2010/04/01

Estimation of Background contributions to Tau analyses via Fake–Rate technique

J. Conway, E. K. Friis, and C. Veelken
Department of Physics
University of California
One Shields Avenue
Davis, CA 95616–8677

Abstract

The fake–rate technique provides a data–driven way of estimating background contributions in physics analyses involving tau leptons in the final state. The technique is applicable in any analysis in which dominant background contributions are expected to arise from the misidentification of quark or gluon jets as hadronic tau decays. In this note, we provide a detailed description of the method and demonstrate its applicability to estimate background contributions using the analysis of $Z \rightarrow \tau^+\tau^- \rightarrow \mu + \tau$–jet by way of example.

1 Introduction

Many analyses of interest for discovering new physics phenomena at the LHC are based on the capability of the detectors to reconstruct and identify tau leptons. In about two thirds of the cases tau leptons decay hadronically, typically into either one or three charged mesons (predominantly π^+, π^-) plus zero to two neutral pions. The experimental challenge in reconstructing and identifying hadronic tau decays is to discriminate efficiently between tau–jets on the one hand and quark/gluon jets on the other hand. Details of the algorithm used to identify and reconstruct tau lepton hadronic decays can be found in [2].

Since quark and gluon jets are produced with cross–sections which are in general some orders of magnitude larger than the cross–sections with which tau leptons are produced, significant background contributions to physics analyses involving tau leptons in the final state may arise from the misidentification of quark and gluon jets as tau–jets, even if the probabilities of such misidentification (“fake–rate”) per jet is on the 10^{-2} to 10^{-3} level [2, 3].

In this note, we describe how knowledge of the probabilities with which quark and gluon jets get misidentified as tau–jets may be utilized to obtain an estimate of background contributions in physics analyses. As an illustrative example and in order to demonstrate the precision achievable with the method, we present results of applying the fake–rate technique to estimate the contributions of QCD, W+jets, $t\bar{t}$+jets and $Z \to \mu^+\mu^-$ backgrounds in the measurement of the $Z \to \tau^+\tau^-$ cross–section, in the channel $Z \to \tau^+\tau^- \to \mu + \tau$-jet. Details of the analysis can be found in reference [4].

The results described in this note were obtained from Monte Carlo simulations of the $Z \to \tau^+\tau^-$ signal and different background processes for a centre–of–mass energy of $\sqrt{s} = 7$ TeV. Analysis of the $\sqrt{s} = 900$ GeV data recorded in 2009 [1] indicate that the probabilities for quark and gluon jets to fake the signatures of hadronic tau decays are well modeled by the Monte Carlo simulation. Once data–samples of sufficient event statistics are available at collision energies of $\sqrt{s} = 7$ TeV, fake–rates at the higher centre–of–mass energy will be measured in data and the values obtained from data will henceforth be used for the purpose of estimating background contributions via the fake–rate technique.

2 Parametrization of fake–rates

Efficiencies and fake–rates of the tau identification algorithm based on requiring no tracks of $P_T > 1$ GeV and ECAL energy deposits of $P_T > 1.5$ GeV reconstructed within an “isolation cone” of size $dR_{iso} = 0.5$ and outside of a “shrinking signal cone” of size $dR_{sig} = 5.0/E_T$ as it is used in the $Z \to \tau^+\tau^- \to \mu + \tau$-jet analysis are displayed in figure 1. In order to account for the visible P_T and η dependence, we parametrize the fake–rates in bins of transverse momentum and pseudo–rapidity. As we will show in section 3, the parametrization of the fake–rates by P_T and η makes it possible to not only estimate the total number of background events contributing to physics analyses, but to model the distributions of kinematic observables with a precision that is sufficient to extract information on the background shape.

We add a third quantity, the E_T-weighted jet–width R_{jet}, to the parametrization in order to account for differences between the fake–rates of quark and gluon jets. The jet width is defined as

$$R_{jet} = \sqrt{E\left(\eta^2\right) + E\left(\phi^2\right)} \tag{1}$$

where $E\left(\eta^2\right)$ $\left(E\left(\phi^2\right)\right)$ is the second η (ϕ) moment of the jet constituents, weighted by constituent transverse energy. Analyses performed by the CDF collaboration [5] found that sys-

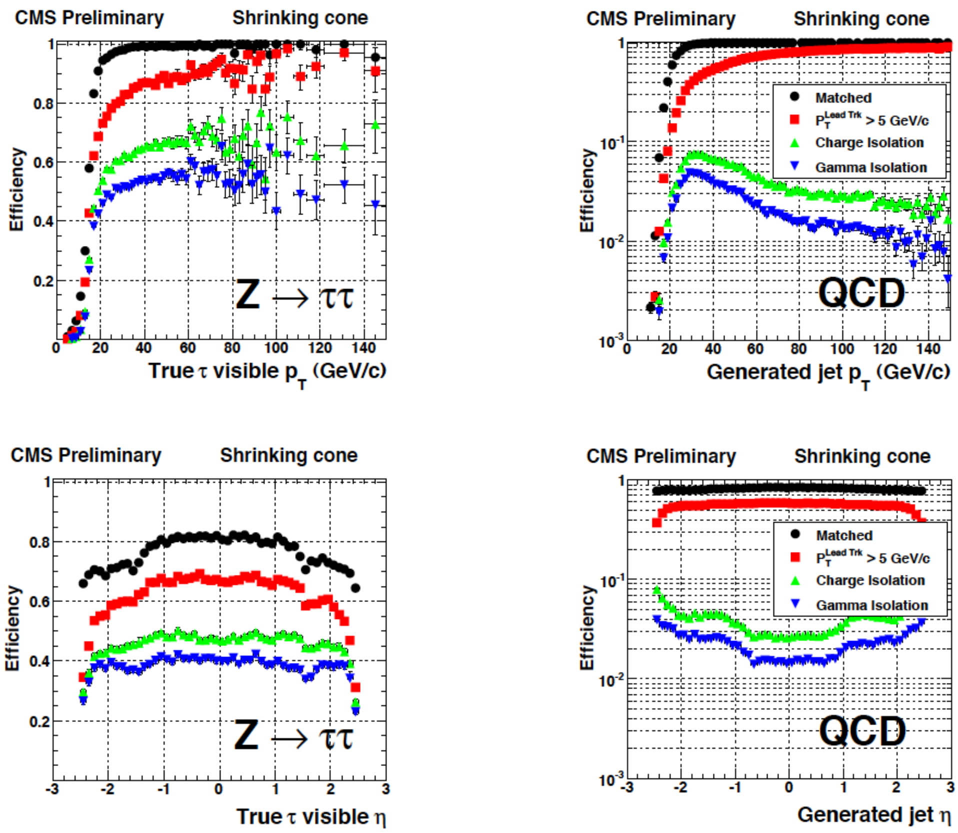

Figure 1: Cumulative efficiencies (left) and fake–rates (right) of successively applied tau identification cuts of the "shrinking signal cone" particle–flow based tau identification algorithm described in [2] as function of P_T^{jet} (top) and η_{jet} (bottom) of tau–jet candidates. The efficiencies/fake–rates for the complete set of tau identification criteria are represented by the blue (downwards facing) triangles.

tematic uncertainties on background estimates obtained from the fake–rate method are reduced in case differences between quark and gluon jets get accounted for in this way.

Efficiencies and fake–rates are then obtained by counting the fraction of tau–jet candidates passing all tau identification cuts and discriminators in a given bin of P_T^{jet}, η_{jet} and R_{jet}:

$$P_{fr}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right) := \frac{N_{jets}\left(P_T^{jet}, \eta_{jet}, R_{jet} | \text{all tau id. cuts and discriminators passed}\right)}{N_{jets}\left(P_T^{jet}, \eta_{jet}, R_{jet} | \text{preselection passed}\right)} \tag{2}$$

The preselection in the denominator of equation 2 in general refers to P_T and η cuts, which are applied with thresholds matching those applied on the final analysis level, but may as well include loose tau identification criteria (which may be applied e.g. already during event skimming).

Different sets of fake–rates are determined for the highest P_T and for the second highest P_T jet in QCD di–jet events, for jets in a QCD event sample enriched by the contribution of heavy quarks and gluons by requiring the presence of a muon reconstructed in the final state, and for jets in "electroweak" events selected by requiring a W boson in the final state.

Tau identification efficiencies are obtained from a $Z \to \tau^+\tau^-$ Monte Carlo sample.

3 Application to background estimation

Knowledge of the tau identification efficiencies and fake–rates as function of the parameters P_T^{jet}, η_{jet} and R_{jet} as defined by equation 2 is utilized to obtain an estimate for the contributions of background processes to physics analyses involving tau lepton hadronic decays in the final state. The basic idea is to replace tau identification cuts and discriminators by appropriately chosen weights.

Application of the fake–rate technique consists of two stages. The first stage consists of loosening the tau identification cuts and discriminators and applying only the preselection requirements defined by the denominator of equation 2, in order to obtain an event sample dominated by contributions of background processes, which are expected to increase by the inverse of the (average) fake–rate, typically by a factor $\mathcal{O}(100)$. In the second stage, weights are applied to all events in the background dominated control sample, according to the probabilities $P_{fr}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right)$ for jets to fake the signature of a hadronic tau decay. After application of the weights, an estimate for the total number of background events passing the tau identification cuts and discriminators and thus contributing to the final analysis sample is obtained.

The fake–rate technique works best if all background contributions to the analysis arise from misidentification of quark and gluon jets as hadronic tau decays. Corrections to the estimate obtained from the fake–rate technique are needed in case of background processes contributing to the final analysis sample which either produce genuine tau leptons in the final state (e.g. $t\bar{t}+$ jets) or in which tau–jet candidates are due to misidentified electrons or muons (e.g. $Z \to \mu^+\mu^-$, $Z \to e^+e^-$), as the latter may fake signatures of hadronic tau decays with very different probabilities than quark and gluon jets.

In the "simple" fake–rate method described in more detail in the next section, the corrections are taken from Monte Carlo simulations. Corrections based on Monte Carlo are needed also to compensate for signal contributions to the background dominated control sample.

An alternative to Monte Carlo based corrections is to utilize additional information contained in the background dominated control sample. The modified version is described in section 3.2. It has been used to estimate background contributions in searches for Higgs boson production with subsequent decays into tau lepton pairs performed by the CDF collaboration in TeVatron run II data [5]. We will refer to the modified version as "CDF–type" method in the following.

3.1 "Simple" weight method

In the "simple" method all tau–jet candidates within the background dominated event sample are weighted by the probabilities of quark and gluon jets to fake the signature of a hadronic tau decay:

$$w_{jet}^{simple}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right) := P_{fr}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right) \tag{3}$$

These weights are applied to all jets in the background dominated control sample which pass the preselection defined by the denominator of equation 2. Note that the weights defined by equation 3 can be used to estimate the contributions of background processes to distributions of tau–jet related observables. They cannot be used as event weights.

In order to compare distributions of event level quantities or per–particle quantities for particles of types different from tau leptons decaying hadronically, event weights need to be defined. Neglecting the small fraction of background events in which multiple tau–jet candidates pass the complete set of all tau identification cuts and discriminators, event weights can be computed by summing up the per–jet weights defined by equation 3 over all tau–jet candidates in

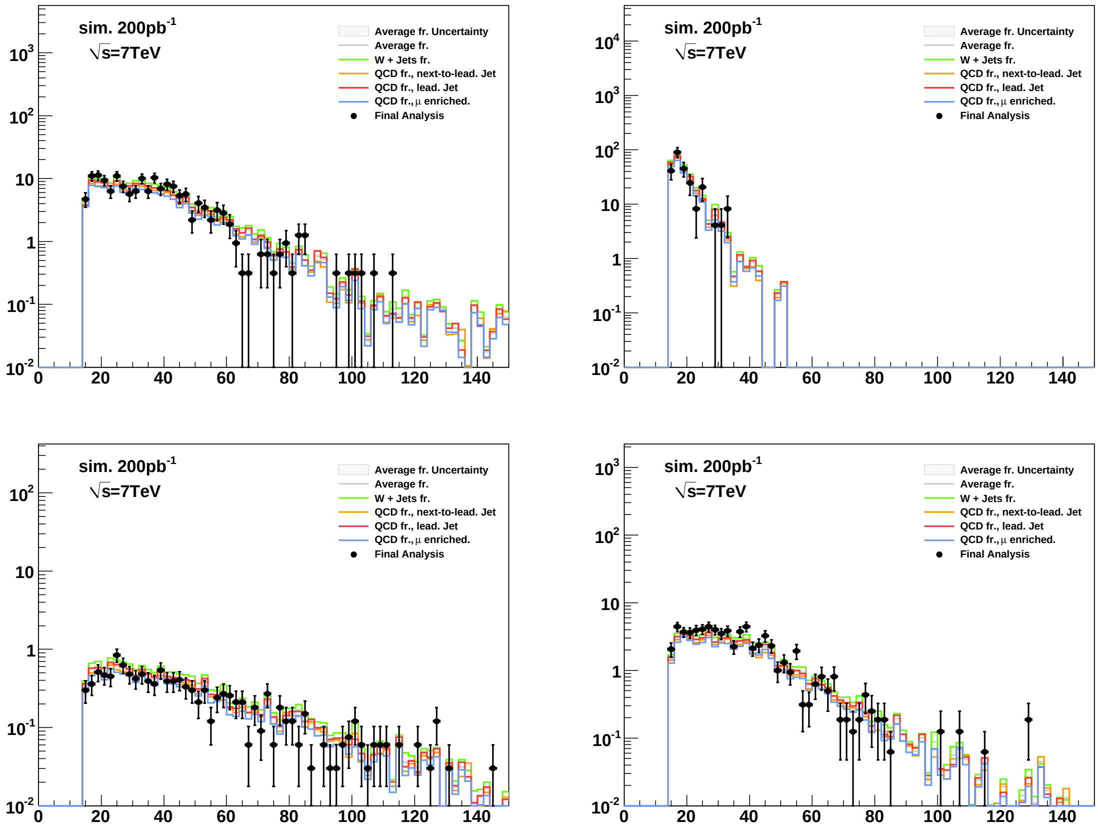

Figure 2: Distributions of muon transverse momentum in W+jets (top left), QCD (top right), $t\bar{t}$+jets (bottom left) and $Z \to \mu^+\mu^-$ (bottom right) background events which pass all selection criteria of the $Z \to \tau^+\tau^- \to \mu + \tau$-jet cross–section analysis compared to the estimate obtained from the fake–rate technique, computed according to equation 4. The expected contribution of background processes is indicated by points. Lines of different colors represent the estimates obtained by applying fake–rate weights determined for different compositions of light quark, heavy quark and gluon jets, as described in section 2. The maximum (minimum) estimate is interpreted as upper (lower) bound. The difference between the bounds is taken as systematic uncertainty on the estimate obtained from the "simple" fake–rate method and is represented by the gray shaded area.

the event which pass the preselection:

$$W_{event}^{simple} := \Sigma w_{jet}^{simple} \tag{4}$$

A bit of care is needed in case one wants to compare distributions of observables related to "composite particles" the multiplicity of which depends on the multiplicity of tau–jet candidates in the event (e.g. combinations of muon + tau–jet pairs in case of the $Z \to \tau^+\tau^- \to \mu + \tau$–jet analysis). Per–particle weights need to be computed for such "composite particles", depending on P_T^{jet}, η_{jet}, R_{jet} of its tau–jet candidate constituent, according to:

$$w_{comp-part}^{simple}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right) := w_{jet}^{simple}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right) \tag{5}$$

Results of applying the "simple" fake–rate weights defined by equations 3, 4 and 5 to the $Z \to \tau^+\tau^- \to \mu + \tau$-jet cross–section analysis are displayed in figure 2, 3 and 4, showing

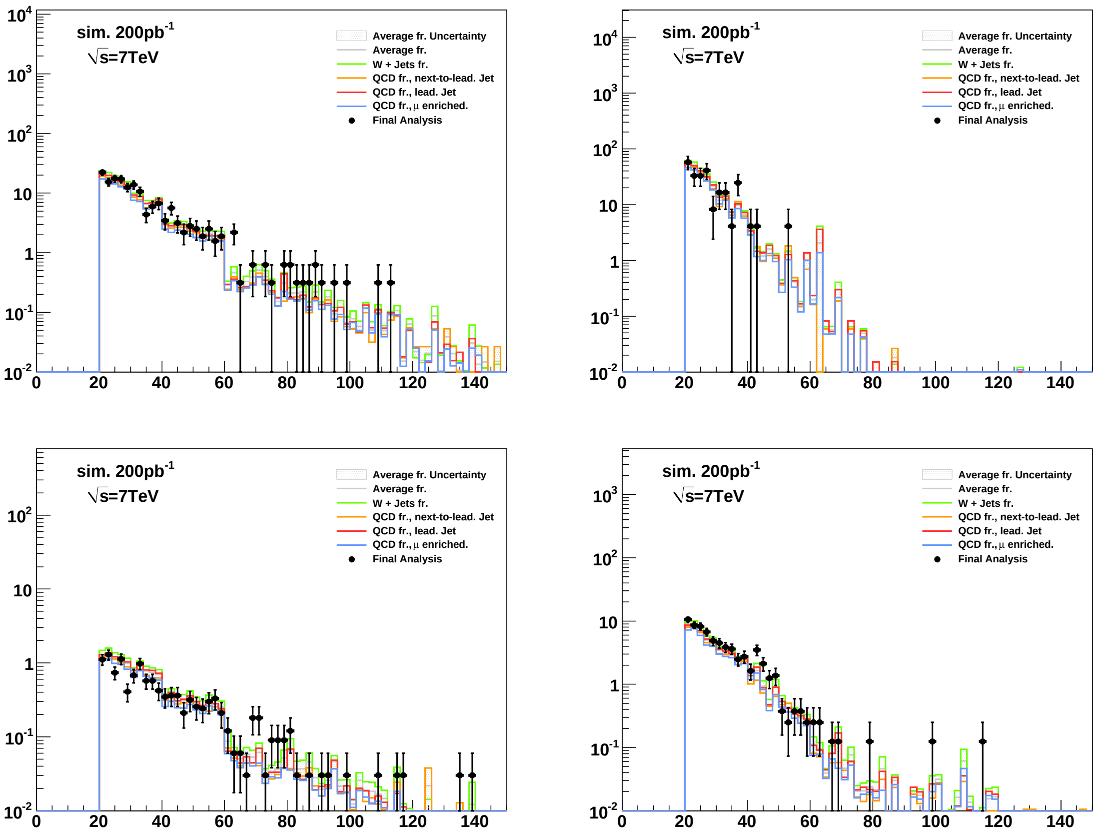

Figure 3: Distributions of transverse momenta of the tau–jet candidates in W+jets (top left), QCD (top right), $t\bar{t}$+jets (bottom left) and $Z \to \mu^+\mu^-$ (bottom right) background events which pass all selection criteria of the $Z \to \tau^+\tau^- \to \mu + \tau$-jet cross–section analysis compared to the estimate obtained from the fake–rate technique, computed according to equation 3. The expected contribution of background processes is indicated by points. Lines of different colors represent the estimates obtained by applying fake–rate weights determined for different compositions of light quark, heavy quark and gluon jets, as described in section 2. The maximum (minimum) estimate is interpreted as upper (lower) bound. The difference between the bounds is taken as systematic uncertainty on the estimate obtained from the “simple” fake–rate method and is represented by the gray shaded area.

distributions of the muon transverse momentum, the transverse momentum of the tau–jet candidates and the visible invariant mass of muon plus tau–jet “composite particles” respectively. The total number of background events estimated by equation 4 is compared to the expected number of QCD, W+jets, $t\bar{t}$+jets and $Z \to \mu^+\mu^-$ events which pass all event selection criteria of the $Z \to \tau^+\tau^- \to \mu + \tau$-jet analysis in table 1. The estimates obtained from the fake–rate method are in good agreement with the contributions expected for QCD and W+jets background processes. Perhaps surprisingly, the fake–rate technique also predicts the contributions of the $t\bar{t}$+jets and $Z \to \mu^+\mu^-$ backgrounds rather precisely. The explanation for the agreement between estimates obtained from the fake–rate method and the expected contributions of $t\bar{t}$+jets and $Z \to \mu^+\mu^-$ backgrounds is illustrated in figure 6. In the figure, it can be seen that most tau–jet candidates in $t\bar{t}$+jets and $Z \to \mu^+\mu^-$ events passing all selection criteria of the final $Z \to \tau^+\tau^- \to \mu + \tau$-jet cross–section analysis do in fact arise from misidentification of quark and gluon jets as hadronic tau lepton decays. $Z \to \mu^+\mu^-$ background contributions arising from misidentification of muons are visible in two bins around the nominal Z mass in

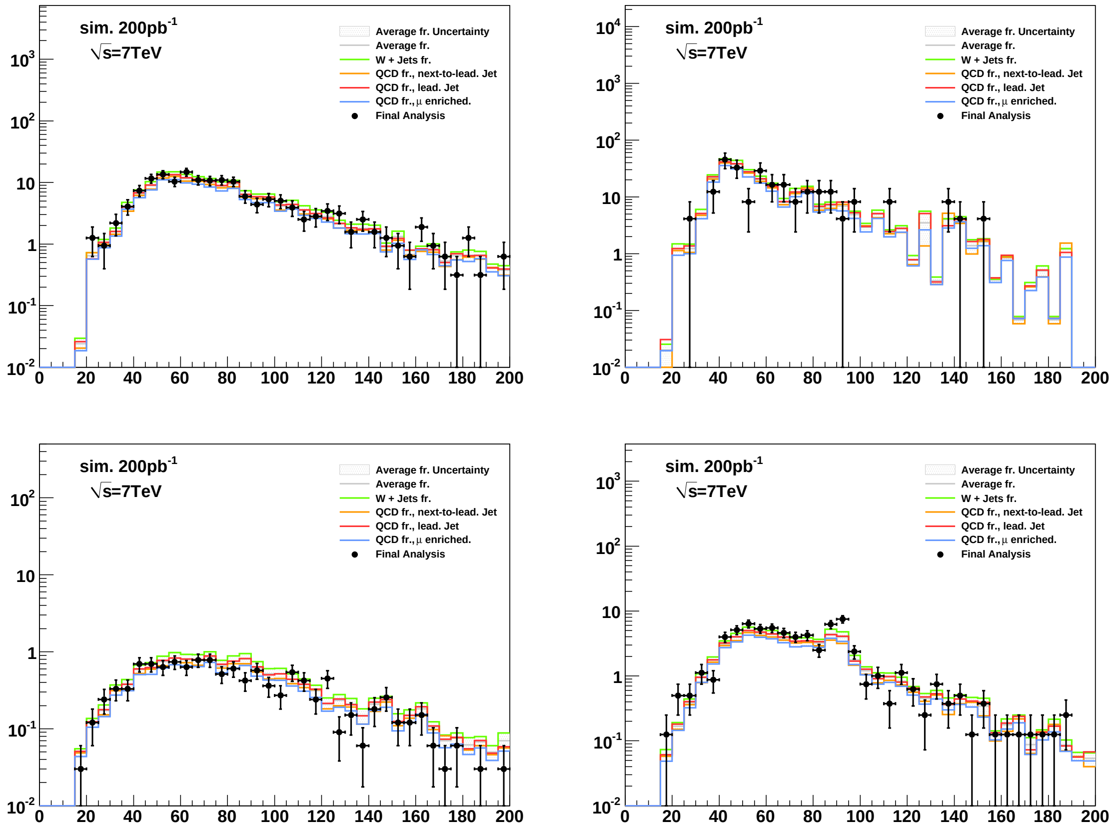

Figure 4: Distributions of the visible invariant mass of muon plus tau–jet in W+jets (top left), QCD (top right), $t\bar{t}$+jets (bottom left) and $Z \rightarrow \mu^+\mu^-$ (bottom right) background events which pass all selection criteria of the $Z \rightarrow \tau^+\tau^- \rightarrow \mu + \tau$-jet cross–section analysis compared to the estimate obtained from the fake–rate technique, computed according to equation 5. The expected contribution of background processes is indicated by points. Lines of different colors represent the estimates obtained by applying fake–rate weights determined for different compositions of light quark, heavy quark and gluon jets, as described in section 2. The maximum (minimum) estimate is interpreted as upper (lower) bound. The difference between the bounds is taken as systematic uncertainty on the estimate obtained from the "simple" fake–rate method and is represented by the gray shaded area.

figure 4 only. Contributions from genuine tau leptons in $t\bar{t}$+jets background events are small and within the variations covered by differences between fake–rate weights determined for varying compositions of light quark, heavy quark and gluon jets as described in section 2.

Different estimates are obtained for the fake–rate probabilities determined for the highest and second highest P_T jet in QCD di–jet events, jets in a muon enriched QCD sample and jets in W+jets events. The arithmetic average of the four estimates together with the difference between the computed average and the minimum/maximum value is given in table 1.

We take the average value as "best" estimate of the background contribution and the difference between the average and the minimum/maximum estimate as its systematic uncertainty. We obtain a value of $\mathcal{O}$ (15%) for the systematic uncertainty and find that the true sum of QCD, W+jets, $t\bar{t}$+jets and $Z \rightarrow \mu^+\mu^-$ background contributions agrees well with the "best" estimate obtained by the fake–rate method within the systematic uncertainty.

Background Process	Expectation	Estimate obtained by applying weights of type: QCD lead. jet	QCD second jet	QCD μ–enriched	W + jets	Average fake–rate estimate
W+jets	163.0 ± 7.1	157.2 ± 2.8	140.9 ± 2.7	129.9 ± 2.5	177.9 ± 3.2	$151.5^{+26.6}_{-21.8}$
QCD	246.4 ± 31.8	269.2 ± 14.0	246.5 ± 14.3	219.7 ± 11.8	300.8 ± 15.2	$259.1^{+44.9}_{-41.7}$
$t\bar{t}$+jets	12.2 ± 0.6	14.3 ± 0.3	12.6 ± 0.3	11.6 ± 0.3	16.5 ± 0.3	$13.8^{+2.7}_{-2.2}$
$Z \to \mu^+\mu^-$	68.6 ± 2.9	58.2 ± 1.3	51.2 ± 1.2	48.5 ± 1.1	65.8 ± 1.4	$55.9^{+10.0}_{-7.5}$
Σ Background	490.4 ± 32.7	499.9 ± 14.4	451.2 ± 14.6	409.7 ± 12.1	561.1 ± 15.6	$480.2^{+82.7}_{-71.9}$
$Z \to \tau^+\tau^-$	–	284.3 ± 3.7	269.0 ± 3.9	256.5 ± 3.3	325.3 ± 4.2	$283.3^{+42.2}_{-27.1}$

Table 1: Number of events from W+jets, QCD, $t\bar{t}$+jets and $Z \to \mu^+\mu^-$ background processes expected to pass all selection criteria of the $Z \to \tau^+\tau^- \to \mu + \tau$-jet cross–section analysis compared to the estimates obtained by weighting events in the background dominated control sample with the "simple" fake–rate weights defined by equation 4.

Note that the estimate for the sum of background contributions which one obtains in case one applies the "simple" fake–rate weights defined by equation 4 to a background dominated control sample selected in data is likely to overestimate the true value of background contributions by a significant amount. The reason is that contributions of the $Z \to \tau^+\tau^-$ signal are non–negligible. In fact, signal contributions to the background dominated control sample are expected to be 14.9% and since the per–jet weights computed by equation 3 are larger on average in signal than in background events, the signal contribution increases by the weighting and amounts to 37.1% of the sum of event weights computed by equation 4 and given in table 1.

The contribution of the $Z \to \tau^+\tau^-$ signal needs to be determined by Monte Carlo simulation and subtracted from the estimate obtained by applying the "simple" fake–rate method to data, in order to get an unbiased estimate of the true background contributions.

3.2 "CDF–type" weights

Instead of subtracting from the estimate obtained for the sum of background contributions a correction determined by Monte Carlo simulation, the signal contribution to the background dominated event sample selected in data can be corrected for by adjusting the weights, based solely on information contained in the analyzed data sample, hence avoiding the need to rely on Monte Carlo based corrections.

In the "CDF–type" method, additional information, namely whether or not tau–jet candidates pass or fail the tau identification cuts and discriminators, is drawn from the data. The desired cancellation of signal contributions is achieved by assigning negative weights to those tau–jet candidates which pass all tau identification cuts and discriminators, i.e. to a fair fraction of genuine hadronic tau decays, but to a small fraction of quark and gluon jets only. The small reduction of the background estimate by negative weights assigned to quark and gluon jets is accounted for by a small increase of the positive weights assigned to those tau–jet candidates for which at least one of the tau identification cuts or discriminators fails. In this way, an unbiased estimate of the background contribution is maintained.

To be specific, the "CDF–type" weights assigned to tau–jet candidates are computed as (for a

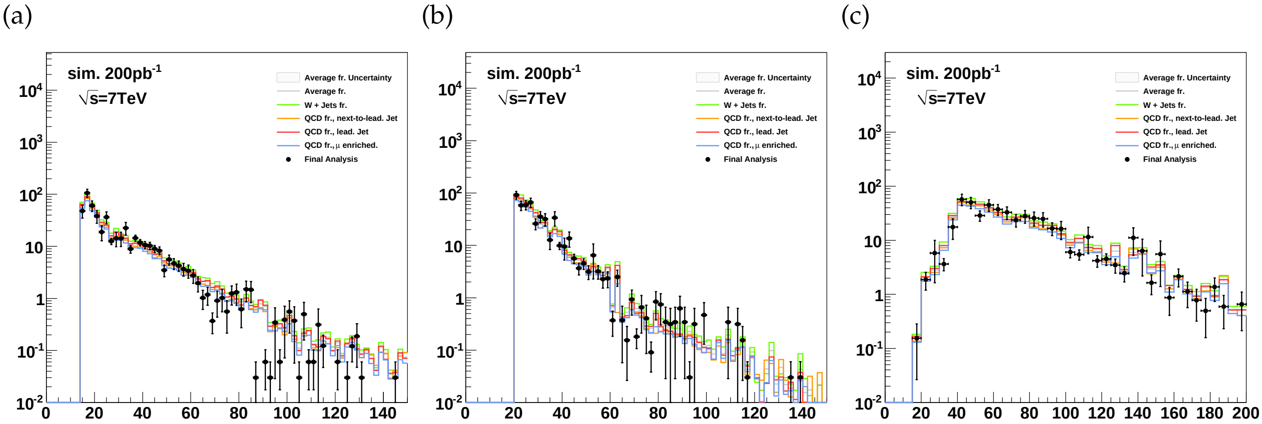

Figure 5: Distributions of muon transverse momentum (a), transverse momenta of the tau–jet candidates (b) and of the visible invariant mass of muon plus tau–jet (c) for the sum of Standard Model background processes W+jets (top left), QCD, $t\bar{t}$+jets and $Z \to \mu^+\mu^-$ passing all selection criteria of the $Z \to \tau^+\tau^- \to \mu + \tau$-jet cross–section analysis compared to the estimate obtained from the fake–rate technique, computed according to equations 3, 4 and 5. The expected contribution is indicated by points. Lines of different colors represent the estimates obtained by applying fake–rate weights determined for different compositions of light quark, heavy quark and gluon jets, as described in section 2. The maximum (minimum) estimate is interpreted as upper (lower) bound. The difference between the bounds is taken as systematic uncertainty on the estimate obtained from the "simple" fake–rate method and is represented by the gray shaded area.

derivation of the formula, see section 5.1 of the appendix):

$$w_{jet}^{CDF}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right) := \begin{cases} \frac{P_{fr}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right)\cdot\varepsilon\left(P_T^{jet}, \eta_{jet}, R_{jet}\right)}{\varepsilon\left(P_T^{jet}, \eta_{jet}, R_{jet}\right) - P_{fr}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right)} & \text{if all tau id. cuts and discriminators passed} \\ \frac{P_{fr}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right)\cdot\left(1-\varepsilon\left(P_T^{jet}, \eta_{jet}, R_{jet}\right)\right)}{\varepsilon\left(P_T^{jet}, \eta_{jet}, R_{jet}\right) - P_{fr}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right)} & \text{otherwise} \end{cases} \tag{6}$$

Event weights and the weights assigned to "composite particles" are computed in the same way as for the "simple" weights, based on the weights assigned to the tau–jet candidates:

$$\begin{aligned} W_{event}^{CDF} &:= \Sigma w_{jet}^{CDF} \\ w_{comp-part}^{CDF}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right) &:= w_{jet}^{CDF}\left(P_T^{jet}, \eta_{jet}, R_{jet}\right), \end{aligned} \tag{7}$$

where the sums extend over all jets in the background dominated control sample which pass the preselection defined by the denominator of equation 2.

The effect of the negative weights to compensate the positive weights in case the "CDF–type" fake–rate method is applied to signal events containing genuine hadronic tau decays is shown in table 2. As expected, positive and negative weights do indeed cancel in the statistical average.

Figures 7, 8 and 9 demonstrate that an unbiased estimate of the background contribution by the "CDF–type" weights is maintained. Overall, the estimates obtained are in good agreement with the contributions expected for different background processes, indicating that the adjustment of negative and positive weights works as expected for the background as well.

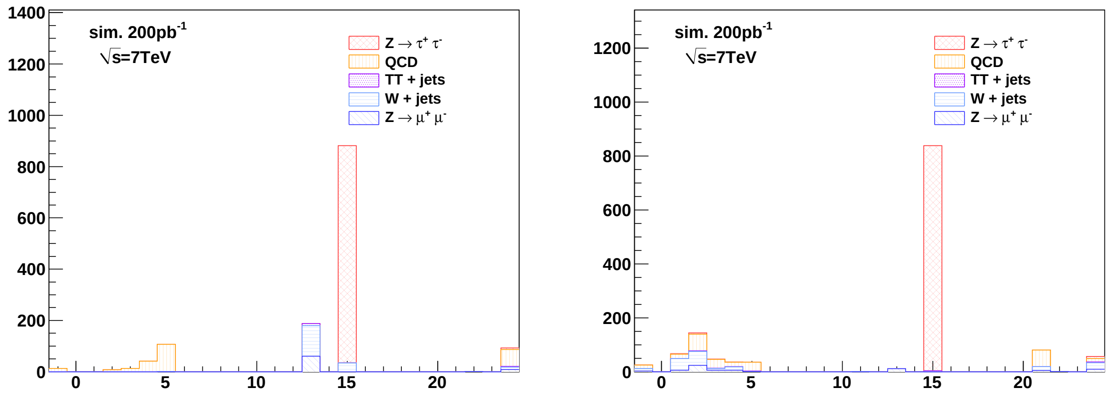

Figure 6: Type of generator level particle which matches muons (left) and tau–jet candidates (right) reconstructed in signal and different types of background events passing all selection criteria of the $Z \rightarrow \tau^+\tau^- \rightarrow \mu + \tau$-jet cross–section analysis. Particle species are encoded according to the definition of the particle data group [6]: numbers 1-5 represent quarks, a value of 11 (13) represents electrons (muons), 15 represents tau leptons (all tau decay modes included) and 22 gluons. Events for which unambiguous matching of reconstructed to generator level particles failed are represented by entries in bins 0 and 24.

Background Process	Expectation	Estimate obtained by applying weights of type: QCD lead. jet	QCD second jet	QCD μ–enriched	W + jets	Average fake–rate estimate
W+jets	163.0 ± 7.1	163.2 ± 3.8	140.6 ± 3.4	128.0 ± 3.1	188.3 ± 4.2	$155.0^{+33.6}_{-27.3}$
QCD	246.4 ± 31.8	300.5 ± 19.5	266.1 ± 19.0	236.0 ± 16.4	335.1 ± 20.4	$284.4^{+55.5}_{-52.0}$
$t\bar{t}$+jets	12.2 ± 0.6	13.1 ± 0.3	11.5 ± 0.3	10.2 ± 0.3	15.4 ± 0.4	$12.6^{+2.8}_{-2.4}$
$Z \rightarrow \mu^+\mu^-$	68.6 ± 2.9	52.7 ± 1.4	46.7 ± 1.4	41.9 ± 1.2	60.3 ± 1.6	$50.4^{+10.1}_{-8.6}$
Σ Background	490.4 ± 32.7	529.5 ± 19.9	464.9 ± 19.3	416.1 ± 16.8	599.1 ± 20.9	$502.4^{+99.4}_{-88.4}$
$Z \rightarrow \tau^+\tau^-$	–	0.3 ± 2.4	-10.6 ± 2.5	3.8 ± 2.0	-10.8 ± 2.8	$-4.3^{+8.4}_{-7.2}$

Table 2: Number of events from W+jets, QCD, $t\bar{t}$+jets and $Z \rightarrow \mu^+\mu^-$ background processes expected to pass all selection criteria of the $Z \rightarrow \tau^+\tau^- \rightarrow \mu + \tau$-jet cross–section analysis compared to the estimates obtained by weighting events in the background dominated control sample with the "CDF–type" fake–rate weights defined by equation 7.

Results obtained by the "CDF–type" fake–rate method are summarized in table 2, in which the total number of background events estimated by equation 7 is compared to the true background contributions. The "best" estimate of the background contribution obtained from the "CDF–type" method is again taken as the arithmetic average of the estimates obtained by applying the fake–rate probabilities for the highest and second highest P_T jet in QCD di–jet events, jets in a muon enriched QCD sample and jets in W+jets events. Systematic uncertainties are taken from the difference between the computed average value and the minimum/maximum estimate. We obtain a value of $\mathcal{O}$ (15–20%) for the systematic uncertainty of the "CDF–type" method, slightly higher than the systematic uncertainty obtained for the "simple" method. The small increase of systematic uncertainties is in agreement with our expectation for fluctuations of the jet–weights in case weights of negative and positive sign are used.

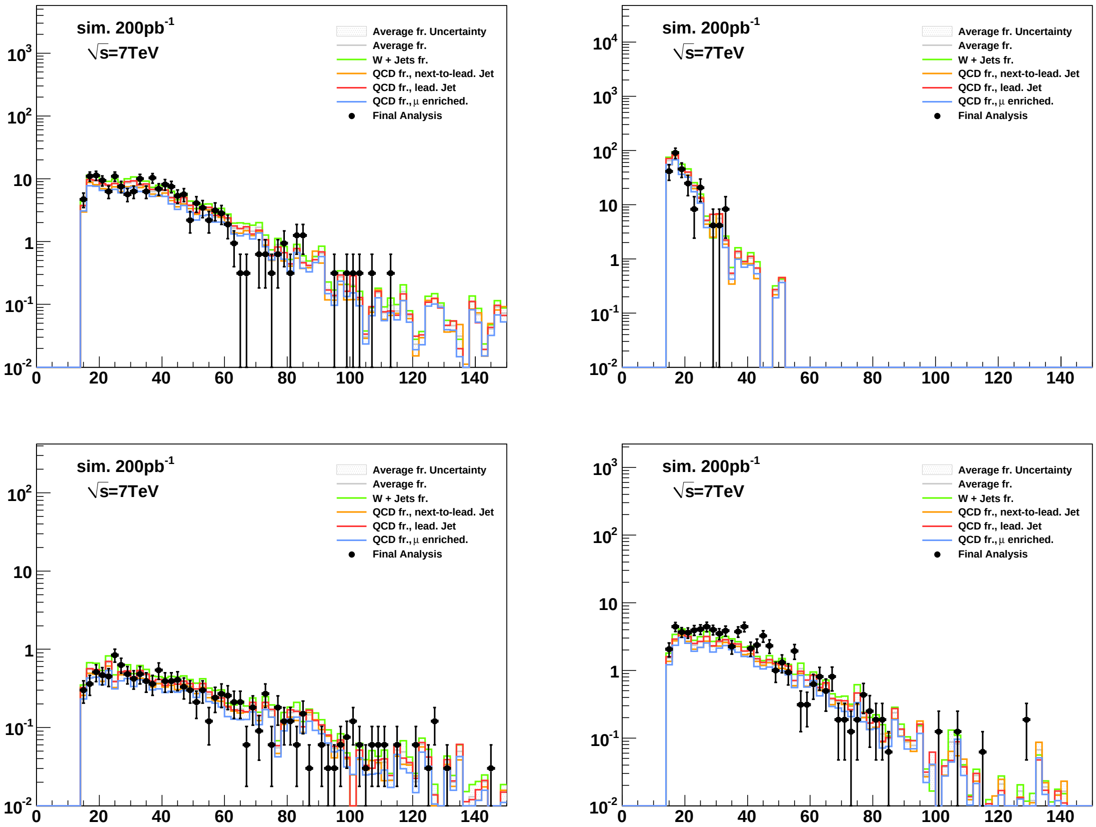

Figure 7: Distributions of muon transverse momentum in W+jets (top left), QCD (top right), $t\bar{t}$+jets (bottom left) and $Z \rightarrow \mu^+\mu^-$ (bottom right) background events which pass all selection criteria of the $Z \rightarrow \tau^+\tau^- \rightarrow \mu + \tau$-jet cross–section analysis compared to the estimate obtained from the fake–rate technique, computed according to equation 7. The expected contribution of background processes is indicated by points. Lines of different colors represent the estimates obtained by applying fake–rate weights determined for different compositions of light quark, heavy quark and gluon jets, as described in section 2. The maximum (minimum) estimate is interpreted as upper (lower) bound. The difference between the bounds is taken as systematic uncertainty on the estimate obtained from the "CDF–type" fake–rate method and is represented by the gray shaded area.

4 Summary

Two different methods for estimating background contributions to the final event sample of physics analyses involving tau leptons in the final state have been presented. The background estimates obtained from the "simple" and "CDF–type" methods are found to be in good agreement with the expectation within the estimated systematic uncertainties of $\mathcal{O}(15\%)$. The "simple" method relies on corrections for signal contributions which need to be determined by the Monte Carlo simulation, while the "CDF–type" method utilizes additional information drawn from the data and does not need Monte Carlo based corrections.

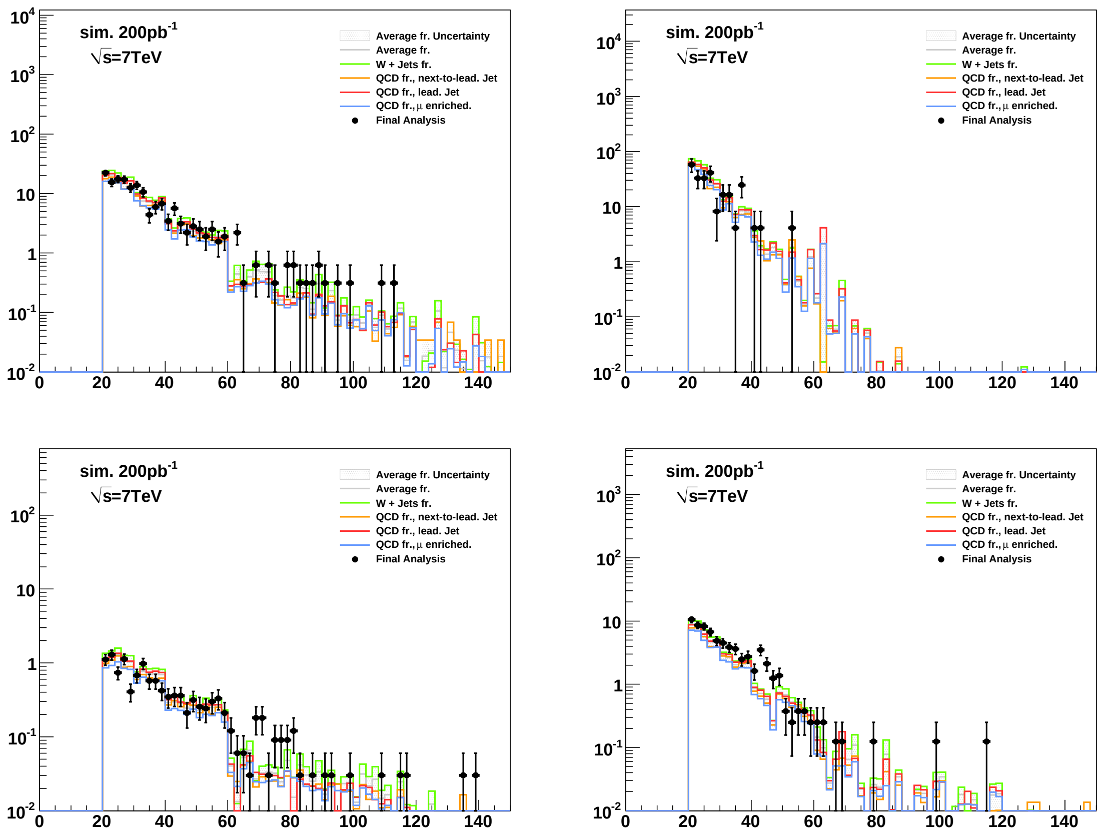

Figure 8: Distributions of transverse momenta of the tau–jet candidates in W+jets (top left), QCD (top right), $t\bar{t}$+jets (bottom left) and $Z \to \mu^+\mu^-$ (bottom right) background events which pass all selection criteria of the $Z \to \tau^+\tau^- \to \mu + \tau$-jet cross–section analysis compared to the estimate obtained from the fake–rate technique, computed according to equation 6. The expected contribution of background processes is indicated by points. Lines of different colors represent the estimates obtained by applying fake–rate weights determined for different compositions of light quark, heavy quark and gluon jets, as described in section 2. The maximum (minimum) estimate is interpreted as upper (lower) bound. The difference between the bounds is taken as systematic uncertainty on the estimate obtained from the "CDF–type" fake–rate method and is represented by the gray shaded area.

5 Appendix

5.1 Derivation of "CDF–type" fake–rate weights

The basic idea of the "CDF–type" weights is to assign negative (positive) weights to tau–jet candidates passing all tau identification cuts and discriminators (failing at least one cut or discriminator), such that signal contributions of genuine hadronic tau decays to the background dominated control sample on average cancel after application of the weights, while providing an unbiased estimate of the contribution of background processes arising from misidentification of quark and gluon jets.

For the derivation of equation 6 for the "CDF–type" weights assigned to tau–jet candidates, we will use the following notation: Let n_τ (n_{QCD}) denote the total number of tau–jets (quark and gluon jets) in a certain bin of transverse momentum P_T^{jet}, pseudo–rapidity η_{jet} and jet–width R_{jet} and n_τ^{sel} (n_{QCD}^{sel}) denote the number of tau–jets (quark and gluon jets) in that bin which pass

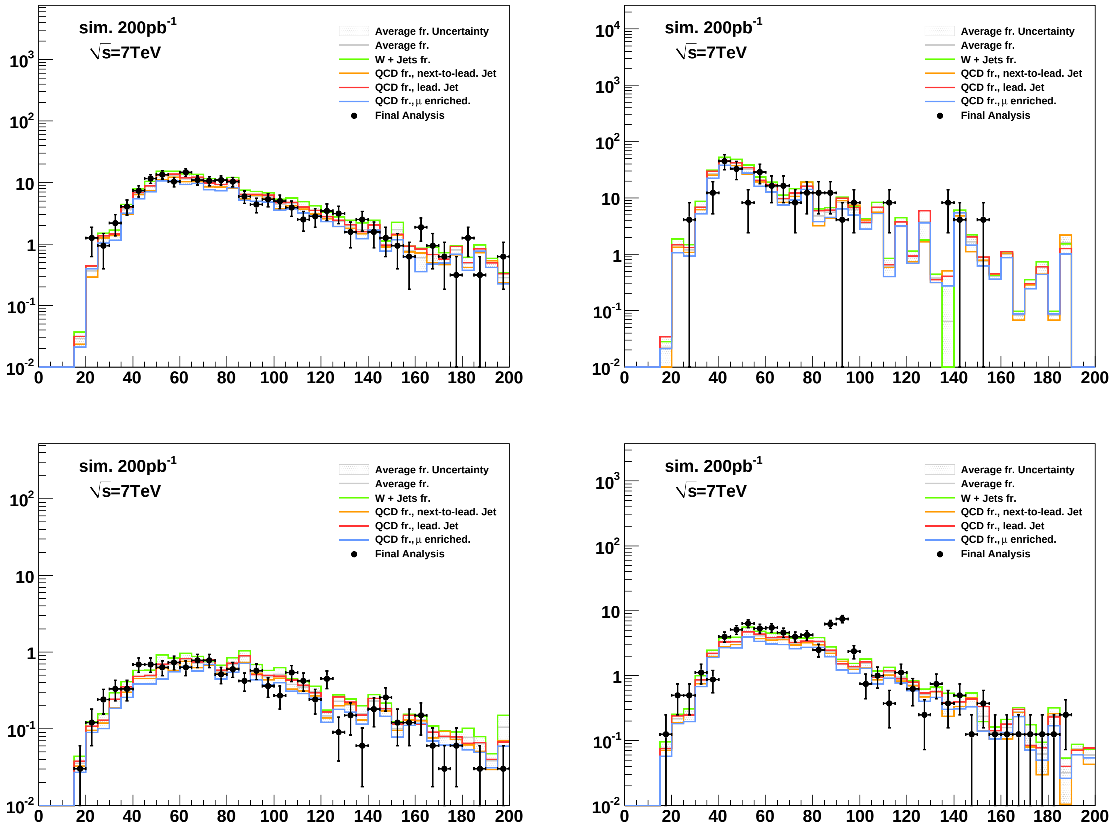

Figure 9: Distributions of the visible invariant mass of muon plus tau–jet in W+jets (top left), QCD (top right), $t\bar{t}$+jets (bottom left) and $Z \to \mu^+\mu^-$ (bottom right) background events which pass all selection criteria of the $Z \to \tau^+\tau^- \to \mu + \tau$-jet cross–section analysis compared to the estimate obtained from the fake–rate technique, computed according to equation 7. The expected contribution of background processes is indicated by points. Lines of different colors represent the estimates obtained by applying fake–rate weights determined for different compositions of light quark, heavy quark and gluon jets, as described in section 2. The maximum (minimum) estimate is interpreted as upper (lower) bound. The difference between the bounds is taken as systematic uncertainty on the estimate obtained from the "CDF–type" fake–rate method and is represented by the gray shaded area.

all tau identification cuts and discriminators.

By definition of the tau identification efficiency $\varepsilon := \varepsilon\left(P_T^{jet}, \eta_{jet}, R_{jet}\right)$ and fake–rate $f := f\left(P_T^{jet}, \eta_{jet}, R_{jet}\right)$:

$$\begin{aligned} n_\tau^{sel} &= \varepsilon \cdot n_\tau \\ n_{QCD}^{sel} &= f \cdot n_{QCD}. \end{aligned} \quad (8)$$

Depending on whether or not a given tau–jet candidate passes all tau identification cuts and discriminators or not, we will assign a weight of value w_{passed} or w_{failed} to it.

The values of the weights w_{passed} and w_{failed} shall be adjusted such that they provide an unbiased estimate of the background contribution:

$$w_{passed} \cdot f \cdot n_{QCD} + w_{failed} \cdot (1 - f) \cdot n_{QCD} \equiv n_{QCD}^{sel} = f \cdot n_{QCD} \quad (9)$$

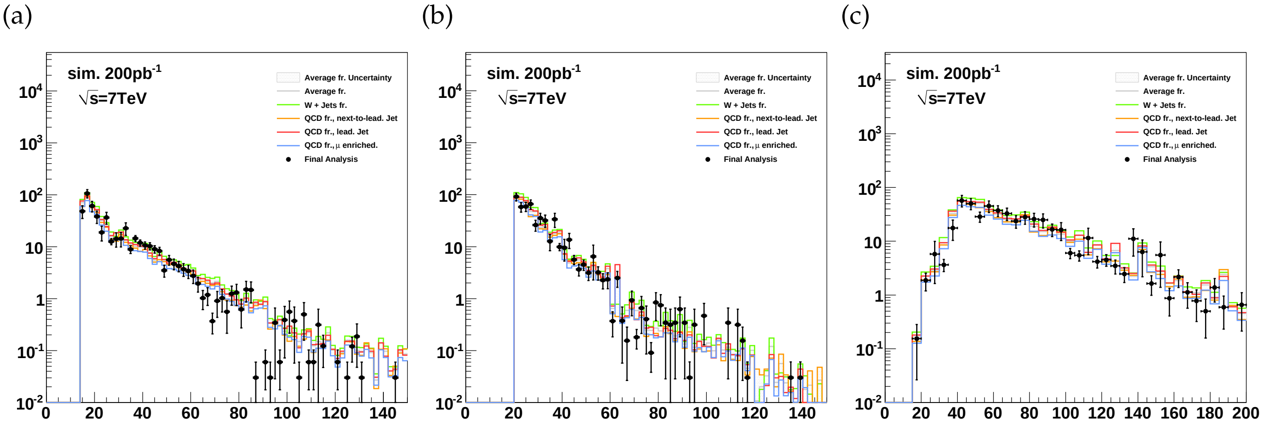

Figure 10: Distributions of muon transverse momentum (a), transverse momenta of the tau–jet candidates (b) and of the visible invariant mass of muon plus tau–jet (c) for the sum of Standard Model background processes W+jets (top left), QCD, $t\bar{t}$+jets and $Z \rightarrow \mu^+\mu^-$ passing all selection criteria of the $Z \rightarrow \tau^+\tau^- \rightarrow \mu + \tau$-jet cross–section analysis compared to the estimate obtained from the fake–rate technique, computed according to equations 6 and 7. The expected contribution is indicated by points. Lines of different colors represent the estimates obtained by applying fake–rate weights determined for different compositions of light quark, heavy quark and gluon jets, as described in section 2. The maximum (minimum) estimate is interpreted as upper (lower) bound. The difference between the bounds is taken as systematic uncertainty on the estimate obtained from the "CDF–type" fake–rate method and is represented by the gray shaded area.

while averaging to zero for genuine hadronic tau decays:

$$w_{passed} \cdot \varepsilon \cdot n_\tau + w_{failed} \cdot (1 - \varepsilon) \cdot n_\tau \equiv 0.$$

The latter equation yields the relation:

$$w_{passed} = -\frac{1-\varepsilon}{\varepsilon} \cdot w_{failed}, \tag{10}$$

associating the two types of weights. By inserting relation 10 into equation 9 we obtain:

$$\begin{aligned} & -\frac{1-\varepsilon}{\varepsilon} \cdot w_{failed} \cdot f \cdot n_{QCD} + w_{failed} \cdot (1-f) \cdot n_{QCD} = f \cdot n_{QCD} \\ \Rightarrow \quad & \left(\frac{-f + \varepsilon \cdot f + \varepsilon - f \cdot \varepsilon}{\varepsilon} \right) \cdot w_{failed} = f \\ \Rightarrow \quad & w_{failed} = \frac{f \cdot \varepsilon}{\varepsilon - f} \end{aligned}$$

and

$$w_{passed} = -\frac{f \cdot (1-\varepsilon)}{\varepsilon - f} \tag{11}$$

which matches exactly equation 6 for the "CDF–type" weights applied to tau–jet candidates given in section 3.2.

5.2 Usage of the fake–rate weights in CMSSW analyses

In this section we aim to provide a recipe how to use the "simple" fake–rate weights defined by equations 3, 4 and 5 and the "CDF–type" fake–rate weights defined by equations 6 and 7 in CMSSW analyses.

Package	Version
RecoTauTag/Configuration	V00–20–00
RecoTauTag/RecoTau	V00–20–00
RecoTauTag/TauAnalysisTools	V00–04–00
RecoTauTag/TauTagTools	V00–20–00

Table 3: CMSSW software packages used in addition to release CMSSW_3_3_6_patch5.

The probabilities defined by equation 2 for quark and gluon jets to fake the signature of a hadronic tau decay may depend to a large extent upon the sample preselection, exact tau identification cuts, and discriminator requirements applied on the analysis level. We therefore include a prescription for modifying the configuration to redetermine these probabilities, either from simulated Monte Carlo samples or from data.

In general, the usage of the fake–rate weights described in this note requires five steps to be taken:

- production of Ntuples containing tau–jet candidates
- determination of fake–rate probabilities
- production of either "simple" or "CDF–type" weights
- addition of weights to pat::Tau
- application of weights in the analysis code

We will describe the individual steps in detail in the following.

We aimed to make the recipe for how to use the fake–rate weights in CMSSW analyses as much as possible independent of a specific CMSSW software version. We expect, however, that the technical details may depend (at least to some extent) on the actual software version used, and for this reason we would like to mention that the recipe given in the following has been tested with release CMSSW_3_3_6_patch5. Additional packages which we have used to obtain the results presented in this note are listed in table 3.

5.2.1 Ntuple production

The fake–rate technique workflow begins with the production of Ntuple files which contain information about the selected sample of reconstructed tau–jet candidates. Each Ntuple entry represents one tau–jet candidate. The entry contains information about P_T, $|\eta|$ and R_{jet} of the tau–jet candidates, as well as of all individual tau discriminators. The observables used for parametrization and the discriminators associated to each tau–jet candidate are defined in `RecoTauTag/TauAnalysisTools/python/tools/ntupleDefinitions.py`.

The Ntuple files are produced by configuration files stored in `RecoTauTag/TauAnalysisTools/test/fakeRate`. The Ntuple production is performed on the grid via CRAB, the CRAB jobs being configured to return the Ntuple data files to the user. The DBS names of RECO/AOD files produced for $\sqrt{s} = 7$ TeV centre–of–mass energy with CMSSW_3_1_2 (Summer'09 Monte Carlo production) and of the CRAB crab configuration files for producing the Ntuples which we used to determine fake–rate probabilities and tau identification efficiencies are listed in table 4.

We have produced Ntuple files for QCD di–jet, muon enriched QCD and W+jets background samples and for the $Z \to \tau^+\tau^-$ signal. The Ntuples are available on castor, at the locations

Sample (Steering file)	Events
QCD Dijet (separately for highest and second highest P_T jets; `multicrab.cfg`)	
`/QCDDiJet_Pt20to30/Summer09-MC_31X_V3_7TeV_AODSIM-v1/AODSIM`	1008752
`/QCDDiJet_Pt30to50/Summer09-MC_31X_V3_7TeV_AODSIM-v1/AODSIM`	1010858
`/QCDDiJet_Pt50to80/Summer09-MC_31X_V3_7TeV_AODSIM-v1/AODSIM`	108517
`/QCDDiJet_Pt80to120/Summer09-MC_31X_V3_7TeV_AODSIM-v1/AODSIM`	581928
`/QCDDiJet_Pt120to170/Summer09-MC_31X_V3_7TeV_AODSIM-v1/AODSIM`	52062
Muon enriched QCD (`crab_ppmux.cfg`)	
`/InclusiveMu15/Summer09-MC_31X_V3_7TeV_AODSIM-v1/AODSIM`	5261979
W+jets (`crab_wjets.cfg`)	
`/WJets-sherpa/Summer09-MC_31X_V3_7TeV_AODSIM-v1/AODSIM`	10980000
$Z \to \tau^+\tau^-$ (`crab_ztt.cfg`)	
`/Ztautau/Summer09-MC_31X_V3_7TeV_AODSIM-v1/AODSIM`	2099189

Table 4: Samples used to produce the Ntuples used in fake rate computation for the various channels. The configuration steering files given in brackets are located in the directory `RecoTauTag/TauAnalysisTools/test/fakeRate`.

Sample	Tau–jet candidates	Ntuple locations
QCD highest P_T jet	688627	`V000400/DijetXX_YY`
QCD second highest P_T jet	240004	`V000400/DijetXX_YY`
Muon enriched QCD	1219239	`V000400/InclusiveMu15`
W+jets	286876	`V000400/WJets`
$Z \to \tau^+\tau^-$	414770	`V000400/Ztautau`

Table 5: Location of Ntuple files used for determination of fake–rates in QCD di–jet, muon enriched QCD and W+jets background and for determination of tau identification efficiencies in $Z \to \tau^+\tau^-$ signal events. The Ntuples were produced from RECO/AOD Monte Carlo sample generated and reconstructed for $\sqrt{s} = 7$ TeV centre–of–mass energy with CMSSW_3_1_2 and are located in subdirectories of `/castor/cern.ch/user/friis/TauFakeRate`. For the QCD di–jet samples, the ntuples are binned by generator level $\hat{P}_T$, and must be weighted to correspond to the same integrated luminosity when the used to produce the fake–rate histograms.

specified in table 5.

As described in section 2, different sets of fake–rate probabilities are computed for the highest P_T ("leading") and second highest P_T jet in QCD di–jet events, for jets in QCD events in which a muon is reconstructed and for jets in "electroweak" events selected by requiring a W boson in the final state. The idea of computing separate sets of fake–rates for these samples is to "cover" differences in fake–rates between light quark, heavy quark and gluon jets in this way. We suggest to use these different sets to compute corresponding background estimates separately for each set, compute the "final" background estimate obtained from the fake–rate method as the average of the four sets and take the difference between the average and the minimum/maximum estimate as systematic uncertainty.

Filename	Content
Fake–rates	
`dijet_highpt_histograms.root`	for the highest P_T jet in QCD di–jet events
`dijet_secondpt_histograms.root`	for the second highest P_T jet in QCD di–jet events
`ppmux_histograms.root`	for jets in muon enriched QCD events
`wjets_histograms.root`	for jets in W+jets events
Tau identification efficiencies	
`ztt_histograms.root`	for $Z \rightarrow \tau^+\tau^-$ events

Table 6: Location of ROOT files containing fake–rate probability (top part) and tau identification efficiency (bottom part) histograms. The ROOT files are located in subdirectories of `/afs/cern.ch/user/f/friis/public/TauFakeRateMar11/`.

5.2.2 Fake–rate determination

The fake–rates are stored in three–dimensional histograms (of ROOT type TH3F) which are used as look–up tables. The histograms are parametrized by P_T (x–axis), $|\eta|$ (y–axis) and R_{jet} (z–axis) as described in section 2. Fake–rate probabilities and tau identification efficiencies are computed according to equation 2, separately for each bin.

Per default, the three–dimensional histograms are defined by the bin–edges:

- $P_T =$ 20, 25, 30, 40, 60, 120
- $|\eta| =$ 0, 0.5, 1.0, 1.2, 1.5, 2.0, 2.5
- $R_{jet} =$ 0, 0.0125, 0.025, 0.05, 0.075, 0.10, 0.15, 0.20, 0.30

in the P_T, $|\eta|$, R_{jet} parameter space.

The binning has been optimized for the $Z \rightarrow \tau^+\tau^- \rightarrow \mu + \tau$-jet analysis. A finer binning has been used in the region $P_T \gtrsim 20$ GeV, since both the tau identification efficiency and the fake–rate values considerably vary in the vicinity of the P_T threshold. Note that you will need to change the binning in case you are analyzing tau leptons of significantly higher P_T.

The fake–rate probabilities for jets in QCD di–jet, muon enriched QCD and W+jets background events are stored in separate ROOT files, which are available on castor at the locations given in table 6. Each ROOT file contains histograms for various combinations of tau identification cuts and discriminator requirements. For each such requirement, a total of three histograms is contained in the ROOT file, storing the number of events passing the requirements of the numerator and the denominator in equation 2, and the fake–rate probability and tau identification efficiency values that are computed by dividing the numerator by the denominator histogram. For the ROOT files available on castor, the denominator is defined by the selection $P_T > 20$ GeV, $|\eta| < 2.1$, leading track $P_T > 5$ GeV and discriminators against electrons and muons passed.

In order to produce sets of histograms matching the tau identification criteria used in your analysis, you need to edit the configuration files specifying the selection criteria for numerator and denominator. The selection criteria for the numerator are defined in `RecoTauTag/TauAnalysisTools/python/fakeRate/histogramConfiguration.py`, while the selection criteria for the denominator are defined in `RecoTauTag/TauAnalysisTools/test/fakeRate/sources.py`. The binning, location of the Ntuple files, desired output location and the luminosity normalization factors are also defined in the latter configuration file. Note that the selection criteria applied in the denominator must not be looser than the requirements applied during event skimming.

After editing the configuration files, you can start executing the *cmsRun* jobs which produce the three–dimensional histograms. Parallel execution of those jobs on a condor compatible batch system is supported. The jobs are submitted by the command:

```
cd $CMSSW_BASE/src/RecoTauTag/TauTagTools/test/fakeRate
condor_submit SubmitHistograms.jdl
```

Once all jobs have finished execution, you need to merge the output of different jobs. Execute:

```
 cd $CMSSW_BASE/src/RecoTauTag/TauTagTools/test/fakeRate
./mergeHistograms.sh
```

In the case that the user does not have access to a batch system supporting condor, the histograms can be merged sequentially by executing:

```
cd $CMSSW_BASE/src/RecoTauTag/TauTagTools/test/fakeRate
python make_histograms.py
```

Finally, in order to use the sets of histograms as look–up tables for production of per–jet and of event weights, edit the configuration file `RecoTauTag/TauAnalysisTools/PFTauEfficiencyAssociator_cfi.py` and set the *filename* attributes of *efficiencySources* configuration parameters to the location of the ROOT files produced by executing *mergeHistograms.sh*. (our recommendation is to copy the ROOT files into the *public* subdirectory of your home directory that you have on `/afs/cern.ch/user/..`).

5.2.3 Weight production

The production of the per–jet and of the event weights proceeds in two stages. In the first stage, per–jet weights are computed according to equations 3 (6) and stored in the RECO/AOD event–content as pat::LookupTableRecord objects associated to the reco::PFTau collection. In the second stage, event weights are computed according to equations 4 (7). are embedded into the pat::Tau collection.

In order to compute the per–jet weights, you first need to add:

```
from RecoTauTag.TauAnalysisTools.PFTauEfficiencyAssociator_cfi import *
from TauAnalysis.BgEstimationTools.fakeRateJetWeightProducer_cfi.py import *
```

to your top–level python configuration file (the one ending with *_cfg.py*) and add the sequence *associateTauFakeRates* plus the module *bgEstFakeRateJetWeights* to your analysis path. You then need to edit the file `TauAnalysis/BgEstimationTools/python/fakeRateJetWeightProducer_cfi.py` and adapt the following configuration parameters to the needs of your analysis:

- *method*
  the string type parameter which specifies if per–jet weights are to be computed via either equation 3 ("simple") or equation 6 ("CDF")
- *allTauJetSource*
  the InputTag that specifies the reco::PFTau collection stored in the RECO/AOD event–content for which tau identification efficiency and fake–rate probabilities are to be

computed. Per default, tau identification efficiencies and fake–rates are computed for the reco::PFTau collection reconstructed by the particle–flow based tau identification algorithm for a "shrinking signal cone" of size $dR_{sig} = 5.0/E_T$. Note that the value of this configuration parameter needs to match the collection used when determining the tau identification and fake–rate probabilities as described in section 5.2.2 as well as the collection used as input for pat::Tau production, specified in the configuration file
`PhysicsTools/PatAlgos/python/producersLayer1/tauProducer_cfi.py`

- *preselTauJetSource*
  the InputTag which specifies the collection of tau–jet candidates passing the preselection requirements defined by the denominator of equation 2. The collection of preselected tau–jet candidates may either be of type std::vector<reco::PFTau> or of type std::vector<pat::Tau>
- *frTypes*
  the configuration parameter set which specifies the different sets of fake–rate probabilities for which per–jet weights are to be computed. The parameter which you need to adapt to your analysis is the InputTag of the tau–jet discriminator collection which indicates whether tau–jet candidates have passed the preselection or not. The tau–jet discriminator collection needs to be of type reco::PFTauDiscriminator. The value of this discriminator is expected to be 1 if a given tau–jet candidates passes all preselection requirements and 0 otherwise. Note that a single discriminator collection needs to contain the information concerning all cuts and tau identification discriminators applied during preselection. You may need to produce such discriminator collection prior to adding the *bgEstFakeRateJetWeights* module to your analysis module (an example for this can e.g. be found in
  `TauAnalysis/RecoTools/python/recoPFTauIdentification_cfi.py`)

(the other configuration parameters can be kept at their default values).

In order to compute the event weights, you need to add:

```
from TauAnalysis.BgEstimationTools.fakeRateEventWeightProducer_cfi.py import *
```

to your top–level python configuration file and add the module *bgEstFakeRateEventWeights* to your analysis path. Per default, all configuration parameters for the *bgEstFakeRateEventWeights* module are taken from the parameters of the *bgEstFakeRateJetWeights* module which you already configured. This guarantees consistency between the configuration parameters of the two modules and simplifies the configuration to the extend that there is no need to edit the
`TauAnalysis/BgEstimationTools/python/fakeRateEventWeightProducer_cfi.py`
file.

As the pat::LookupTableRecord objects are needed for production of the pat::Tau collection, the first stage needs to proceed before the PAT–tuple production sequence, while the second stage proceeds best after the PAT–tuple production sequence.

5.2.4 Addition of weights to pat::Taus

In order to add the per–jet weights produced as described in section 5.2.3 to the pat::Tau objects, such that they can be used in your analysis later, you simply need to add:

```
from PhysicsTools.PatAlgos.producersLayer1.tauProducer_cfi.py import *
from RecoTauTag.TauAnalysisTools.PFTauEfficiencyAssociator_cfi import *
```

```
from TauAnalysis.BgEstimationTools.fakeRateJetWeightProducer_cfi.py import *

process.allLayer1Taus.efficiencies = cms.PSet()
build_pat_efficiency_loader(shrinkingConeZTTEffSimAssociator,
                    None, process.allLayer1Taus.efficiencies)
build_pat_efficiency_loader(shrinkingConeWJets,
                            None, process.allLayer1Taus.efficiencies)
build_pat_efficiency_loader(shrinkingConeMuEnrichedQCDAssociator,
                            None, process.allLayer1Taus.efficiencies)
build_pat_efficiency_loader(shrinkingConeDiJetHighPt,
                            None, process.allLayer1Taus.efficiencies)
build_pat_efficiency_loader(shrinkingConeDiJetSecondPt,
                            None, process.allLayer1Taus.efficiencies)
frTypes = getPSetAttributes(process.bgEstFakeRateJetWeights.frTypes)
for frType in frTypes:
  frLabel = "".join(["bgEstFakeRateJetWeight", "_", frType])
  frInputTag = cms.InputTag('bgEstFakeRateJetWeights', frType)
  setattr(process.allLayer1Taus.efficiencies, frLabel, frInputTag)

process.allLayer1Taus.addEfficiencies = cms.bool(True)
```

to your top–level python configuration file.

5.2.5 Application of weights in the analysis

As described in section 3, usage of the per–jet and of the event weights for the purpose of estimating background contributions via the fake–rate technique requires to separate steps.

In the first step, you need to loosen the tau identification cuts and discriminators and apply only the preselection requirements for which you have computed the fake–rates, matching the denominator of equation 2.

In the second step, you will then fill histograms of tau–jet related observables with jet–weights, histograms of per–particle quantities for particles of types different from hadronically decaying tau leptons with event weights and sum the event weights in order to obtain an estimate for the total number of background events passing all event selection criteria applied on the final analysis level. The following code snippet demonstrates how the event and per–jet weights can be accessed from within an EDAnalyzer:

```
edm::Handle<double> evtWeight;
edm::InputTag evtWeightSource('bgEstFakeRateEventWeights', 'qcdMuEnriched');
evt.getByLabel(evtWeightSource, evtWeight);

std::cout << "event weight = " << (*evtWeight) << std::endl;

edm::Handle<pat::TauCollection> patTaus;
evt.getByLabel('cleanLayer1Taus', patTaus);

for ( std::vector<pat::Tau>::const_iterator patTau = patTaus->begin();
      patTau != patTaus->end(); ++patTau ) {
  std::cout << "per-jet weight = "
            << patTau->efficiency("bgEstFakeRateJetWeight_qcdMuEnriched").value()
            << std::endl;
  std::cout << "(tau id. efficiency = "
            << patTau->efficiency("effByStandardChainZTTsim").value()
            << "," << " fake-rate = "
```

Label	Instance
bgEstFakeRateEventWeights	qcdDiJetLeadJet qcdDiJetSecondLeadJet qcdMuEnriched WplusJets

Table 7: Labeling schema for "simple"/"CDF–type" event weights defined by equations 4/ 7. The left (right) column specifies the label (instance) name parameter of the InputTag which you need to pass to the edm::Event::getByLabel function in order to access the value of the weight for a given event (e.g. edm::InputTag("bgEstFakeRateEventWeights", "qcdMuEnriched")).

Prefix	Suffix
bgEstFakeRateJetWeight_	qcdDiJetLeadJet qcdDiJetSecondLeadJet qcdMuEnriched WplusJets

Table 8: Labeling schema "simple"/"CDF–type" per–jet weights defined by equations 3/ 6. The argument which you need to pass to the pat::Tau::efficiency function in order to access the values of per–jet weights is given by concatenation of the left and right columns (e.g. patTau→("bgEstFakeRateJetWeight_qcdMuEnriched")).

```
               << patTau->efficiency("frByStandardChainMuEnrichedQCDsim").value()
               << ")" << std::endl;
  }
```

The names of the InputTags which you need to use in order to access the event weights computed for different sets of fake–rate probabilities described in section 5.2.2 are composed as detailed in table 7.

Table 8 describes the schema how the string type arguments which need to be passed to the pat::Tau::efficiency function are composed.

For debugging purposes it may be useful for you to access the "raw" tau identification and fake–rate probabilities taken from the TH3s, before the values are passed to the *bgEstFakeRateJetWeights* module. Per default, those probabilities are available via the pat::Tau::efficiency function. The string type arguments which need to be passed to the pat::Tau::efficiency function in order to access the "raw" tau identification and fake–rate probabilities are composed as described in table 9.

Prefix	Discriminator	Sample
eff	ByStandardChain	ZTTsim
fr	ByStandardChain	DiJetHighPtsim DiJetSecondPtsim MuEnrichedQCDsim WJetssim

Table 9: Labeling schema for tau identification efficiencies and fake–rates entering the computation of the per–jet weights defined by equations 3/ 6. You can access the tau identification efficiency and fake–rate values by calling the pat::Tau::efficiency function, passing as argument the concatenation of the "Prefix" with the "Discriminator" and the "Sample" column (e.g. patTau→("effByStandardChainZTTsim")).

References

[1] M. Bachtis *et al.*, "Particle-flow reconstruction of 0.9 TeV and 2.36 TeV collision events in CMS", CMS AN–2010/031

[2] S. Gennai *et al.*,"Tau reconstruction and identification with particle-flow techniques using the CMS detector at LHC", CMS AN–2008/043

[3] J. Conway, E. Friis, C. Veelken, "Tau identification using the Tau Neural Classifier Algorithm", CMS AN in preparation

[4] S. Bagliesi et al., "Tau production from Z decays in pp collisions at $\sqrt{s} = 7$ TeV", EWK–09–015; in preparation

[5] D. Jang, "Search for MSSM Higgs decaying to τ pairs in $p\bar{p}$ collision at $\sqrt{s} = 1.96$ TeV at CDF", PhD thesis, Rutgers University, 2006; A. Abulencia *et al.* (CDF Collaboration), "Search for neutral MSSM Higgs bosons decaying to tau pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", Phys. Rev. Lett. **96** (2006) 011802; C. Cuenca Almenar, "Search for the neutral MSSM Higgs bosons in the $\tau\tau$ decay channels at CDF Run II", PhD thesis, Universitat de Valencia, 2008

[6] C. Amsler *et al.* (Particle Data Group), Physics Letters B **667**, 1 (2008)