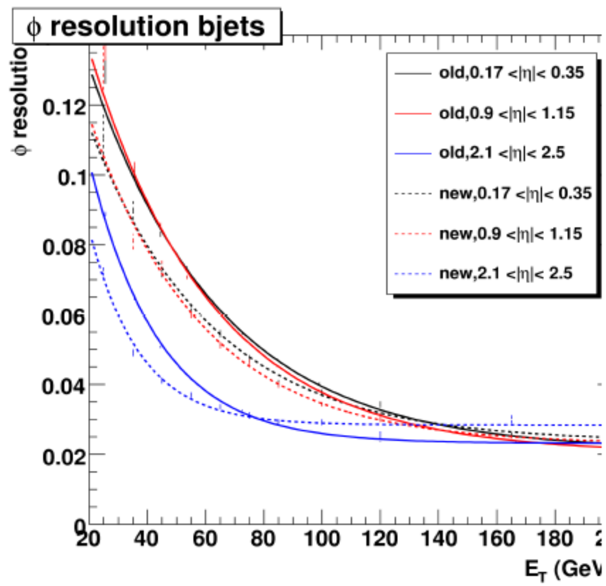


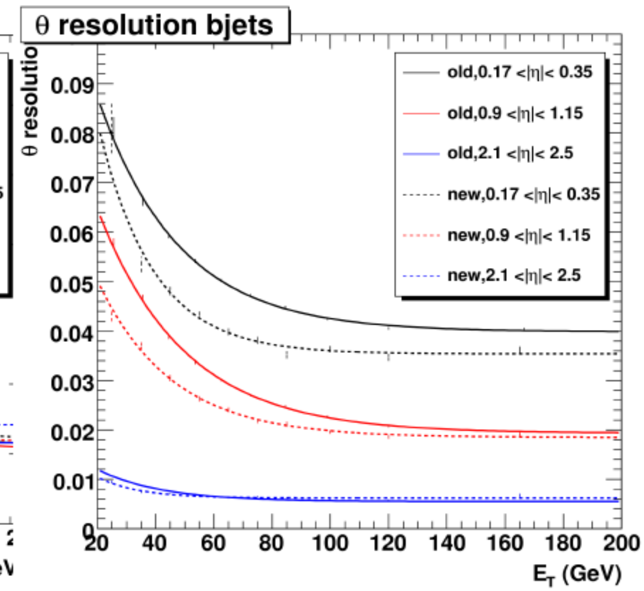
- Method is linear (slope of 0.77± 0.02 for light jets and 0.87 ± 0.03 for b jets → to avoid bias: corrections to be estimated should not deviate too much from 0)
- Method is robust against process and comb. background
- Method is also robust against smeared jet resolutions

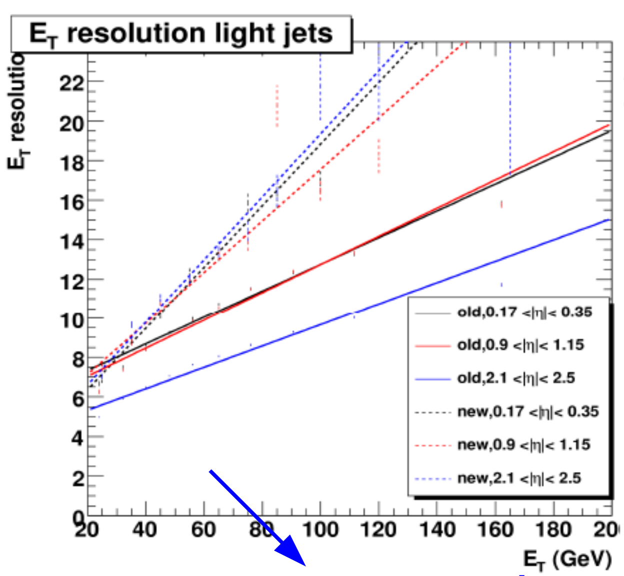
- Performance of method depends on Δm_t from Tevatron

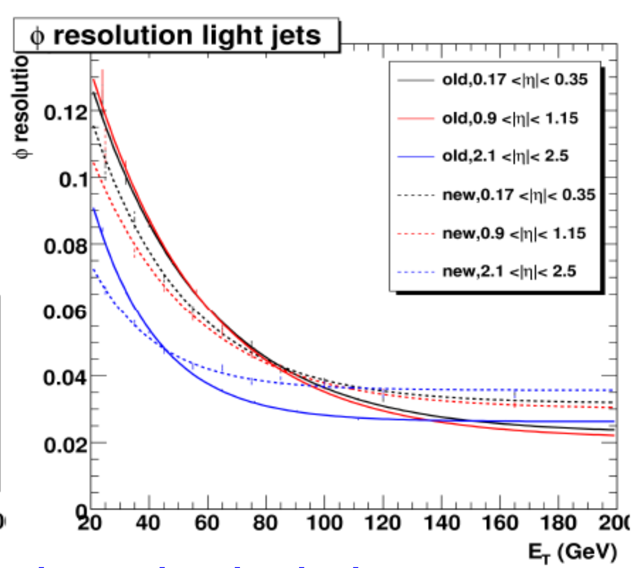
CMSSW_2_1_9: Resolutions

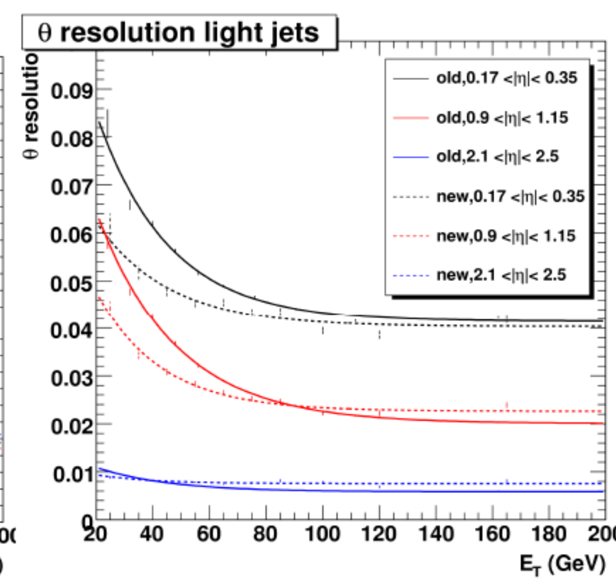












wrong quark energies! to be checked!!!

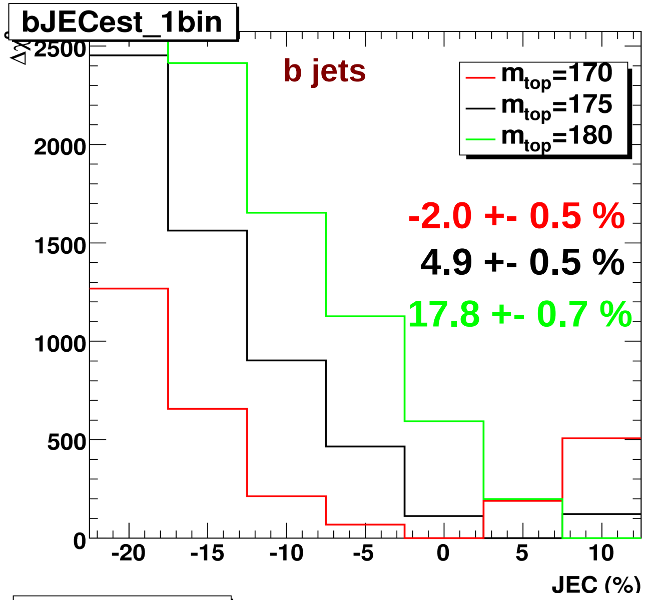
CMSSW_2_1_9: Analysis code

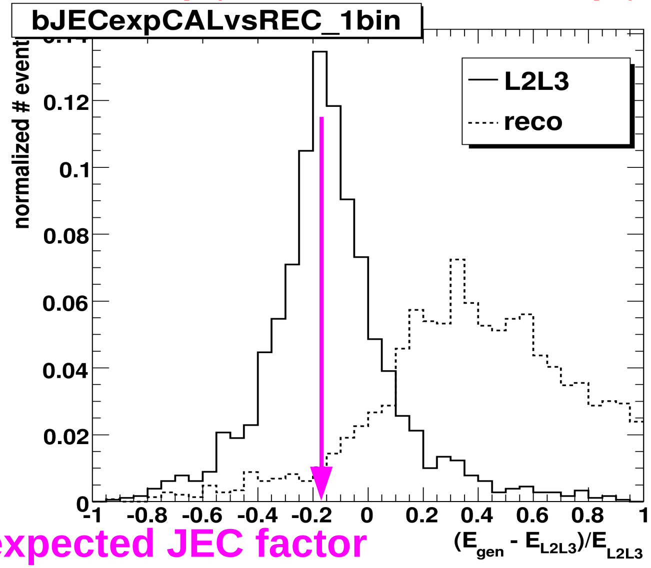
- In TQAF Layer 2, the event solutions (according to the possible jet-parton associations) are build using the EDProducer TtSemiEvtSolutionMaker
  - Every solution is a member of the class TtSemiEvtSolution
- On the existing structure of the Layer 2, I build a Layer 3 for my analysis (~60 hours for 5000 events with 25 JEC factors and 5 m_t):
  - A producer provides for every solution a vector containing all the χ^2-values for the different JEC factors and top masses
  - The provided "new" solutions are based on a class which inherits from TtSemiEvtSolution: all methods of TtSemiEvtSolution can be used + some extra methods to read out the vector containing the χ^2-values.
- On the Layer 3 the EDAnalyzer can be run to estimate the best JEC factors:
  - In the configuration file the η- and p_T-bins are specified
  - Among the 12 solutions, the best jet-parton association is picked (now the Monte-Carlo best jet-parton association is used: smallest sum of $\Delta R_{(jet,parton)}$)
  - The EDAnalyzer provides the estimated JEC factors for every (p_T,η)-bin and every top mass m_t
  - The EDAnalyzer provides also the plots with the $\Delta\chi^2$-values
- A separate EDAnalyzer is used to make plots of variables used in the event selection and ROOT macros are used to make the plots more fancy

CMSSW_2_1_9: Illustration (1 (p_T,η)-bin)

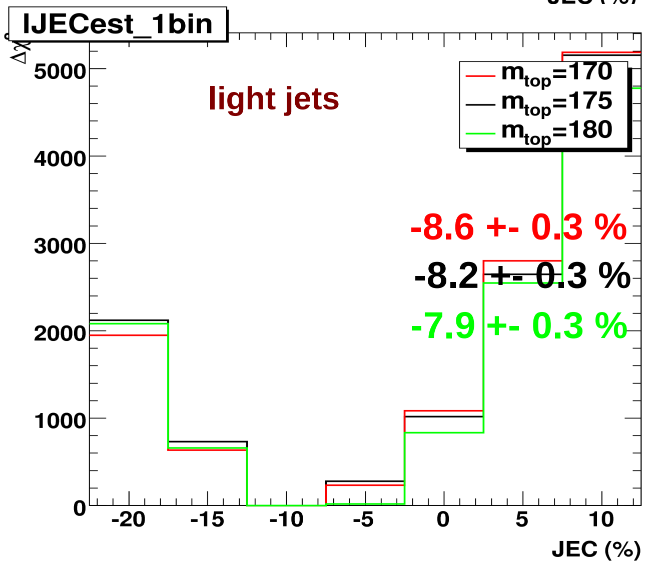
cuts: $E_{T(jets)}$>30 GeV/c; $p_{T(\mu)}$>30 GeV/c; $|\eta_{(jets,\mu)}|$<2.5

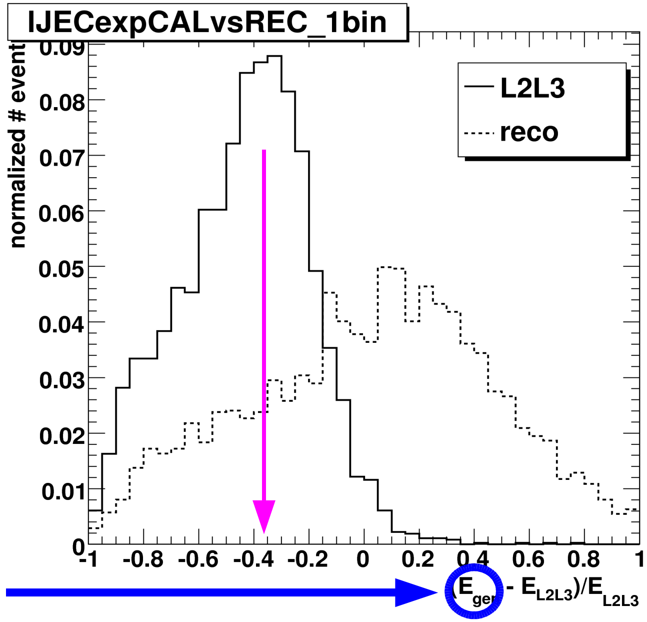
1 bin: 30<$p_{T(jets)}$<200 GeV/c and 0<$|h_{(jets)}|$<2.5





Fit with Gaussian:
expectation value=expected JEC factor



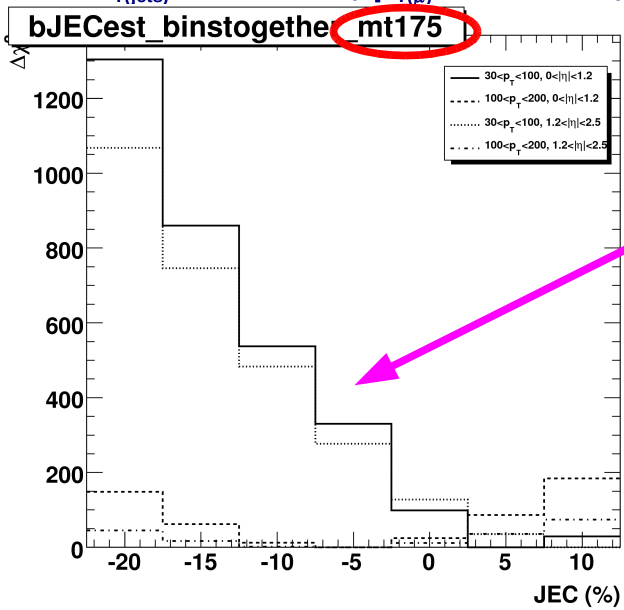


wrong quark energy
to be checked!!!

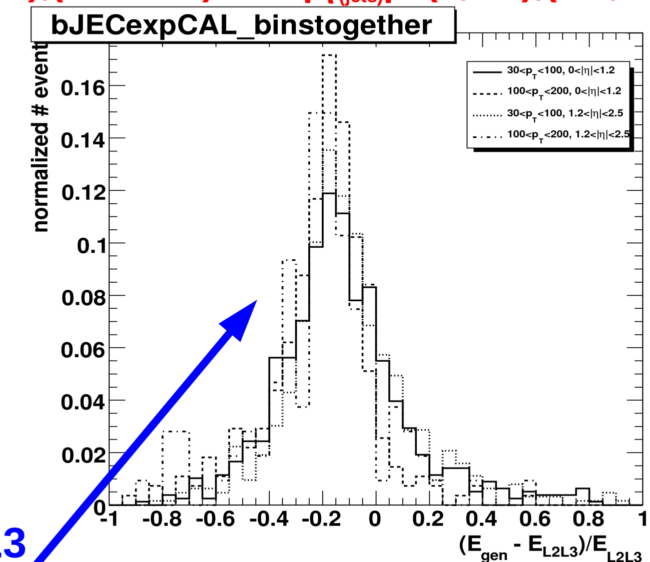
CMSSW_2_1_9: Illustration (4 (p_T,η)-bins)

cuts: $E_{T(jets)}>30$ GeV/c; $p_{T(\mu)}>30$ GeV/c; $|\eta_{(jets,\mu)}|<2.5$

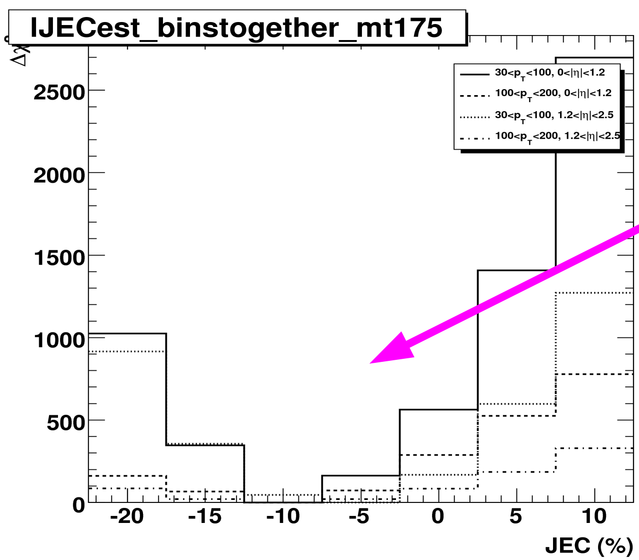
4 bins: $p_{T(jets)} \rightarrow (30\text{-}100),(100\text{-}200)$ and $|\eta_{(jets)}| \rightarrow (0,1.2),(1.2,2.5)$



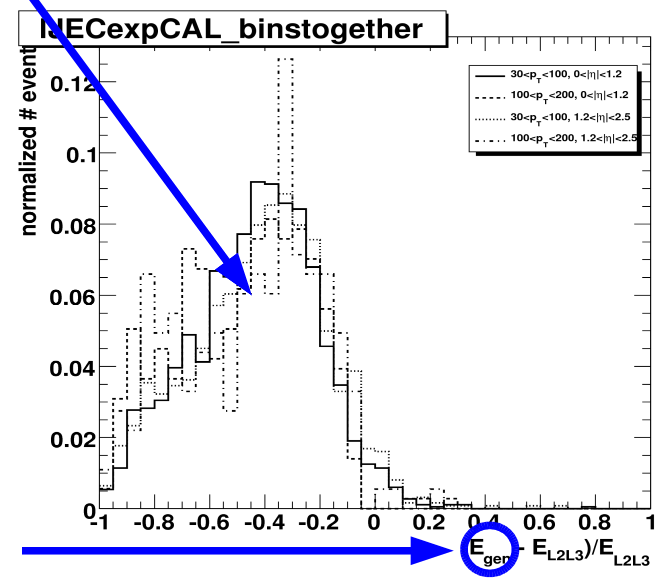
- for $p_T \in (30,100)$, $|\eta| \in (0,1.2)$: **6.9 +- 0.7 %**
- for $p_T \in (100,200)$, $|\eta| \in (0,1.2)$: **-5.8 +- 1.2%**
- for $p_T \in (30,100)$, $|\eta| \in (1.2,2.5)$: **10.6 +- 0.9%**
- for $p_T \in (100,200)$, $|\eta| \in (1.2,2.5)$: **-6.9 +- 1.9%**



Expected values are close to one other (L2L3 corrections are applied)



- for $p_T \in (30,100)$, $|\eta| \in (0,1.2)$: **-8.5 +- 0.4 %**
- for $p_T \in (100,200)$, $|\eta| \in (0,1.2)$: **-11.4 +- 0.7%**
- for $p_T \in (30,100)$, $|\eta| \in (1.2,2.5)$: **-6.3 +- 0.5%**
- for $p_T \in (100,200)$, $|\eta| \in (1.2,2.5)$: **-9.9 +- 1.1%**



wrong quark energy to be checked!!!

Outlook: towards CMSSW_2_2_X

- **With new crab-jobs (wider range, smaller stepsize, 5 different top masses)**
  - **Apply all selection cuts**
  - **Get results as function of p_T and η**
  - **Try to get a combined measurement of top quark mass and the JEC factors**

- **CMSSW_2_2_X:**
  - **Make new jet resolutions in CMSSW_2_2_X (from the produced Pat-tupple)**
  - **Develop method to give resolutions to kinematic fit (previous method obsolete)**
  - **Event hypothesis changes in CMSSW_2_2_X:**
    - **Need to change code (my class inherits from TtSemiEvtSolution class which is obsolete)**
  - **Use MVA discriminator tools to find the best jet-parton association**

- **Two main problems last weeks:**
  - **"matrix is singular" in kinematic fit: problem to invert a singular matrix.**
    - → **applied another method to invert the matrix solves the problem**
    - → **check consistency between 2 methods**
  - **Grid-CRAB problems: Bari server, queue too short, ... (should be solved)**

Smallest ΔR(jet,μ) after cut on RelIso

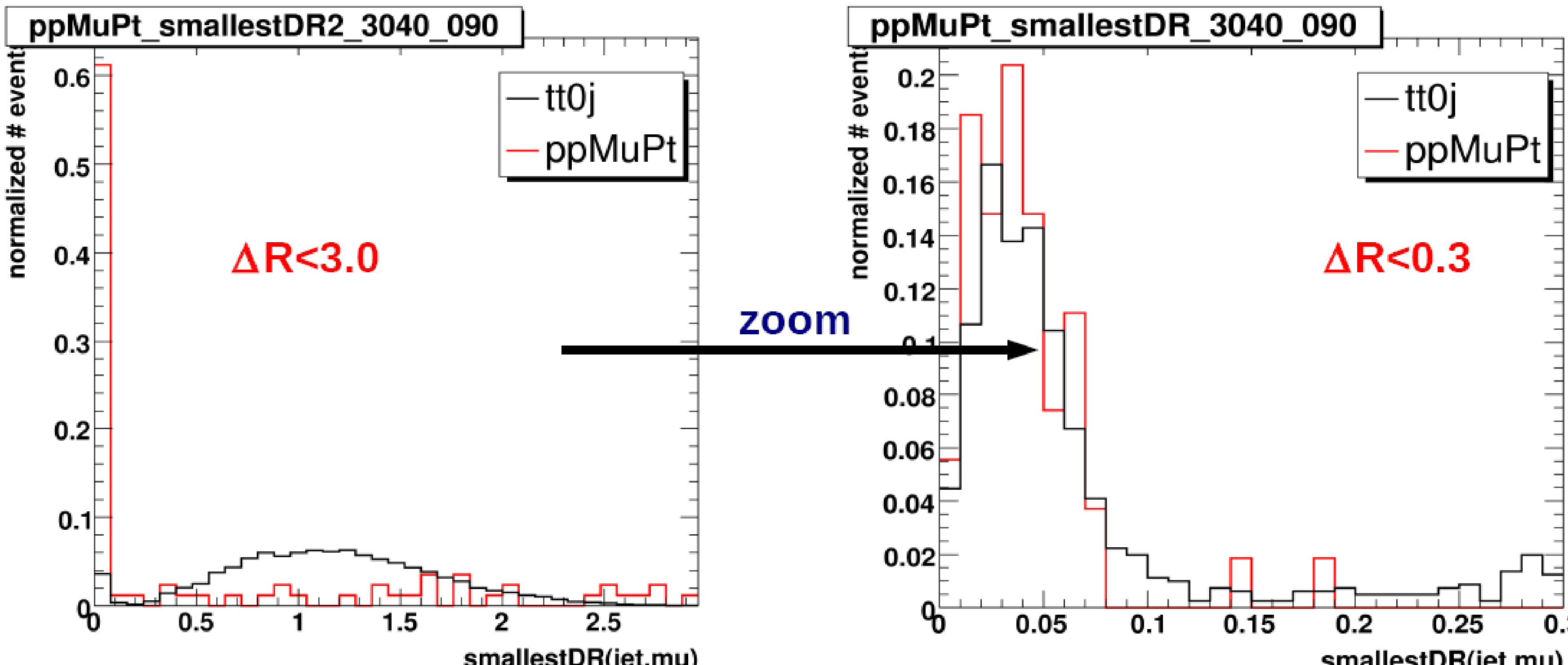
ttbar = all ttbar matched with genMuon (ΔR<0.1)

After applying the isolation cuts on the μ, there are still a lot of QCD events passing. Applying a cut on the smallest angle between jets and μ removes these events.

Plots:

- **RelIso > 0.90**
- **p_T^{μ} > 30 GeV/c, |η|<2.1**
- **E_T^{jets} > 40 GeV, |η|<2.4 (all jets!)**
- **Normalized**

Question: how can we remove these QCD events without removing ttbar?



Several variables were studied to check which kind of events are passing the isolation criterion, but have a muon is closer to the jet than 0.3.

- E_T closest jet → not interesting
- p_T muon → not interesting (higher values for ttbar compared to QCD)
- d0 muon → not interesting (slightly broader distribution for QCD)
- number of hits for muon → not interesting
- χ^2/ndf for muon → not interesting
- # constituents of closest jet → bigger values for ttbar, bigger values for $\Delta R > 0.3$
- emEt: energy deposited in ECAL in a cone of 0.3, with exclusion of towers crossed by the muon + all towers in a cone of 0.07 → not so interesting
- HadEt: energy deposited in HCAL in a cone of 0.3, with exclusion of towers crossed by the muon + all towers in a cone of 0.1 → not so interesting
- emEt+hadEt → difference between ttbar and QCD
- emEt/emEt+hadEt → not so interesting
- emS9: energy deposited in 3x3 ECAL crystal shape around the crossed crystal

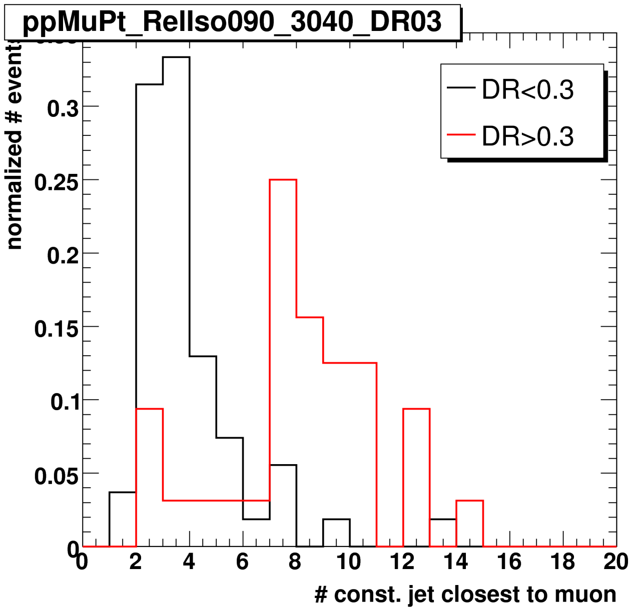
→ not so interesting

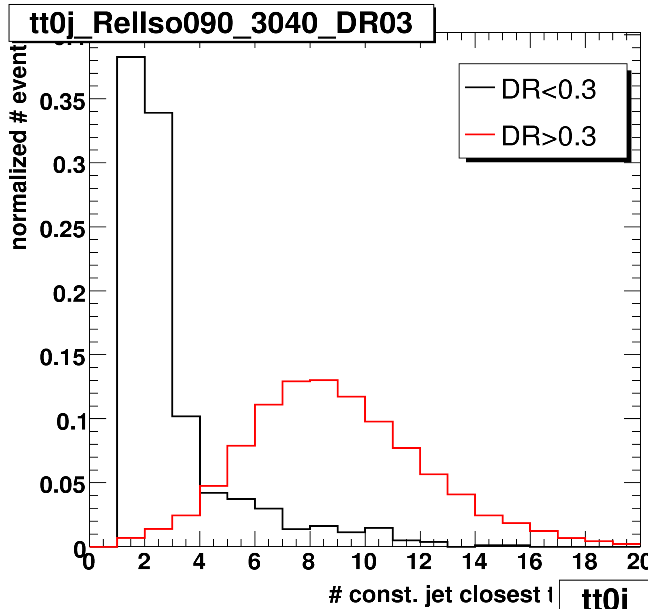
- hadS9: energy deposited in 3x3 HCAL tower shape around the crossed tower

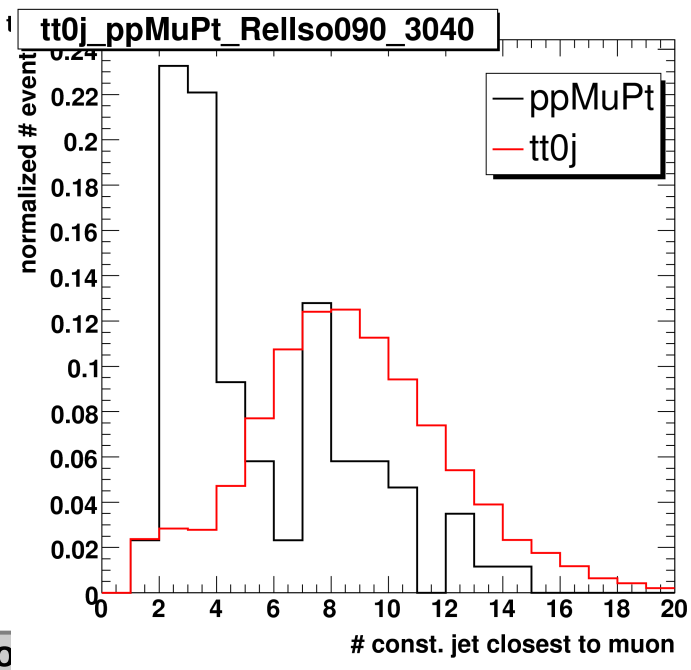
→ not so interesting

- emS9+hadS9 → not so interesting
- emS9/emS9+hadS9 → difference for $\Delta R <$ or > 0.3, difference QCD/ttbar
- MIP compatibility: are the muon deposits consistent with the deposits of a minimum ionizing particle? → difference for $\Delta R <$ or > 0.3, difference QCD/ttbar

constituents closest jet

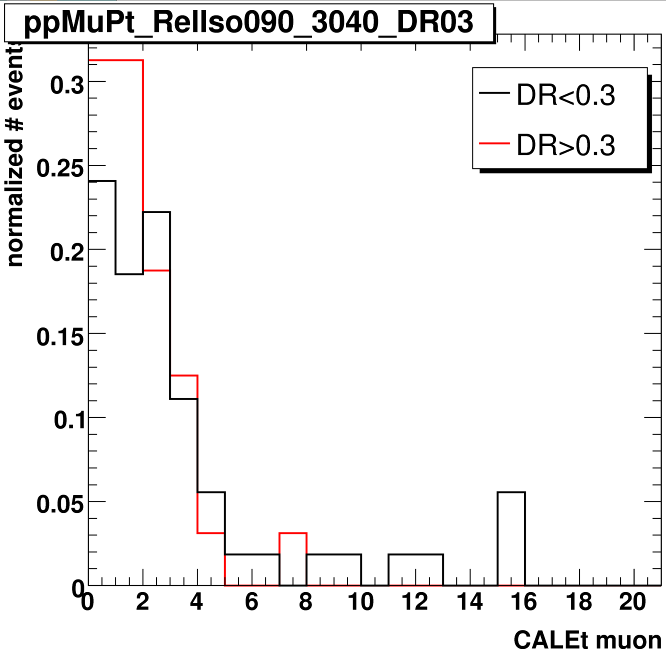


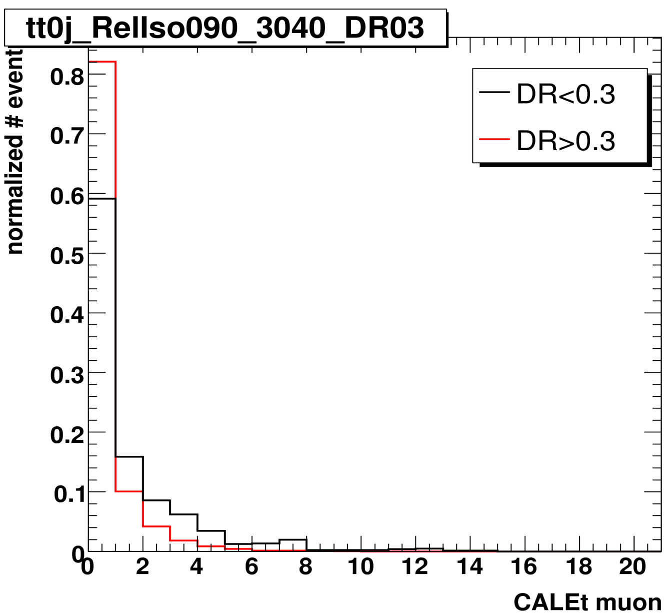




Clearly difference for ΔR< 0.3 and ΔR > 0.3
Also clear difference between ppMuPt and tt0j

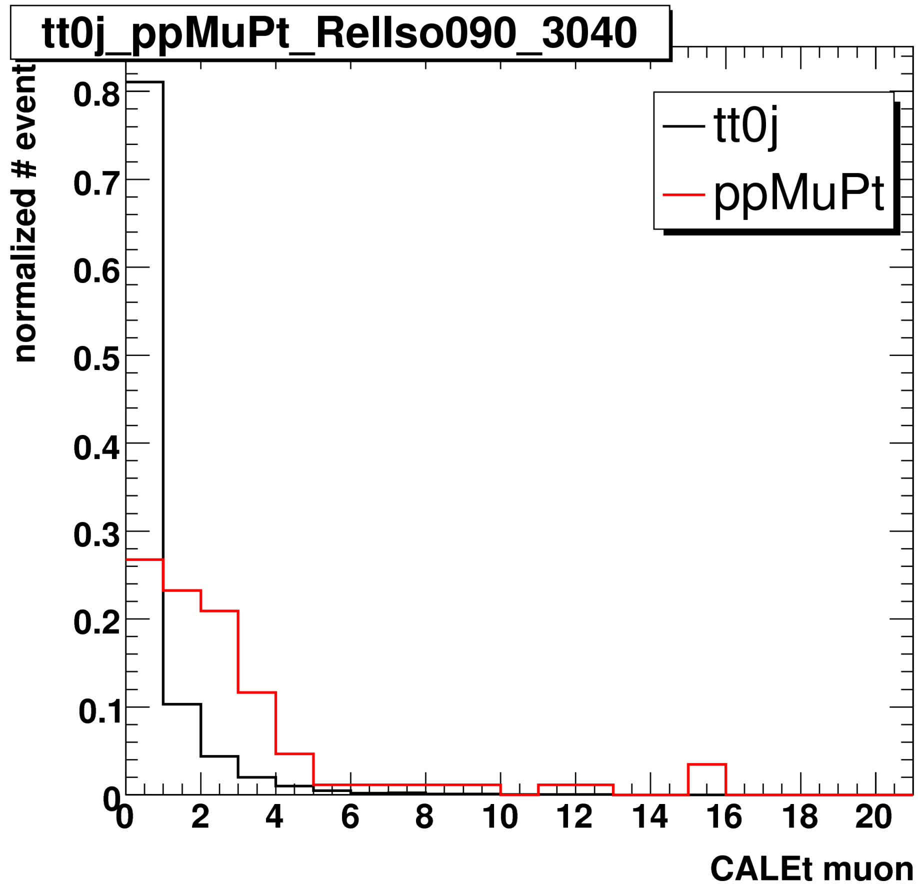
emEt+hadEt



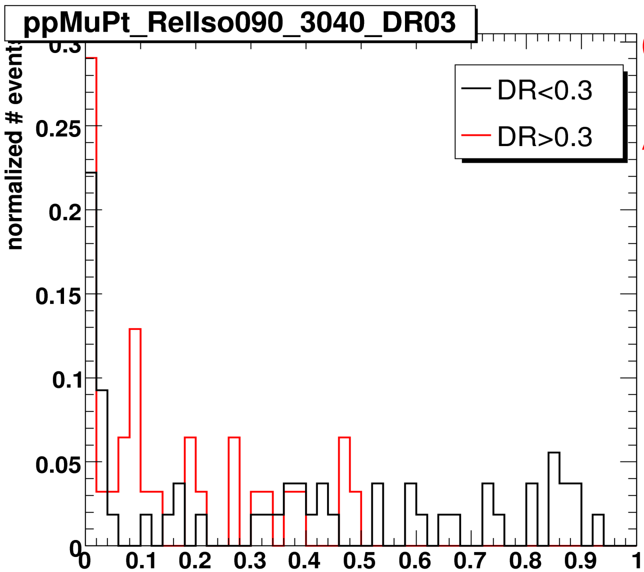


sum smaller for ΔR>0.3 and ΔR<0.3

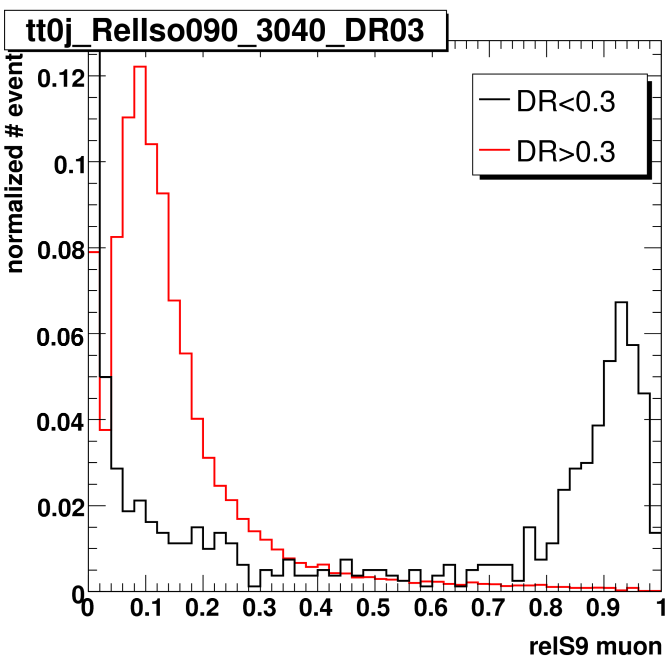
Difference between ppMuPt and tt0j

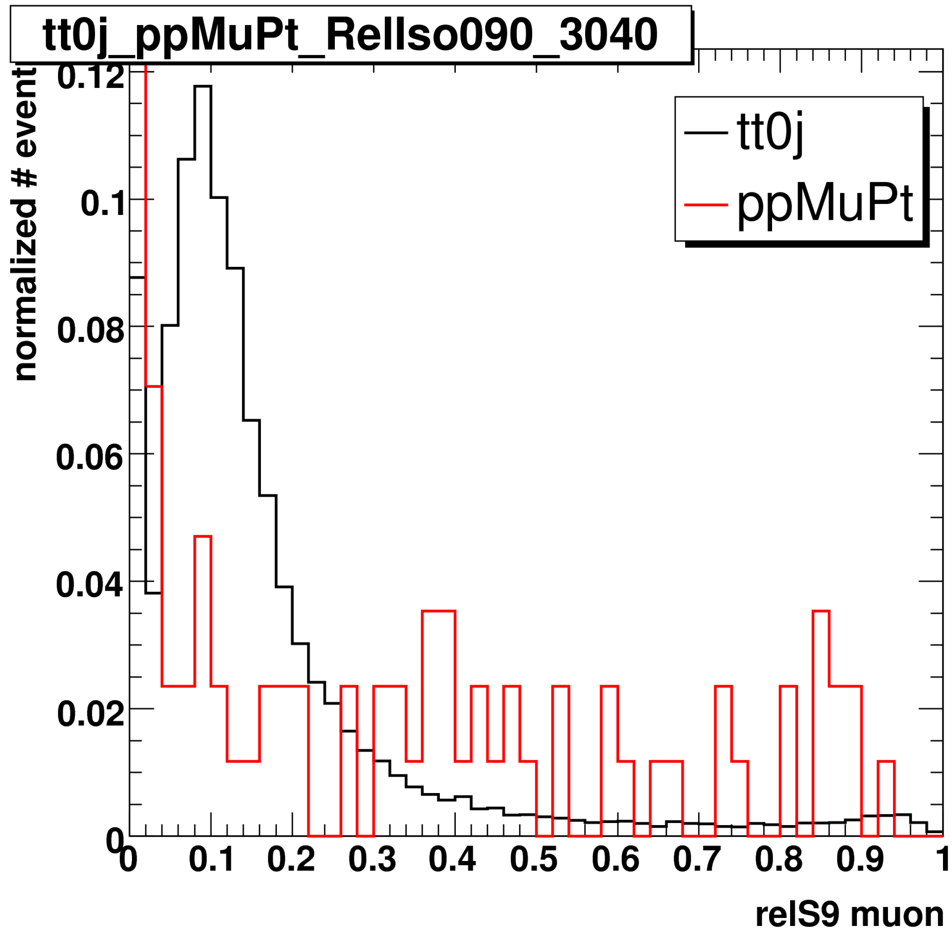


emS9/emS9+hadS9



Great separation between DR>0.3 and DR<0.3 for ttbar
less good for ppMuPt
Also difference between ppMuPt and tt0j

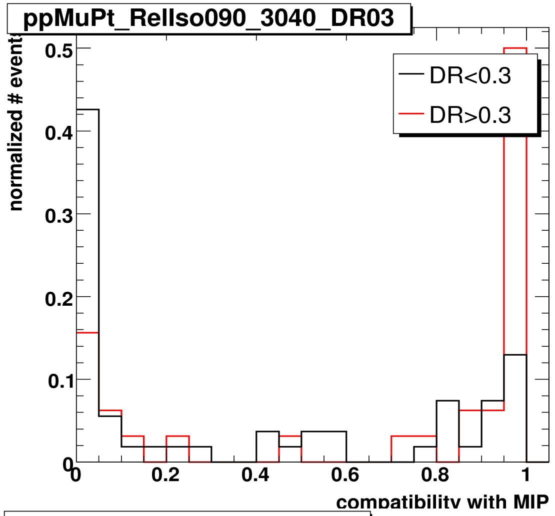


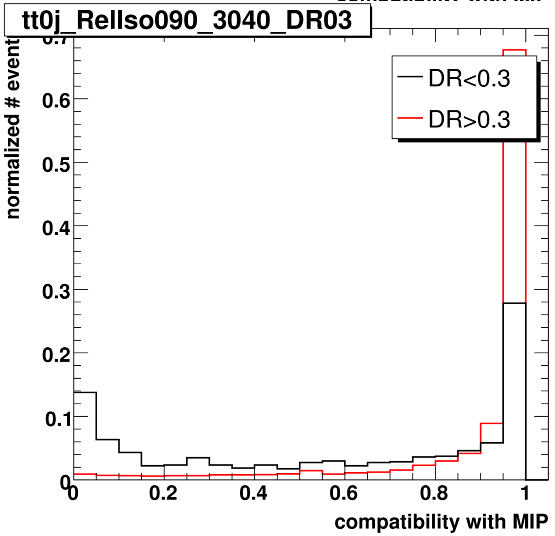


Note: X-axis was bigger, so normalization not really visible on these plots

MIP compatibility

MIP compatibility: are the muon deposits consistent with deposits from a minimum ionizing particle?





Great separation between DR>0.3 and DR<0.3
Also difference between ppMuPt and tt0j