CMS Physics Analysis Summary

Contact: cms-pog-conveners-pflowtau@cern.ch 2010/07/29

Study of tau reconstruction algorithms using $pp$ collisions
data collected at $\sqrt{s} = 7$ TeV

The CMS Collaboration

Abstract

Proton–proton collision events collected with the CMS experiment at LHC at a center–
of–mass energy of $\sqrt{s} = 7$ TeV in 2010 are used to commission the algorithms for
reconstruction and identification of tau lepton hadronic decays. Four different types
of algorithms are considered: three based on particle–flow event reconstruction and
one based on combinations of tracks and calorimeter clusters. Probabilities for quark
and gluon jets to pass the tau identification criteria of the different algorithms are
measured in data dominated by QCD multi–jet events and compared to predictions
of Monte Carlo simulations.
1 Introduction

Tau leptons are expected to play a significant role in discovering new physics phenomena at LHC [1]. Many physics analyses are expected to benefit from an efficient reconstruction of tau leptons which should be complemented by a good performance in rejecting possible background contaminations.

In about two thirds of cases tau leptons decay hadronically, typically into either one or three charged mesons (predominantly $\pi^+$, $\pi^-$) in presence of up to two neutral pions, decaying via $\pi^0 \rightarrow \gamma\gamma$. Similar signature is expected for generic, quark and gluon, QCD jets production. Since the cross-section of jet production exceeds the cross-section of tau lepton production by several orders of magnitude, the experimental challenge in reconstructing and identifying hadronic tau decays is to discriminate efficiently between genuine tau lepton hadronic decays (tau jets) and quark/gluon jets misreconstructed as tau candidates.

The $pp$ collision data collected with the CMS experiment [2] so far are expected to contain very few tau leptons. A dataset corresponding to an integrated luminosity of 70 nb$^{-1}$, collected in $pp$ collisions at a center-of-mass energy of $\sqrt{s} = 7$ TeV in 2010 can be used to test the level of agreement between data and Monte Carlo (MC) simulation for events containing tau jet candidates. As an example, $Z \rightarrow \tau^+\tau^-$ candidates in the muon plus tau jet final state are selected by requiring the presence of a global muon [3] and a tau jet reconstructed as described in Section 2. No isolation is applied, for both the muon and tau jet candidate, to enhance the contributions of $W^\pm \rightarrow \mu^\pm\nu$ and QCD multi-jet background processes. Distributions of the transverse mass $M_T(\mu, MET)$ and of the visible mass of muon plus tau jet for events selected in the analyzed dataset are compared to MC predictions in Fig. 1. The visible mass distributions for events containing muon and tau jet candidates of like-sign (LS) and opposite-sign (OS) charge match within statistical uncertainties, in agreement with the expectation for a background dominated event sample. When tight isolation cuts are applied on muon and tau jet candidate, a single event remains in data, with expected $Z \rightarrow \tau^+\tau^-$ purity of 75%. A display of this event is shown in the appendix.

Figure 1: Distributions of the transverse mass of muon plus MET (a) and of the muon plus tau jet visible mass (b). Comparison of the muon plus tau jet visible mass distributions for events containing a muon plus tau jet candidate pair of like-sign (LS) versus opposite-sign (OS) charge (c).

The number of tau leptons present in the $pp$ collision data collected so far is not sufficient for a measurement of tau identification efficiencies. Instead, the focus of the results presented in the following will be to measure the probabilities (“fake–rates”) with which generic QCD jets pass the selection criteria of four tau identification algorithms. The measurement of fake–rates
2 Results

The fake–rates of the TCTau, shrinking cone, HPS and TaNC tau identification algorithm are measured as function of jet $p_T$ and $\eta$ by counting the fraction of quark and gluon jets passing the identification criteria of one particular algorithm in a given $p_T^\text{jet}$ or $\eta^\text{jet}$ bin:

$$P_{fr}(\text{bin}) := \frac{N_{\text{jets}}(\text{bin})}{N_{\text{jets}}(\text{bin})},$$

(1)
where \( N_{\text{jets}} \) denotes the number of jets reconstructed in that bin. Jets considered in the fake-rate measurement are required to have transverse momenta \( P_T^{\text{jet}} > 10 \ \text{GeV}/c \) and to be within the geometrical acceptance of the tracking detectors, \(|\eta| < 2.5\).

The fake-rates measured for the four tau identification algorithms are presented in Fig. 2. Also shown in the figure are the efficiencies with which the algorithms identify genuine hadronic tau decays, estimated using a sample of simulated \( Z \to \tau^+ \tau^- \) events. The efficiency is determined for tau lepton hadronic decays into tau jets of visible transverse momentum \( P_T^{\text{vis}} > 10 \ \text{GeV}/c \) and pseudo-rapidity \(|\eta| < 2.5\), selected on generator level.

![Figure 2](image-url)

Figure 2: Left: Measured probabilities of quark/gluon jets to pass the tau candidate selection criteria of the TCTau (black stars), shrinking cone (red squares), HPS (brown downward facing triangles) and TaNC (blue upward facing triangles) tau identification algorithms as function of jet \( P_T \) (a) and \( \eta \) (c). Right: Efficiencies of the algorithms to identify genuine tau lepton hadronic decays as function of transverse momentum (b) and pseudo-rapidity (d) of the visible tau decay products on generator level, estimated using a sample of simulated \( Z \to \tau^+ \tau^- \) events.

The simple cone based algorithms have very similar fake-rates and efficiencies. Averaged over

\[ \text{ CMS Preliminary } \\
\text{ Data, } L = 8.4nb \]
the $p_T$ and $\eta$ distribution of jets in the analyzed data sample, the fake–rate of the TCTau and shrinking cone algorithms amounts to about 2%, for an estimated efficiency of about 45% to identify hadronic tau decays in $Z \to \tau^+\tau^-$ events.

More complex algorithms analyzing the constituents of tau jet candidates in terms of individual tau lepton hadronic decay modes provide a significantly lower fake–rate. The exact fake–rate and efficiency values of the HPS and TaNC algorithms depend on the chosen working–point. Medium tight selection criteria are used for comparing the algorithms. The “medium isolation” requirement on the $p_T$ of jet constituents not associated with tau decay products by the HPS algorithm yields a fake–rate of on average 0.2%, for a tau identification efficiency of about 27%, while a medium tight cut on the neural network output of the TaNC algorithm, denoted as “0.5%” working–point, yields an average fake–rate of 0.3% for a tau identification efficiency of about 36%.

The fake–rates as well as the efficiencies exhibit a significant dependency on $p_T$. The $p_T$ dependence varies significantly between different algorithms. A common feature of all algorithms is that fake–rate as well as efficiencies rise steeply at low $p_T$. For higher $P_{vis}$ the efficiency then reaches a plateau and remains approximately constant, whereas the fake–rate decreases with jet $p_T$, slowly for some algorithms and more rapidly for others. The dependency of fake–rates and efficiencies on $\eta$ is significantly less pronounced than the dependency on $p_T$.

The measured fake–rates are compared to MC predictions in Figs. 3 and 4. The predictions of the MC simulation are seen to underestimate the fake–rate values measured for different jet $p_T$ and $\eta$ by a small amount. Preliminary studies indicate that the difference in fake–rate observed between data and simulation may be to the modeling of hadronization processes in the Monte Carlo generator. Investigations are ongoing to what extent different PYTHIA tunes [7, 8] have an effect on the fake–rates predicted by the MC simulation. Considering the fact that the tau identification algorithms select collimated jets of low particle multiplicity, atypical for the bulk of quark/gluon jets, the agreement between data and MC simulation, on the level observed, is remarkable.

4 Conclusion

The fake–rate of four tau reconstruction algorithms, measured using a data sample corresponding to an integrated luminosity of 8.4 nb$^{-1}$ collected with the CMS experiment at a center–of–mass energy of $\sqrt{s} = 7$ TeV in 2010, was analyzed as function of jet transverse momentum and pseudo–rapidity.

Averaged over the transverse momentum and pseudo–rapidity distribution of jets in the analyzed dataset, the probability of generic QCD jets to pass the tau candidate selection criteria of the simple cone based TCTau and shrinking cone algorithms amount to about 2%, for an average estimated efficiency to reconstruct and identify genuine hadronic tau decays in simulated $Z \to \tau^+\tau^-$ events of about 45%. The more complex TaNC and HPS algorithms, which analyze the constituents of tau jet candidates in terms of individual tau lepton hadronic decay modes, provide a reduction in fake–rate by a factor of 5 to 10, for less than a factor 2 loss in tau identification efficiency.

The demonstrated performance of the algorithms in terms of tau identification efficiency versus fake–rate satisfy the requirements of CMS analyses of processes which include tau lepton hadronic decays in the final state.
Figure 3: Probabilities for quark/gluon jets to pass the tau candidate selection criteria of the TCTau (a), shrinking cone (b), HPS (c) and TaNC (d) tau identification algorithms as function of jet $p_T$. Fake–rates measured in data are represented by solid symbols and compared to Monte Carlo predictions, represented by open symbols. In case of the HPS and TaNC algorithms, the fake–rates measured for loose, medium and tight “working–point” selection criteria are represented by markers of different colors.
Figure 4: Probabilities for quark/gluon jets to pass the tau candidate selection criteria of the TCTau (a), shrinking cone (b), HPS (c) and TaNC (d) tau identification algorithms as function of jet $\eta$. Fake–rates measured in data are represented by solid symbols and compared to Monte Carlo predictions, represented by open symbols. In case of the HPS and TaNC algorithms, the fake–rates measured for loose, medium and tight “working–point” selection criteria are represented by markers of different colors.
5 Appendix A. Event display

Figure 5: Display of the $Z \rightarrow \tau^+ \tau^-$ candidate event passing all selection criteria of the muon plus tau jet final state observed in a dataset corresponding to an integrated luminosity of 70 nb$^{-1}$, collected in $pp$ collisions at a center–of–mass energy of $\sqrt{s} = 7$ TeV in 2010. Two vertices are reconstructed in this event. The tracks of muon and tau jet candidate are compatible with originating from the same vertex. The second vertex is interpreted as belonging to a pile–up event, produced by another $pp$ interaction in the same bunch–crossing.

References


