New kinematic reconstruction for the $t\bar{t}$ events in the dilepton channel with the CMS experiment at $\sqrt{s} = 8$ TeV

Irene Sánchez Arribas, Universidad de Granada, Spain
Supervised by Ganna Dolinska
DESY-CMS group

Abstract

The efficiency of the kinematic reconstruction of the $t\bar{t}$ events in the dilepton channel with the CMS experiment in the LHC at $\sqrt{s} = 8$ GeV is analysed. The study is performed in the dilepton decay channel using data corresponding to an integrated luminosity of 19.7 fb^{-1}. In order to do this, the efficiency of finding any solution, the correct solution (well matched jets) and the correct solution but considering also the possibility of swapped jets in the kinematic reconstruction is analysed for different number of smearings of the energy of the lepton and jets, for both signal and background events. Finally, also the weight distribution is analysed for correct assigned jets and for swapped jets.

Contents

1 Introduction

1.1 Top quark pair production

The top quark is the most massive of all observed elementary particles so far. Its existence (and the botton quark one) was postulated in 1973 by Makoto Kobayashi and Toshihide Maskawa to explain the observed CP violations in kaon decay. Finally, in 1995, it was discoverd an Fermilab by the experiments CDF and DØ.

The measurements of top quark properties are used as a constraint for the Standard Model and it is a sensitive probe of New Physics. The top quark is essencial to study the higgs properties and to measure top Yukawa coupling. So, the top quark is a tool for precise SM measurements.

The top quark is the heaviest particle ever observed, with a mass of $m_t \simeq 173\,GeV$. Its lifetime is $\tau_t \simeq 5 \cdot 10^{-25}\,s$ while the typical time to create a bound top-antitop quark state is $2 \cdot 10^{-24}\,s$. Because of this, the top quark is the only quark that decays before hadronizing. In fact, the existence of the top quark is inferred by the observation of its decay products.

We can obtain a top quark pair via stron interaction in two diferent ways, by gluon-gluon fusion or quark-antiquark annihilation. The first one is the most important at the LHC energy scale ($\sim 80\%$), while the second one is the most important in the TEVATRON scale ($\sim 85\%$). Feynman diagrams for this two productions processes are shown in Figure 1.

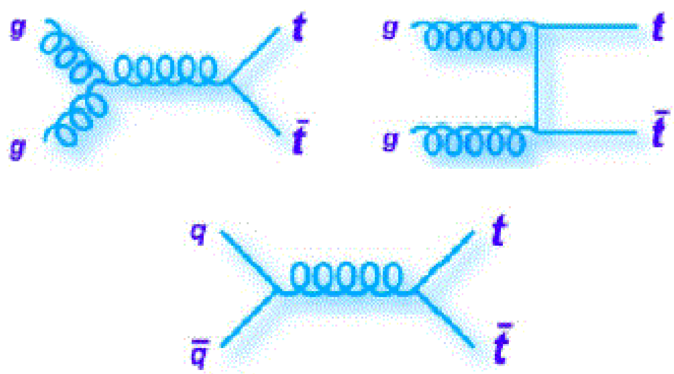

Figure 1: Top pair production

1.1.1 $t\bar{t}$ decay

The top quark decays via electroweak interaction into a charged W boson and a b quark.

$$t \rightarrow W + b$$

The b quark becomes a hadronic jet in the final state. Each W boson also decays shortly after being produced, therefore signatures of top pair production depend on the decay of the two W bosons in the event. It can decays in a hadronic or leptonic decay mode:

$$W \rightarrow qq'$$

$$W \rightarrow l\nu_l$$

However, we have to take into acount the conservation of electromagnetic charge. Because of this, the top and antitop decay is finally:

$$t \rightarrow W^+ + b$$

$$\bar{t} \rightarrow W^- + b$$

In the leptonic decay mode of the W boson, the charge of the lepton is positive if it comes from a t quark and negative if it comes from a $\bar{t}$. In order to the W decay mode, we can categorize the $t\bar{t}$ decay in three channels: all-jet channel (fully-hadronic decay), lepton-jet channel and dileptonic channel.

1. All-jet channel: it consits of two b-jets and four light jets. It occurs when the two W bosons decay hadronically ($\sim 46\%$ of the time).
2. Lepton + jet channel (semileptonic decay): in this case, in the final state there is one lepton, one neutrino, two light jets and two b-jets ($\sim 34\%$ of the time).
3. Dileptonic channel: both W bonsons decay leptonically ($\sim 6\%$ of the time), so at the end of the process two leptons, two neutrinos and two b-jets are found.

2 Kinematic Reconstruction of tt events in the dileption channel with the CMS experiment

The lepton, neutrino and b quark from the decay of the top quark (and subsequent decay of the W) have some particular kinematic properties. The lepton for example tends to be very energetic compared to leptons coming from other decays. These differences can be used to select which events are top quarks. To study top quarks scientists select candidate events that are similar to what the Standard Model would predict if a top and anti-top quark were produced in the collision. Despite care, this carefully selected sample still contains a large fraction of other types of physics events called background. The major background is when W bosons and jets are created without coming from the top quark together with other light particles. Considering the three channels, the fully hadronic channel has a huge background, the semileptonic channel a moderate one, and in the dileptonic channel the background is low. Many techniques are applied to suppress the background further until we consider that we have a cleaner sample of top events.

DESY group investigates $t\bar{t}$ production cross sections as it is a constraint of the SM. In our group, the dilepton channel is analysed. Using a kinematic reconstruction, top quark properties are measured. For this kinematic reconstruction an analytical solution of the $t\bar{t}$ dilepton equations is used. The top quark and antiquark production system in the dilepton decay is described by a set of equations which is nonlinear in the unknown neutrino momenta. As in the process there are two undetected neutrinos, we have 6 unknowns p_ν and $p_{\bar{\nu}}$, but it is possible to determinate them using 6 kinematics constraints : $W^{\pm}$and top masses, and MET (E_x^{miss} and E_y^{miss}).

These equations are:

$$\begin{aligned}
\not{E}_x &= p_{\nu_x} + p_{\bar{\nu}_x}, \\
\not{E}_y &= p_{\nu_y} + p_{\bar{\nu}_y}, \\
E_\nu^2 &= m_\nu^2 + p_{\nu_x}^2 + p_{\nu_y}^2 + p_{\nu_z}^2, \\
E_{\bar{\nu}}^2 &= m_{\bar{\nu}}^2 + p_{\bar{\nu}_x}^2 + p_{\bar{\nu}_y}^2 + p_{\bar{\nu}_z}^2, \\
m_{W^+}^2 &= (E_{\ell^+} + E_\nu)^2 - (p_{\ell_x^+} + p_{\nu_x})^2, \\
&\quad -(p_{\ell_y^+} + p_{\nu_y})^2 - (p_{\ell_z^+} + p_{\nu_z})^2, \\
m_{W^-}^2 &= (E_{\ell^-} + E_{\bar{\nu}})^2 - (p_{\ell_x^-} + p_{\bar{\nu}_x})^2, \\
&\quad -(p_{\ell_y^-} + p_{\bar{\nu}_y})^2 - (p_{\ell_z^-} + p_{\bar{\nu}_z})^2, \\
m_t^2 &= (E_b + E_{\ell^+} + E_\nu)^2 - (p_{b_x} + p_{\ell_x^+} + p_{\nu_x})^2, \\
&\quad -(p_{b_y} + p_{\ell_y^+} + p_{\nu_y})^2 - (p_{b_z} + p_{\ell_z^+} + p_{\nu_z})^2, \\
m_{\bar{t}}^2 &= (E_{\bar{b}} + E_{\ell^-} + E_{\bar{\nu}})^2 - (p_{\bar{b}_x} + p_{\ell_x^-} + p_{\bar{\nu}_x})^2, \\
&\quad -(p_{\bar{b}_y} + p_{\ell_y^-} + p_{\bar{\nu}_y})^2 - (p_{\bar{b}_z} + p_{\ell_z^-} + p_{\bar{\nu}_z})^2.
\end{aligned}$$

Figure 3: Set of equations for the kinematic reconstruction
arXiv: hep-ph/0603011

And the final equation is an univariate polynomial of degree four :

$$h_0 p_{\nu_x}^4 + h_1 p_{\nu_x}^3 + h_2 p_{\nu_x}^2 + h_3 p_{\nu_x} + h_4 = 0$$

Which can be solved analitically. The problem is that in each event there is more than one solution ($t\bar{t}$ candidate) because of the multiple solutions of kinematic equations (up to four), the several combinations of leptons and jets and other problems like the detector resolution.

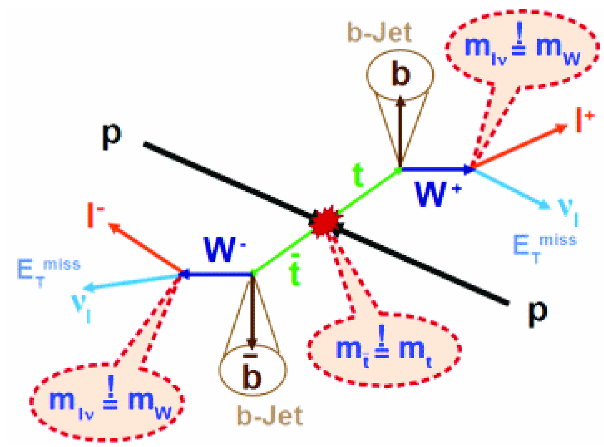

Figure 4: Top quark decay

In this kinematic reconstruction, in order to correct the detector effects, the jet and lepton energies are smeared . The resolution of this smearing is parametrized in MonteCarlo using the cuantities $\frac{E^{true}_{bquark}}{E^{reco}_{jet}}$ and $\frac{E^{true}_{lepton}}{E^{reco}_{lepton}}$. This smearing is also directional , so that the most probable energies are more times taken. The top mass is assumed and fixed to 172.5 GeV as the SM predicts. In the reconstructed level, the W mass is smeared according to the MonteCarlo W mass distribution (true level).

As it is mencioned before, there are more than one solution. As a selection method, only the lepton-jet combination with the largest sum (over the number of smearings) of solution weights $\sum w_i$ (where w_i is the weight for the $i - smearing$) according to the true m_{bl} spectrum and with the smallest $m_{t\bar{t}}$ is taken. Finally, for all smeared points, a weighted average solution is taken. For example, in case of the transverse momentum of the top it is weighted as it follows:

$$p_T^{top} = \frac{1}{\sum_i^{N_{sm}} w_i} \sum_i^{N_{sm}} w_i p_{T_i}^{top}$$

3 Events selection

Events are selected if there are two isolated leptons (only electrons and muons are considered as signal) of opposite charge and at least two jets, one of them, identified as a b-jet (at least one b-tagged jet). Leptons are requierd to be within the pseudorapidity interval $|\eta| < 2.4$, to have a transverse momentum of at least 20 GeV , an invariant mass in the ee channel or$\mu\mu$ channel greater than 20 GeV in order to supress constributions from heavy flavour resonance decay events, and $|M(Z) - M(ee, \mu\mu)| > 15$ GeV for avoiding contributions from the Drell-Yan process, which dominates the ee and $\mu\mu$ channels. The kinematic reconstruction of the $t\bar{t}$ system is required to assign the jets coming from the $t\bar{t}$ pair and therefore identify the additional jet. Jets are required to have a transversal momentum $p_T \geq 30$ GeV and $|\eta| < 2.4$. The transversal missing energy has to be larger than 40 GeV but only for ee and $\mu\mu$ channels.

4 Variation of the number of smearings in Kin Reco

It is important to see how the variation of the number of smearings of the lepton and jet energies affects to the efficiency of the kinematic reconstruction of finding any solution and how much time the program uses while processing. In previous simulations, there were taken 100 smearings. In figure 5, figure 6, figure 7 and figure 8, it is shown the Kin Reco eff. of finding any solution for the signal sample for 3, 70, 100 and 10000 smearings. This efficiency is calculated in the following way:

$$Eff.^{correct-sol} = \frac{N^{events}_{after\,KinReco}}{N^{events}_{before\,KinReco}}$$

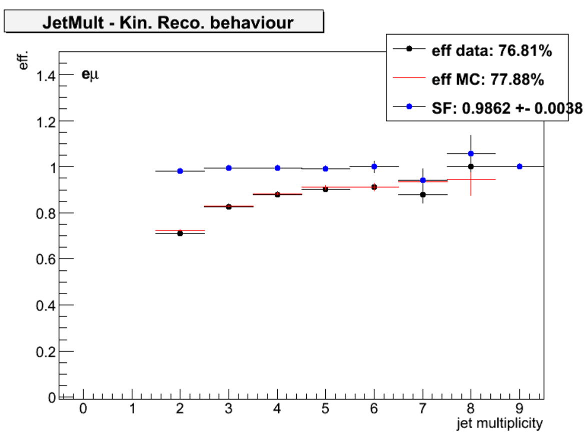

Figure 5: $Eff.^{Kin\,Reco}$ vs jet multiplicity for 3 sm

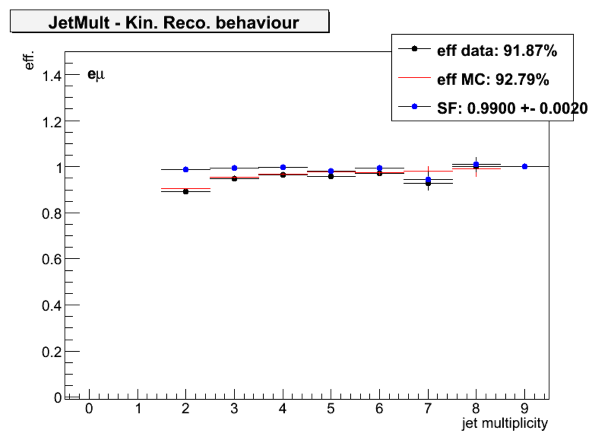

Figure 6: $Eff.^{Kin\,Reco}$ vs jet multiplicity for 70 sm

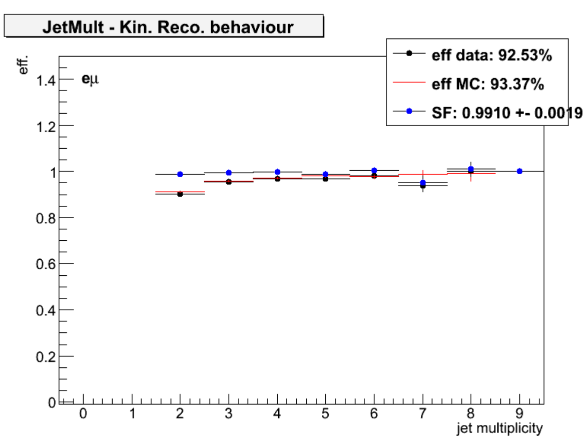

Figure 7: $Eff.^{Kin\,Reco}$ vs jet multiplicity for 100 sm

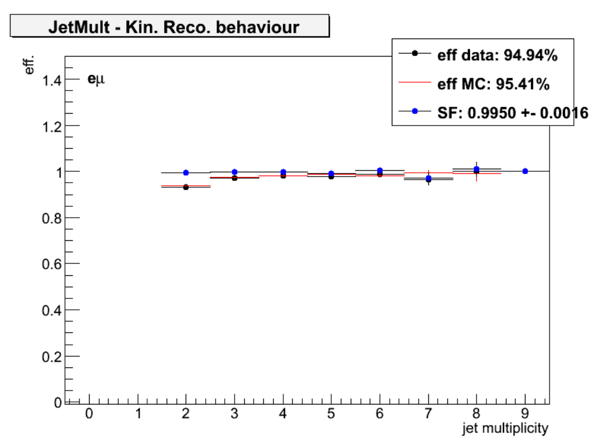

Figure 8: $Eff.^{Kin\,Reco}$ vs jet multiplicity for 10000 sm

It is seen that the efficiency to find any solution increases with more smearings. Another way to check Kin Reco efficiency is to look at the RMS of p_T^{top}. The smaller is the RMS, the better the Kin Reco efficiency is.

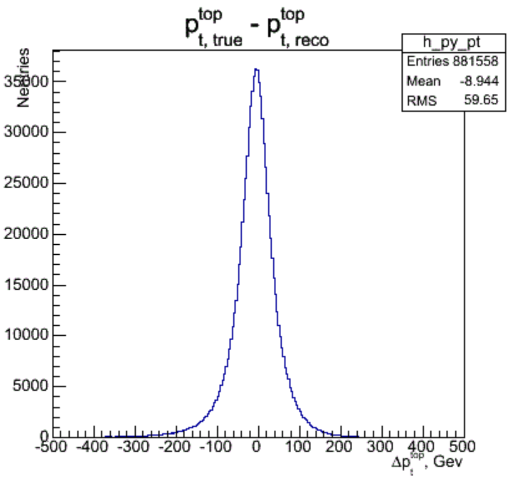

Figure 9: RMS of the p_T^{top} distribution

In figure 10, the RMS of the p_T^{top} distribution is shown as a measure of Kin Reco efficiency for different number of smearings (1, 3, 10, 30, 70, 100, 300,700, 5000 and 10000 smearings) and also the processing time in figure 11:

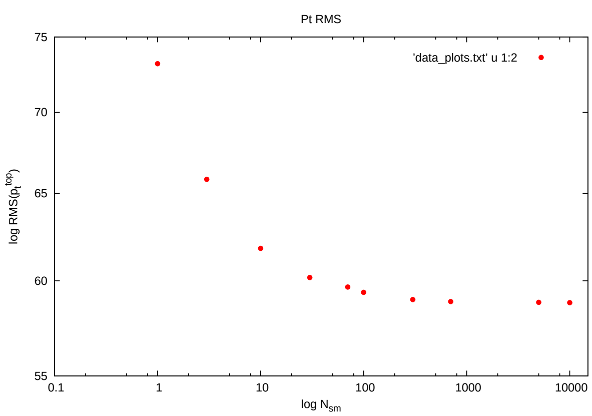

Figure 10: RMS of the p_T^{top} distribution vs N_{sm}

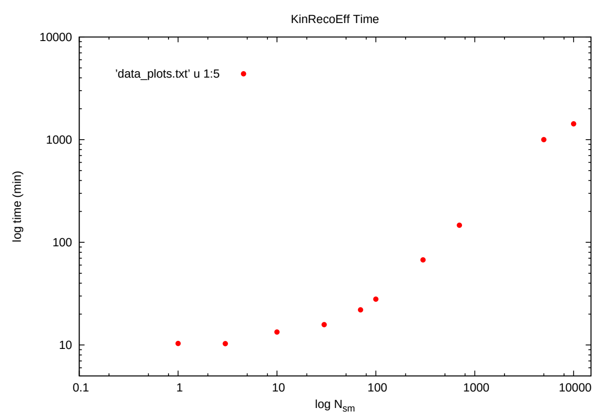

Figure 11: time (min) vs N_{sm}

As it is shown in the plots, from 700 to 10000 there is an almost flat region considering the RMS. Nevertheless, the processing time increases a lot from 100 smearings to 10000 smearings. In the other hand, it is also interesting to see how changes Kin Reco Efficiency not just for the signal, but also for background events. This is shown in the next four figures for 3, 70, 100 and 10000 smearings:

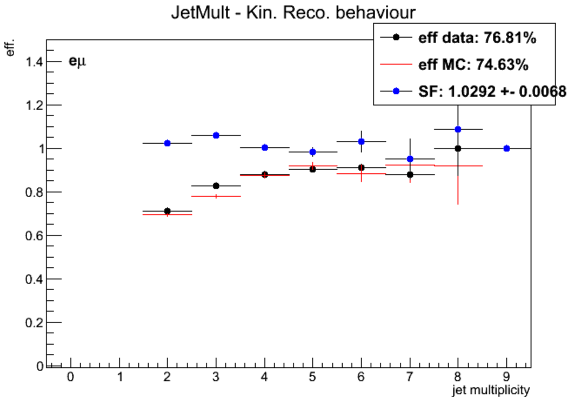

Figure 12: $Eff._{bg}^{Kin\,Reco}$
vs jet multiplicity for 3 sm

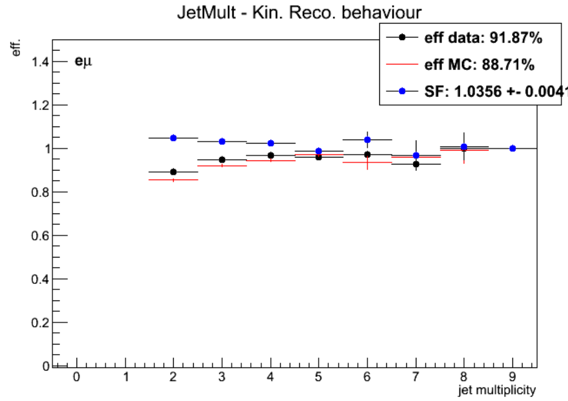

Figure 13: $Eff._{bg}^{Kin\,Reco}$
vs jet multiplicity for 70 sm

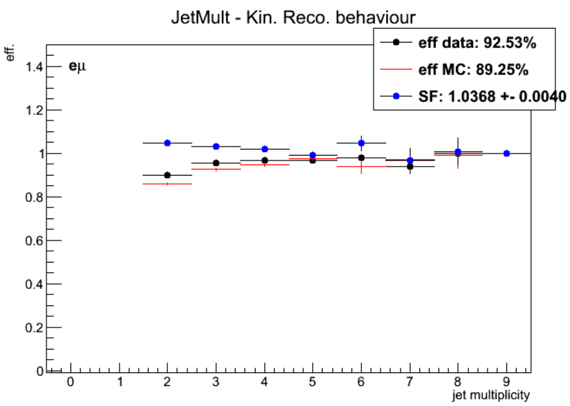

Figure 14: $Eff._{bg}^{Kin\,Reco}$
vs jet multiplicity for 100 sm

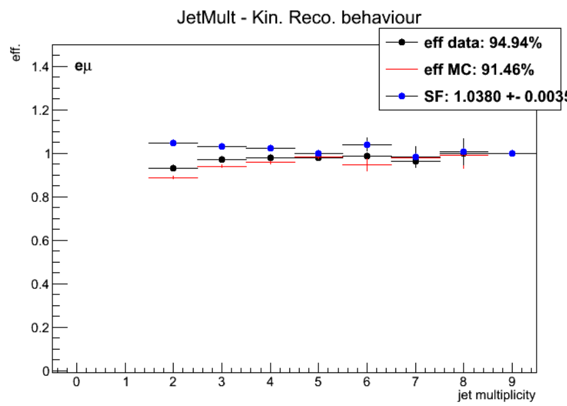

Figure 15: $Eff._{bg}^{Kin\,Reco}$
vs jet multiplicity for 10000 sm

In this case, also the backgound increases when the number of smearings increases. For further analysis, since the efficiency to find any solution increases with more smearings in signal and background to very high numbers it would be also interesting to see how changes the efficiency of finding the correct solution for well matched b-lepton pairs (b - antilepton and $\bar{b}$ - lepton) for different number of smearings in relation to the already taken 100 smearings. This efficiency is calculated as it follows :

$$Eff.^{correct-sol} = \frac{N_{after\,KinReco}^{events}\,give\,correct\,sol}{N_{after\,KinReco}^{events}}$$

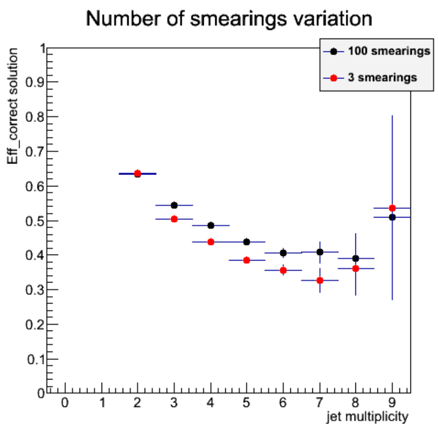

Figure 16: $Eff._{correct-sol}^{Kin\,Reco}$ vs jet multiplicity for 3 sm

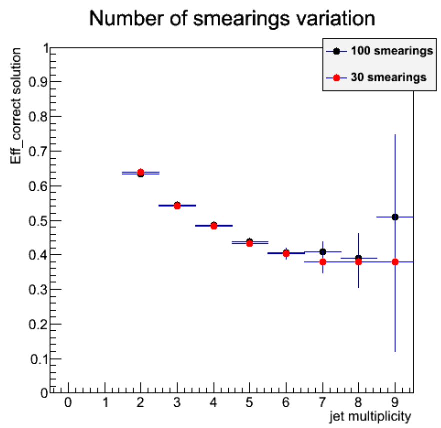

Figure 17:$Eff._{correct-sol}^{Kin\,Reco}$ vs jet multiplicity for 30 sm

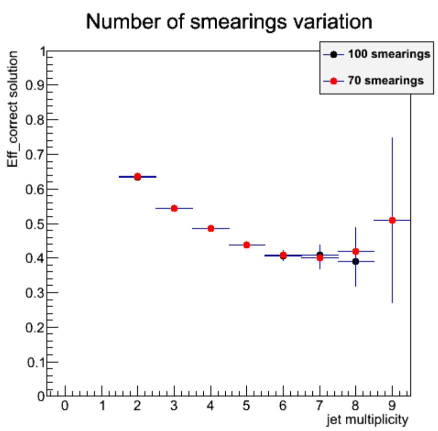

Figure 18: $Eff._{correct-sol}^{Kin\,Reco}$ vs jet multiplicity for 70 sm

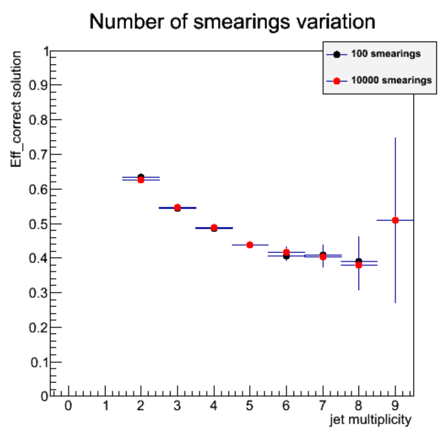

Figure 19: $Eff._{correct-sol}^{Kin\,Reco}$ vs jet multiplicity for 10000 sm

It is seen that the efficiency of finding the correct solution and correct assigned jets is almost at maximum at 30 smearings and later at 70 smearings it is stable at the maximum. Also, for the ttH analysis, it is interesting to see the efficiency of finding the correct solution (well matched jets) or correct solution with swapped jets:

$$Eff.^{correct-sol} = \frac{N_{after\,KinReco}^{events}\, give\, correct\, sol\, and\, swapped\, jets}{N_{after\,KinReco}^{events}}$$

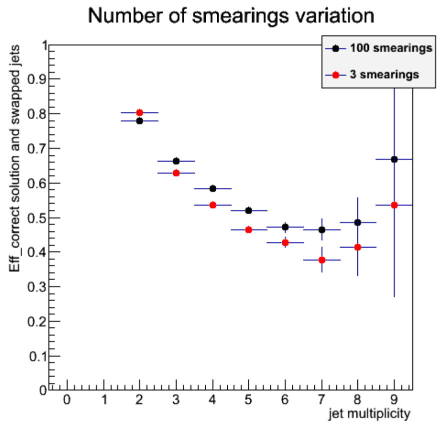

Figure 20: $Eff._{correct-sol+swapped\,jets}^{Kin\,Reco}$ vs jet multiplicity for 3 sm

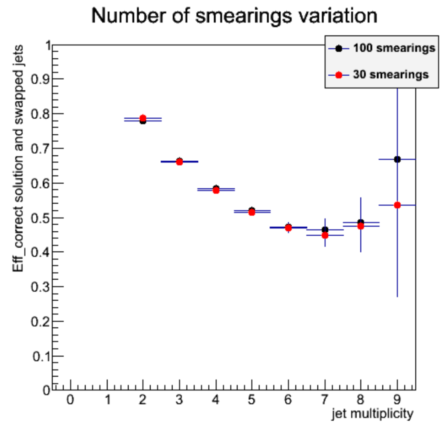

Figure 21:$Eff._{correct-sol+swapped\,jets}^{Kin\,Reco}$ vs jet multiplicity for 30 sm

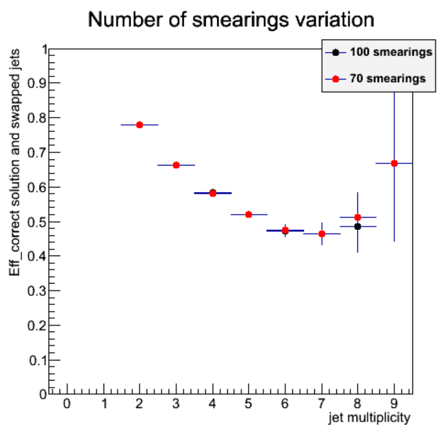

Figure 22:$Eff._{correct-sol+swapped\,jets}^{Kin\,Reco}$ vs jet multiplicity for 70 sm

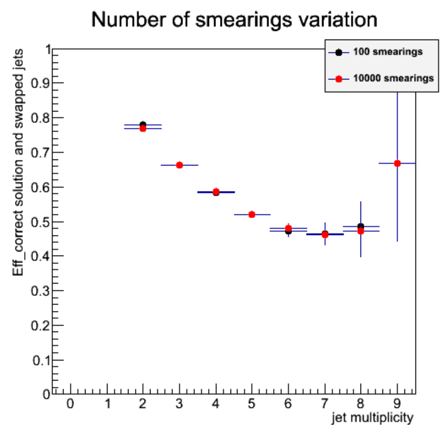

Figure 23: $Eff._{correct-sol+swapped\,jets}^{Kin\,Reco}$ vs jet multiplicity for 10000 sm

5 Weight control plots

As it is explained before, only the lepton-jet combination with the largest sum of solution weights according to the true m_{bl} spectrum and the weighted average solution are taken. For this reason, studying the weight distribution is an important fact in the kinematic reconstruction. The weight of the $i - smearing$ is called w_i, and it is obteined as it follows:

$$w_i = w_{b\bar{l}} w_{\bar{b}l}$$

Where both $w_{\bar{b}l}$ and $w_{b\bar{l}}$ are taken according to the m_{bl} spectrum. In the figures above, there are shown the weight distribution for the reconstructed weight for any solution (the one that is taken for the kinematic reconstruction), the weight distribution for correct solution and well matched jets and the weight distribution for correct solution and swapped jets:

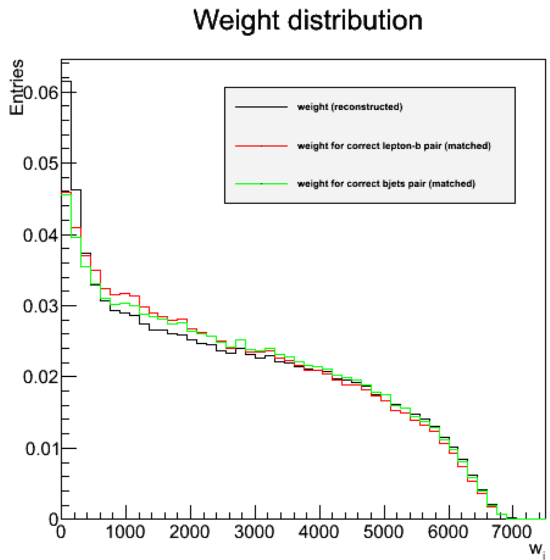

Figure 24: Weight distribution

In the next figure is shown the control plot for the weight distribution:

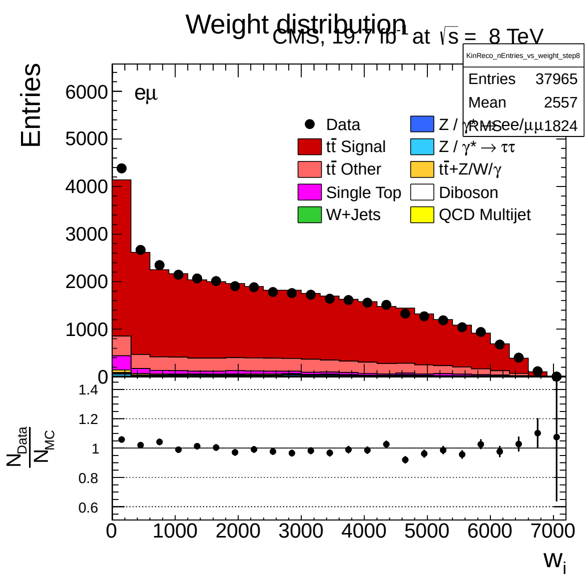

Figure 25: Weight control plot

The black points are the Monte Carlo Data , so that it is seen that they fit very well to the $t\bar{t}$ signal.

6 Conclusions

It is interesting to point out, that the efficiency for finding the correct solution (or the correct two jets) is almost at maximum after 30 smearings, and later at 70 smearings it is stable at the maximum. Since this is the relative fraction, and it is shown that the number of solved events increases continuously with the smearings, in fact it is ganed more events with correct solutions with more smearings. So the only disadvantage of using more smearings is that the background increases, but this is a small effect since the backgrounds are small. The other disadvantage is of course the processing time,but it could be possible to run tests with ~70 smearings, and final results with a huge number.

For further studies and improvements of the kinematic reconstruction, it could be interesting to analyze why some events are not reconstructed: if they are outside the kinematic acceptence or not, and if it is not, try to find out why they are not reconstructed and the kinematic reconstruction fails.

References

[1] CERN Home Page http://home.web.cern.ch/topics/large-hadron-collider

[2] CMS Home Page http://cms.web.cern.ch/

[3] Korol Ievgen, *"Kinematic reconstruction of ttbar events in the dilepton channel with the CMS experiment"*.

[4] http://summerstudents.desy.de/hamburg/e177730/e178988/top_physics_210814.pdf

[5] Lars Sonnenschein, *"Analytical solution of ttbar dilepton equations"* arXiv: hep-ph/0603011

[6] http://conferences.fnal.gov/lp2003/forthepublic/topquark/