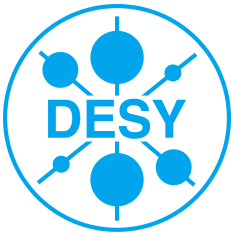

Measurement of the Top Quark Mass from $t\bar{t}$ Events with Additional Jets

María Vieites Díaz[1]
Supervised by Carmen Diez Pardos

[1] *Universidade de Santiago de Compostela, Spain*

Summer student work report

06 September 2013

Abstract

A measurement of the top quark mass from $t\bar{t}$ events with additional jets is performed in pp collisions at $\sqrt(s) = 8$ TeV with the CMS detector using data recorded in 2012 corresponding to an integrated luminosity of 19.6 fb^{-1}. The measurements are performed in the dilepton decay channels ($e^+e^-, \mu^+\mu^-$ and $e^\pm\mu^\mp$) of the top-antitop quark pairs. The mass is extracted from the normalised differential top antitop quark cross section as a function of the invariant mass of the $t\bar{t} + jet$ system.

Contents

1 Introduction

The top quark is the heaviest elementary particle observed so far, and a precise measurement of its mass can be used as a constraint for some of the Standard Model parameters. The mass of the top quark has been measured precisely by the Tevatron and LHC experiments, and the current world average is 173.20±0.87 GeV.

In proton-proton collisions the top quarks are mostly produced in pairs $t\bar{t}$ from gluon or quark fusion processes, while they decay most of the time into a b quark and a W boson. Therefore the $t\bar{t}$ events can be catergorized into three channels according to the decay of the W boson (leptonically or hadronically): lepton+jet channel (BR∼30%, only considering electrons and muons), the dilepton channel (BR∼5%) and fully hadronic channel (BR∼44%). For the dilepton channel, the presence of the two leptons in the final state results in a high purity selection, leading this channel to be the one with the largest signal-to-background ratio.

In this report, the results obtained for the top quark mass extracted from the normalised differencial cross secction measured as a fuction of an observable which is proportional to the inverse of the invariant mass of the $t\bar{t} + jet$ system are presented. The measurement is done using the 19.6 fb^{-1} recorded by the CMS detector in 2012 [2].

The observable used in this analysis has been proposed in [1] and it is defined as

$$\rho_s = \frac{2m_0}{\sqrt{s_{t\bar{t}j}}},$$

where m_0 is a constant close to the top quark mass and $\sqrt{s_{t\bar{t}j}}$ is the invariant mass of the $t\bar{t} + jet$ system. In the reference used m_0 has an arbitrarily fixed value of 170 GeV. As it is a funcion of $\sqrt{s_{t\bar{t}j}}$ the whole event needs to be reconstructed to assign correctly the jets from the decay of the tops and identify the additional jets. The measure is not statistically limited due to the large number of $t\bar{t}$ events produced with additional ISR/FSR (∼50%) at LHC energies.

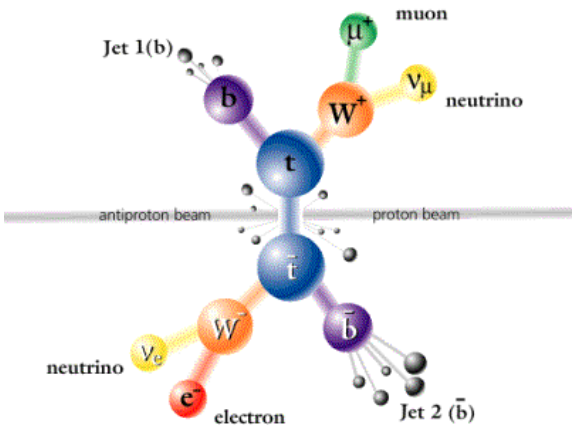

Figure 1: Scheme of a $t\bar{t}$ decaying via the dilepton channel.

2 Event Selection and Kinematic Reconstruction

The events are reconstructed using a particle-flow technique, in which signals from all subdetectors are combined. Charged hadron candidates from pileup events, i.e. originating from a vertex other than the one of the hard interaction, are removed before jet clustering on an event by event basis. Subsequently, the remaining component from neutral particle candidates from pileup events is subtracted at the level of a jet energy correction. The missing transverse energy (E_T^{miss}) is defined as the magnitude of the imbalance of the transverse momentum of all reconstructed particle candidates.

The muons are required to have a transverse momentum $p_T > 20$ GeV within a pseudorapidity region $|\eta| < 2.4$ and are required to be isolated with $I_{rel} < 0.15$. I_{rel} is defined as the sum of the transverse momenta of all neutral and charged reconstructed particle candidates, except the muon itself, inside a cone in $\eta - \phi$ space of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$, divided by the muon transverse momentum.

Electron candidates are required to be within the pseudorapidity interval $|\eta| < 2.4$, to have a transverse energy of at least 20 GeV and fulfil a relative isolation $I_{rel} < 0.15$. Electrons from identified photon conversions are rejected.

Jets are reconstructed by clustering the particle-flow candidates, using the anti-k_T clustering algorithm with size parameter R = 0.5. They are selected in the pseudorapidity interval $|\eta| < 2.4$ and with a transverse momentum of at least 30 GeV. Jets originating from bottom quarks are identified from combined secondary vertex and track-based lifetime information. The b-tagging efficiency of the selected algorithm is about 80-85% and the mis-tag rate around 10%.

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. As the ee and $\mu\mu$ channels are highly dominated by Drell-Yan proceses, events with a lepton pair invariant mass inside a $\pm$15GeV window around the Z-boson mass are removed and E_T^{miss} in the event is required to be larger than 40 GeV. In order to supress contributions from heavy flavour resonance decays events with $M_{ll} < 20$GeV are also removed.

The fully kinematic reconstruction of the $t\bar{t}$ system is required to assign the jets coming from the $t\bar{t}$ pair and thus identify the additional jet. The system of equations obtained is underconstrained due to the presence of the neutrinos so some kinematic constraints are applied: W mass is assumed, top and antitop quarks are required to have the same mass and E_T^{miss} has to be balanced by the two neutrinos of the decay. After this process up to four possible solutions could remain. The ones with more b-tagged jets and with the best reconstructed neutrino energy with respect to Monte Carlo generated spectrum are prefered.

Events are finally selected if they have one additional jet (this meaning not being selected for the kinematic reconstruction) fulfilling a p_T cut (p_T >50GeV) and being in the $|\eta| < 2.4$ region.

3 Event Yield and Control Distributions

The number of observed events, compared to the expected signal and background events before and after requiring the extra jet in the event are shown in Table 1. The requirement of the additional jet reduces the statistics by a factor four; however, the statistical uncertainty remains low. Prompt dileptonic decays are refered to as $t\bar{t}$ *signal*, while other allowed decay channels of the $t\bar{t}$ system (W decaying hadronically or via τ leptons) are named $t\bar{t}$ *BG*. The rest of the background contributions (single top, diboson decays, Drell-Yan, etc) are refered as *BG other*. From the table it can be seen that the dominant backgrounds are the other $t\bar{t}$ allowed decays. Only the contribution from Drell-Yan processes is estimated from data, all other background contributions are estimated from MC simulation.

Figure 2 presents the invariant mass of the $t\bar{t}$ system after the kinematic reconstruction is performed. In this figure, data are compared with the MC predictions generated with a mass of 172.5 GeV. The results are shown for each channel and for the combination (sum) of the three of them. In the lower part of each figure ratios of data over MC are presented. A reasonable agreement is observed, specially in the $e\mu$ channel, which is the best described by the MC. Control distibutions for the additional jet p_T and the observable used are shown in Figures 3 and 4. In general, MC describes data, being the $e\mu$ channel the best described, as mentioned above.

Table 1: Number of expected signal and BG events compared to the event yields for the three dilepton channels. On the left side, just after the kinematic reconstruction, when the extra jet is not required yet; on the right, after this jet has been required.

	After kin.reco.			After Extra jet required		
	$\mu\mu$	ee	$e\mu$	$\mu\mu$	ee	$e\mu$
Data	13106	9697	33618	3790	2790	9393
$t\bar{t}$ signal	7995	6195	25625	2818	2252	7771
$t\bar{t}$ BG	1705	1279	4753	529	420	1418
Other BG	1970	1562	1922	412	424	522

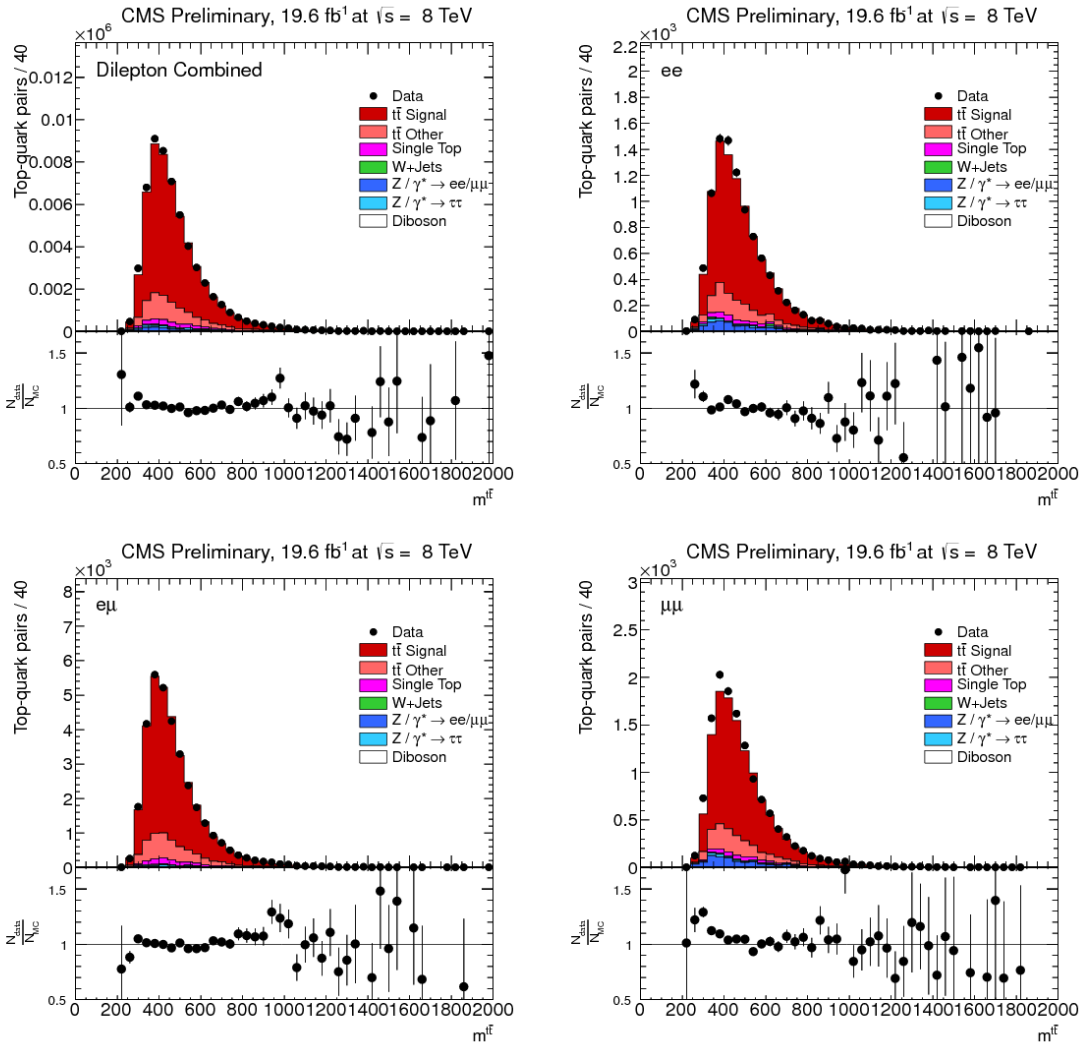

Figure 2: Invariant mass of the $t\bar{t}$ system for the three channels: ee (top right), $e\mu$ (bottom left) and $\mu\mu$ (bottom right). The combination of the three of them is shown on the top left figure. The error in data correspond to the statistical uncertainty.

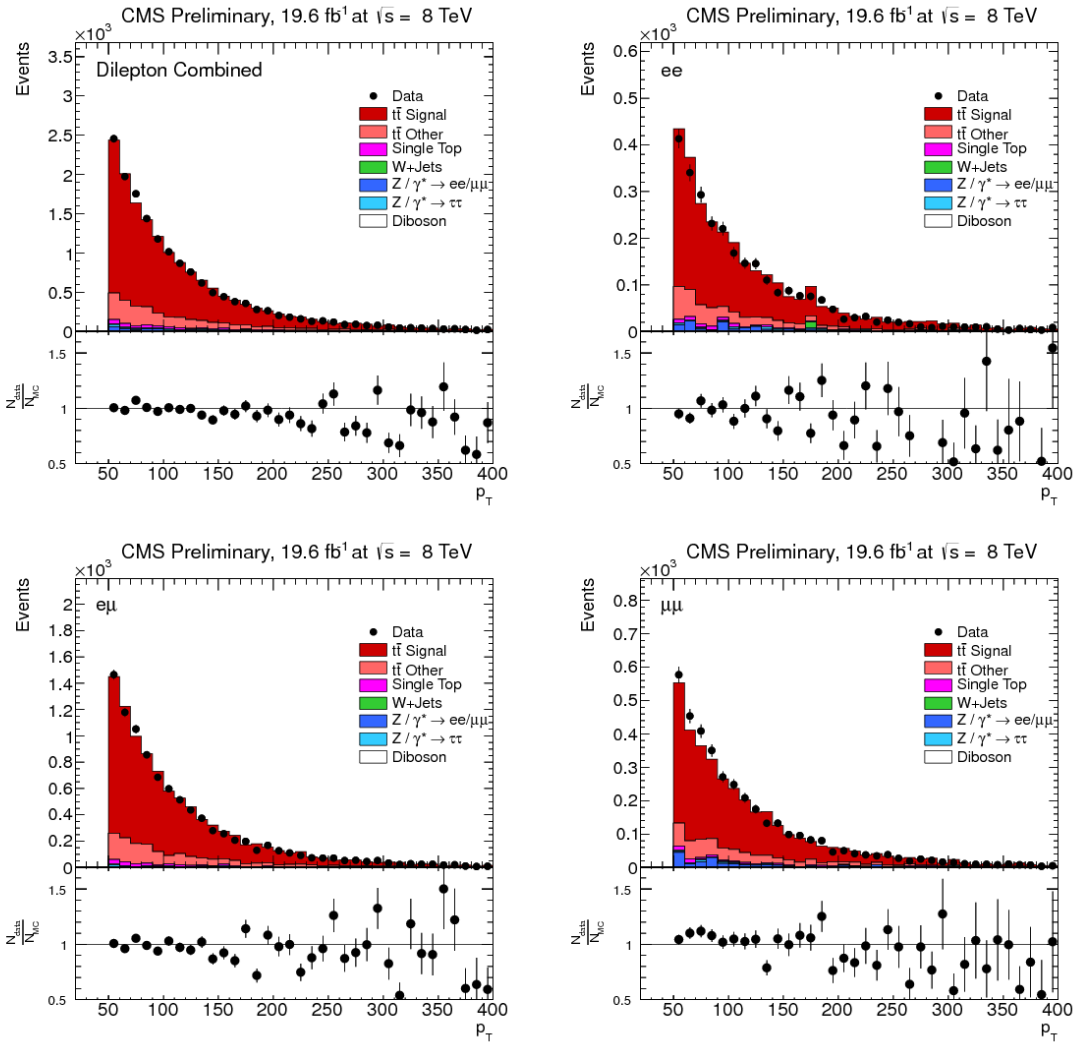

Figure 3: Extra jet p_T distributions. This jet is defined as the leading additional jet with $p_T > 50$ GeV not selected by the kinematic reconstruction. From top to bottom and left to right: all channel combined, ee, $e\mu$ and $\mu\mu$. The error in data correspond to the statistical uncertainty.

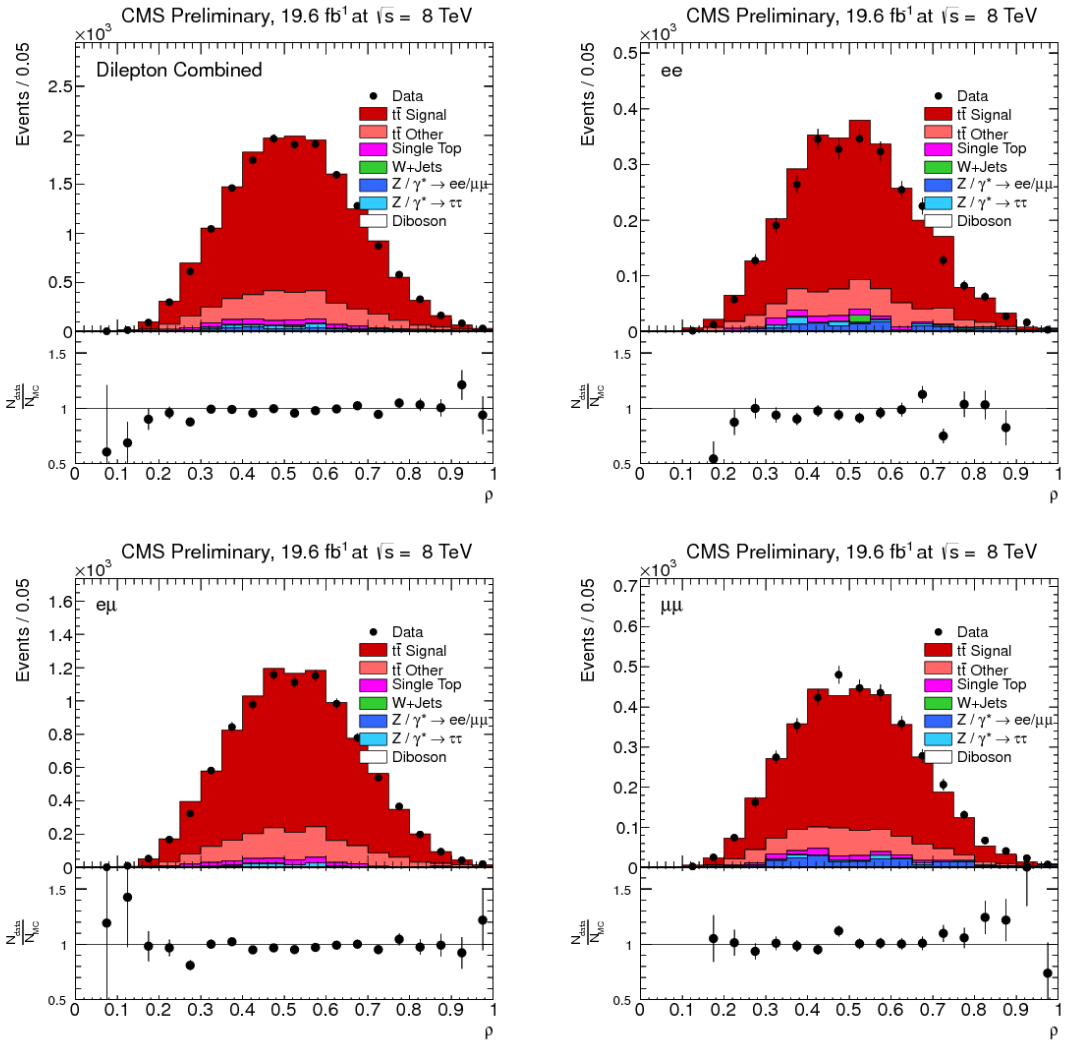

Figure 4: Distributions of the ρ_s observable, calculated with $m_0 = 170$ GeV (used in the reference). The results obtained with the combination of the three channels are shown in the top left figure; results for each channel are also shown: ee (top right), $e\mu$ (bottom left) and $\mu\mu$ (bottom righ). Ratios between data and the expected values are also presented. The error in data correspond to the statistical uncertainty.

4 Measurement of the Normalised Differential Cross Section

The normalised differential cross section as a function of ρ_s is defined as

$$\frac{1}{\sigma}\frac{d\sigma^i}{d\rho_s} = \frac{1}{\sigma}\frac{N^i_{Data} - N^i_{Bkg}}{\varepsilon^i \Delta^i_\chi \mathcal{L}}$$

where σ is the total cross section measured in the same phase space, N^i_{data} is the number of events in data in bin i, N^i_{bkg} is the number of estimated background events, $\mathcal{L}$ is the integrated luminosity, Δ^i_χ is the bin width and ε^i is the factor to correct for detector efficiencies and acceptances in each bin i of the measurement. Due to the normalisation, those systematic uncertainties that are correlated across all bins of the measurement cancel out and only shape uncertainties contribute to the total error.

4.1 Bin-to-bin Migration and Unfolding

Due to finite experimental resolution migrations of events across bin boundaries can occur. This means that events can be produced (generated) in one bin and measured (reconstructed) in another one. These effects are studied in terms of the purtity and stability values in each bin, where these two quantities are defined as follows:

- ◇ Purity: sensitive to migrations into the bin, is the number of particles generated and correctly reconstructed in a certain bin i divided by the total number of reconstructed particles in the same bin
$$p^i = \frac{N^i_{rec\&gen}}{N^i_{rec}}.$$
- ◇ Stability: sensitive to migrations out of the bin, is the number of particles generated and correctly reconstructed in a certain bin i divided by the total number of generated particles in the same bin
$$s^i = \frac{N^i_{rec\&gen}}{N^i_{gen}}.$$

In this analysis the bin width has been choosen such as both, purity and stability, are greater than 40%. In Figure 5 the purity and stability values in each bin before and after the optimisation of the bin width was performed are shown.

In order to correct for these bin migrations effects and efficiency, a regularized unfolding method is applied using the correlation matrix obtained from simulated $t\bar{t}$ events (shown in Figure 5 bottom). A detailed description of this method can be seen in [3].

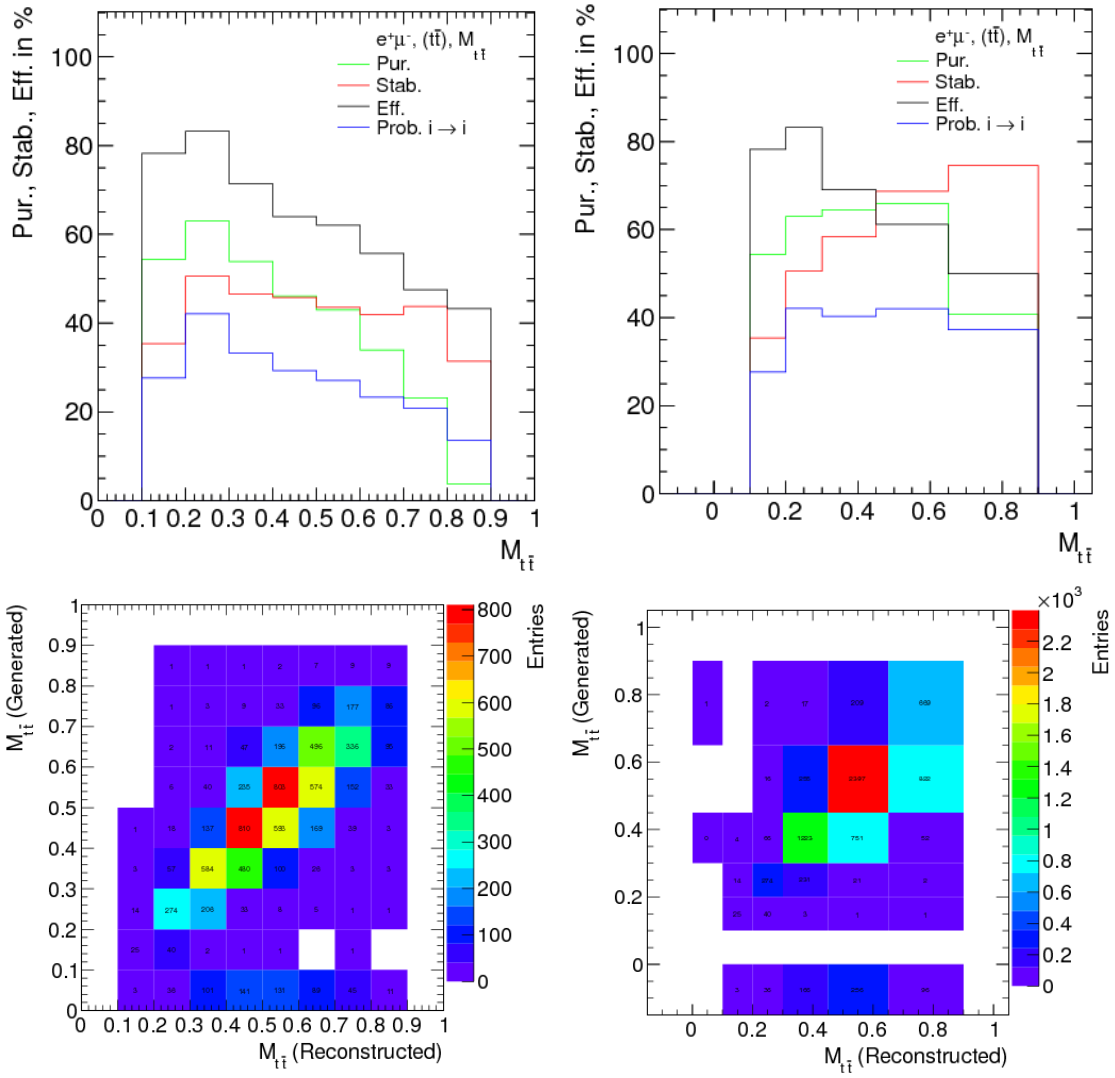

Figure 5: In the upper figures purity and stability before and after the rebining are shown. This rebinning reduces the non diagonal contributions in the correlation matrix, as it is shown by the lower figures, in which both correlation matrices are presented.

4.2 Differential Cross Section: Results

Once bin migrations efects, to correct to particle level, are taken into account, the cross section can be calculated. The reference sample used to extract the cross sections is generated with a top quark mass of 172.5 GeV. In Figure 6 the obtained distributions for each channel and its combination are shown, as well as the predicted values for different top quark masses (from 163.5 to 181.5 GeV in steps of 3 GeV). The distribution is sensitive to mass variations in the region around 0.4 and the region with $\rho_s \geq 0.7$, as expected acording to the reference [1]. Since the latter is more precise, the mass extraction is performed using the results of the last bin.

Different sources of systematic uncertainties are considered, arising from detector effects as well as theoretical uncertainties. Each systematic uncertainty is determined individually in each bin of the measurement, by variation of the corresponding efficiency, resolution, or scale within its uncertainty. For each variation, the measured normalised differential cross section is recalculated and the difference with respect to the nominal result is taken as systematic uncertainty. The overall uncertainty on the measurement is then derived by adding all contributions in quadrature. Theoretical uncertainties are originated due to some assumptions performed on the hadronisation process, the top quark mass considered and other assumptions related to the Q^2 and the matching scales. The experimental sources considered are jet energy scale (JES) and jet energy resolution (JER), background normalisation, b-tag efficiency, pileup modeling and the kinematic reconstruction.

5 Extraction of the Top Quark Mass

As explained before, by using samples generated with variated top quark masses, different values of the cross section were predicted. The top quark mass can be obtained by the comparison between the linear fit of the theory predictions and the cross section measurement. An example of this fit for the combination of all channels can be seen in Figure 5, where data is shown with its statistical and total errors. The top quark mass is extracted for each channel and for the combination of them, obtaining similar results (see Table 2). The most precise measurement is obtained from the $e\mu$ channel. The result is dominated by the systematic uncertainties, being the most relevant the ones coming from the theoretical models. The largest experimental contributions to the systematical uncertainties come from Jet Energy Scale and Resolution. A summary of them is shown in Table 3.

Table 2: Top quark mass results obtained for each channel and for the sum of them. Statistical, systematic and total uncertainties are shown in the table.

Channel	Mass (GeV)	Stat. (GeV)	Syst. (GeV)	Total Uncert. (GeV)
Combined	175.0	1.5	5.8	6.0
ee	191.0	4.1	8.6	9.5
$e\mu$	173.0	2.0	4.3	4.7
$\mu\mu$	171.0	3.0	10.6	11.0

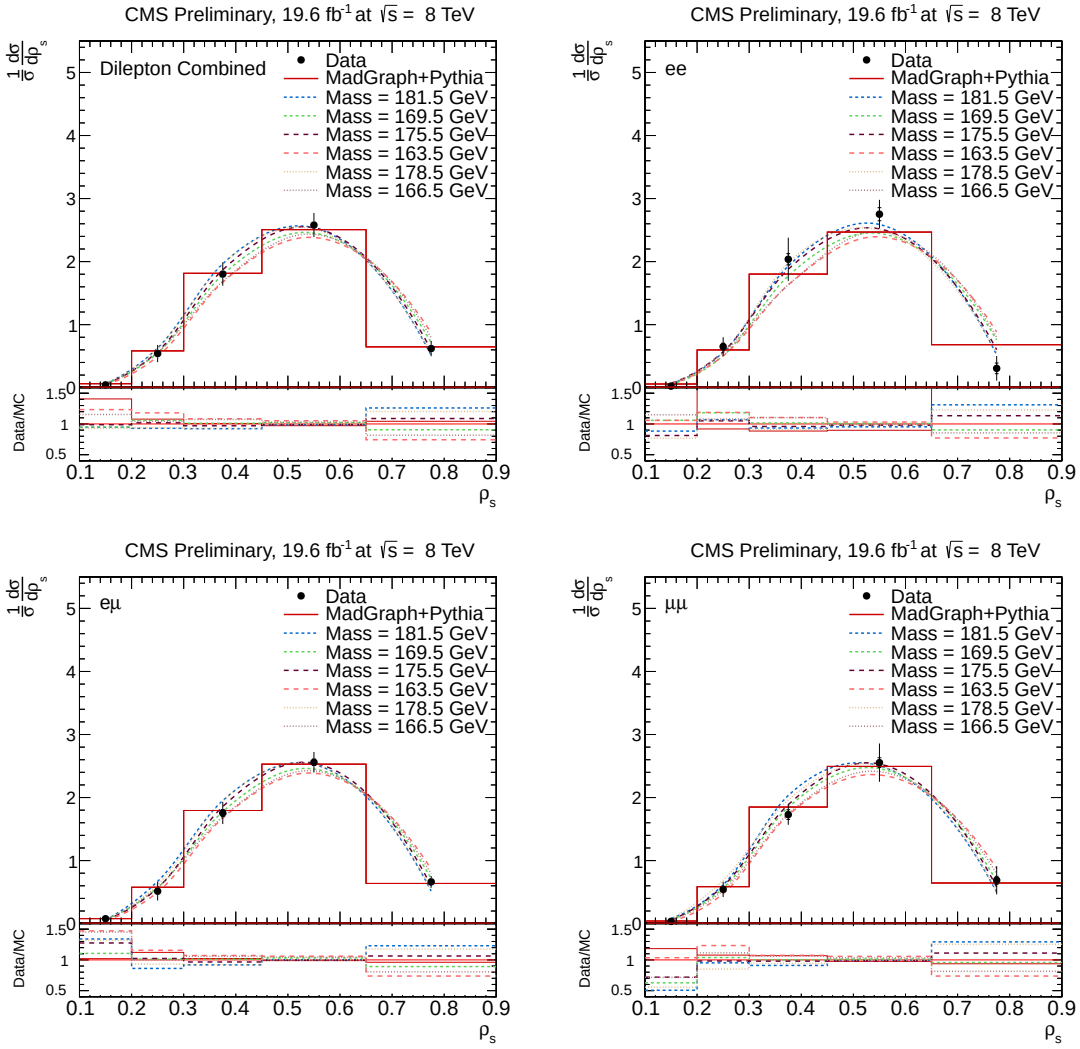

Figure 6: Normalised Differential Cross Section measured for each channel and for the sum (dilepton combined) of the three of them. Data are shown as black dots and the expected cross section value for the nominal value of the top quark mass ($m_{top}^{Nom} = 172.5$ GeV) as red histogram. Dotted lines show the expected cross section values for predictions with different top quark masses (varied between 163.5 and 181.5 GeV in steps of 3 GeV). In the lower part of each plot the ratios of the data over the different MC samples are shown. The errors on the data points indicate the statistical (inner bars) and the total uncertainty

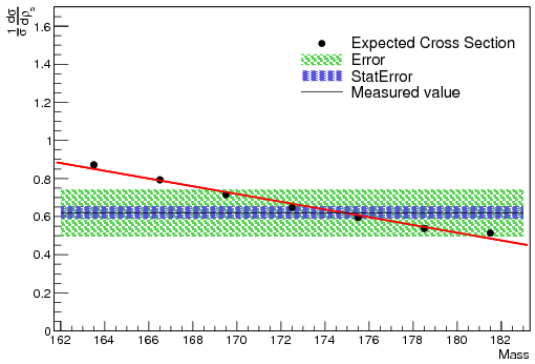

Figure 7: Normalised Differential Cross Section in the bin ρ_s >0.7 as a function of the top quark mass used to generate the simulated sample. Black dots correspond to the expected cross section values for each variaed mass, the red line is the linear fit of them and the black line is the measured value. Blue and green bands represent the statistical and total uncertainties of the measurement.

Table 3: Summary of the systematic uncertainties considered in the analysis

source	**uncertainty (%)**			
	$\mu\mu$	ee	μe	combined
Hadronisation	**3.0**	**1.0**	**1.7**	**1.9**
ME/PS thresh.	**3.5**	**1.8**	**1.1**	**1.8**
Q^2 Scale	**3.5**	**1.4**	**1.1**	**1.7**
B-tagging	0.1	<0.1	0.1	0.1
Kinematic fit	0.4	0.2	0.4	0.3
Trigger/Lepton selection	<0.1	0.4	<0.1	0.1
Backgrounds	0.1	3.5	0.3	0.7
Pile-up	0.3	0.3	<0.1	0.1
Jet energy resolution	0.3	1.2	0.3	0.4
Jet energy scale	1.3	0.6	0.7	0.8

6 Further Studies

In order to prove that the results presented here are independent from the input variables of the method used some further studies were performed. Results were found to be independent (within statistical uncertainties) on small variations of the top quark mass used as reference for the measurement and the m_0 value. As in the kinematic reconstruction of the events the MC generated neutrino energy spectrum is used, it was also checked that the differences in this variable are small enough to have a relevant effect on the result.

6.1 Effect of the reference top quark mass in MC

The differential cross section distributions obtained with variated top quark masses were studied by repeating the measurement with simulated samples generated with different top

masses (from 166.5 to 178.5 GeV in stepts of 3 GeV) instead of the nominal sample with m =172.5 GeV. 8 shows the cross section results for two of these variations, in particular 169.5 GeV and 175.5 GeV, compared to the nominal result. In the figure one can also see the mass obtained using MC samples generated with ±3 GeV or ±6 GeV with respect to the nominal (172.5 GeV) sample. Considering variations of the input mass of ±3 GeV, the uncertainty of the measurement due to the MC mass used as reference is below 1 GeV.

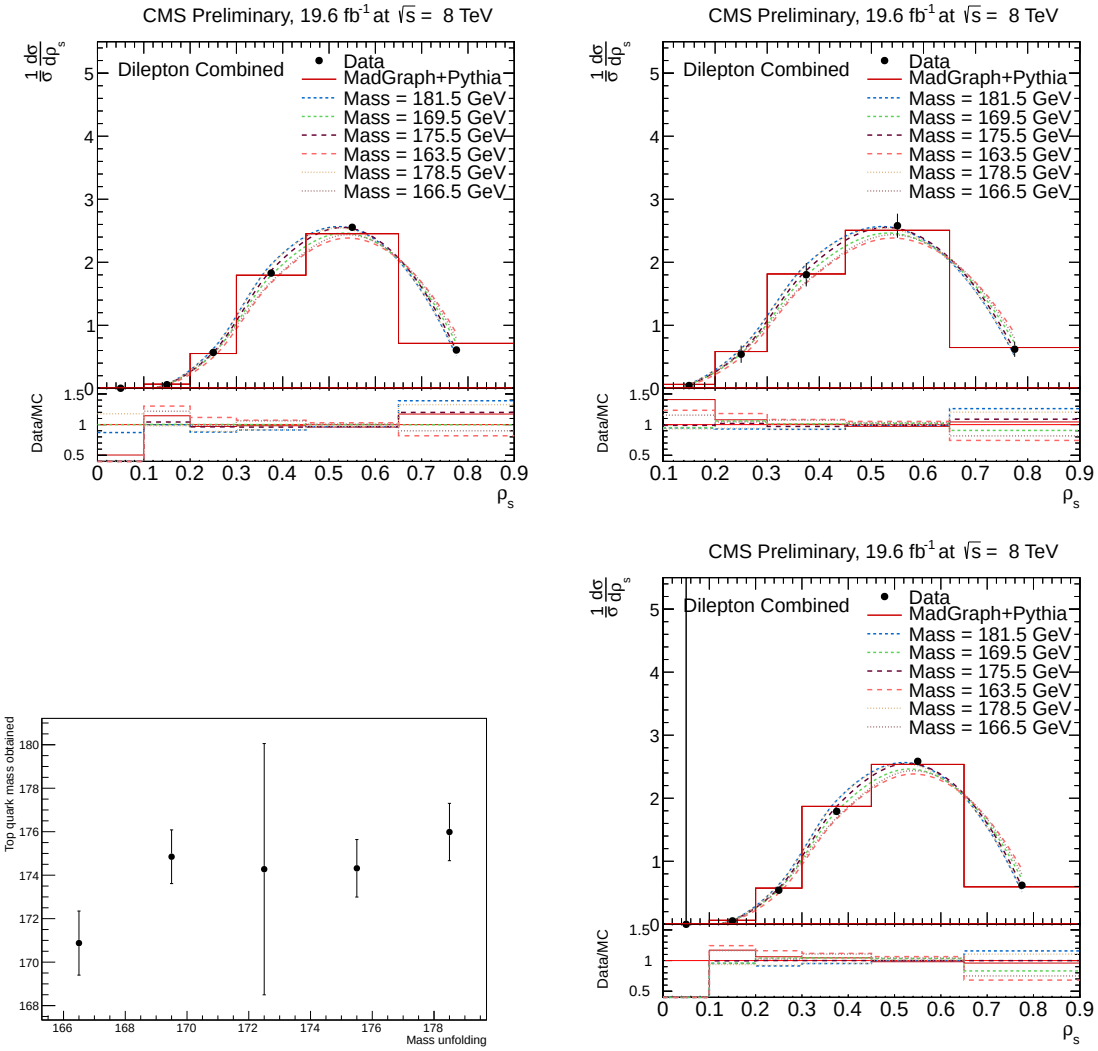

Figure 8: Differential cross section distributions obtained with variated top quark masses. Masses used for the unfolding and $t\bar{t}$ contribution estimated: (top left) m_{top} = 169.5 GeV, (top right) m_{top}^{nom} = 172.5 GeV(nominal value) and bottom right m_{top} = 175.5 GeV. The dependence of the result obtained as a function of the top quark mass used for the unfolding is shown at the bottom left. The central point in this graph includes the total error, while the rest of the measurements shows only the statistical uncertainty. Here the red histogram correspond to the reference sample used (169.5, 172.5 or 175.5 GeV).

6.2 Value of m_0

In the reference, m_0=170 GeV was fixed, but small variations from this value were considered in this work as the only requirement on this constant was to be close to the top quark mass value. The cross section distributions found can be seen in Figure 9, where the expected small changes in the cross section do not lead to significative changes in the top quark mass obtained. The top quark mass values obtained from each distibution are shown in Table 4, where it can also be seen that they are compatible with each other within the statistical uncertainties. Note that the binning is not optimised in terms of purity and stability.

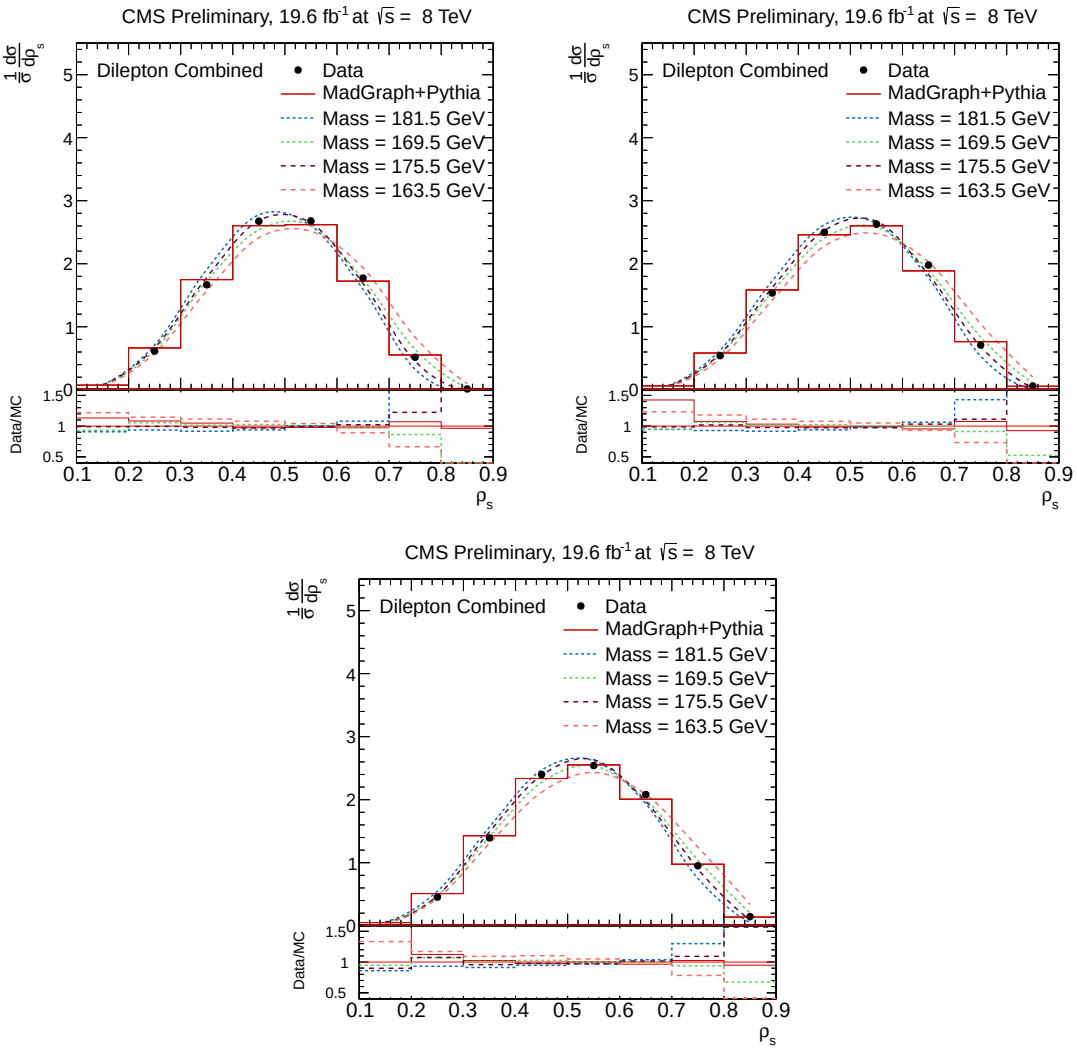

Figure 9: Measurements of the cross section with different m_0 values are shown: m_0 =165 GeV (top left), m_0 =170 GeV (top right) and m_0 =175 GeV (bottom). The errors on the data points corresponds to the statistical uncertainty.

Table 4: Top quark masses obtained for the different m_0 values studied. Results are compatible within the statistical uncertainties.

m_0 (GeV)	Mass (GeV)	Stat. Error (GeV)
165	174.1	0.6
170	174.7	0.8
175	173.6	1.0

6.3 Neutrino p_T spectrum

As explained in Section 2, neutrino energy spectrum is used during the kinematic reconstruction process to select the solution. This spectrum is generated acording the nominal top quark mass value and, therefore, may be affected by top quark mass variations. In order to study this effect, several neutrino p_T spectum were generated using variated top quark masses (again, from 163.5 to 181.5 GeV in stepts of 3 GeV). A comparison between the distributions obtained can be seen in Figure 10.

7 Conclusions

The differential cross section was measured as a function of the ρ_s observable ($\rho_s = 2m_0/\sqrt{s_{t\bar{t}j}}$) and the top quark mass was extracted from the sensitive region to mass variations of this distribution. The results obtained were found to be independant on the MC mass used in the extraction of the cross section and on the m_0 value.

The most precise result, $m_t = 173.0 \pm 2.0(stat.) \pm 4.3(syst.)$, was obtained using the $e\mu$ channel only and it is in good agreement (within systematical uncertainties) with both the resul obtained for the combination of the three channels, $m_t = 175.0 \pm 1.5(stat.) \pm 5.8(syst.)$, and the current reference value, $m_t = 173.20 \pm 0.87$ GeV.

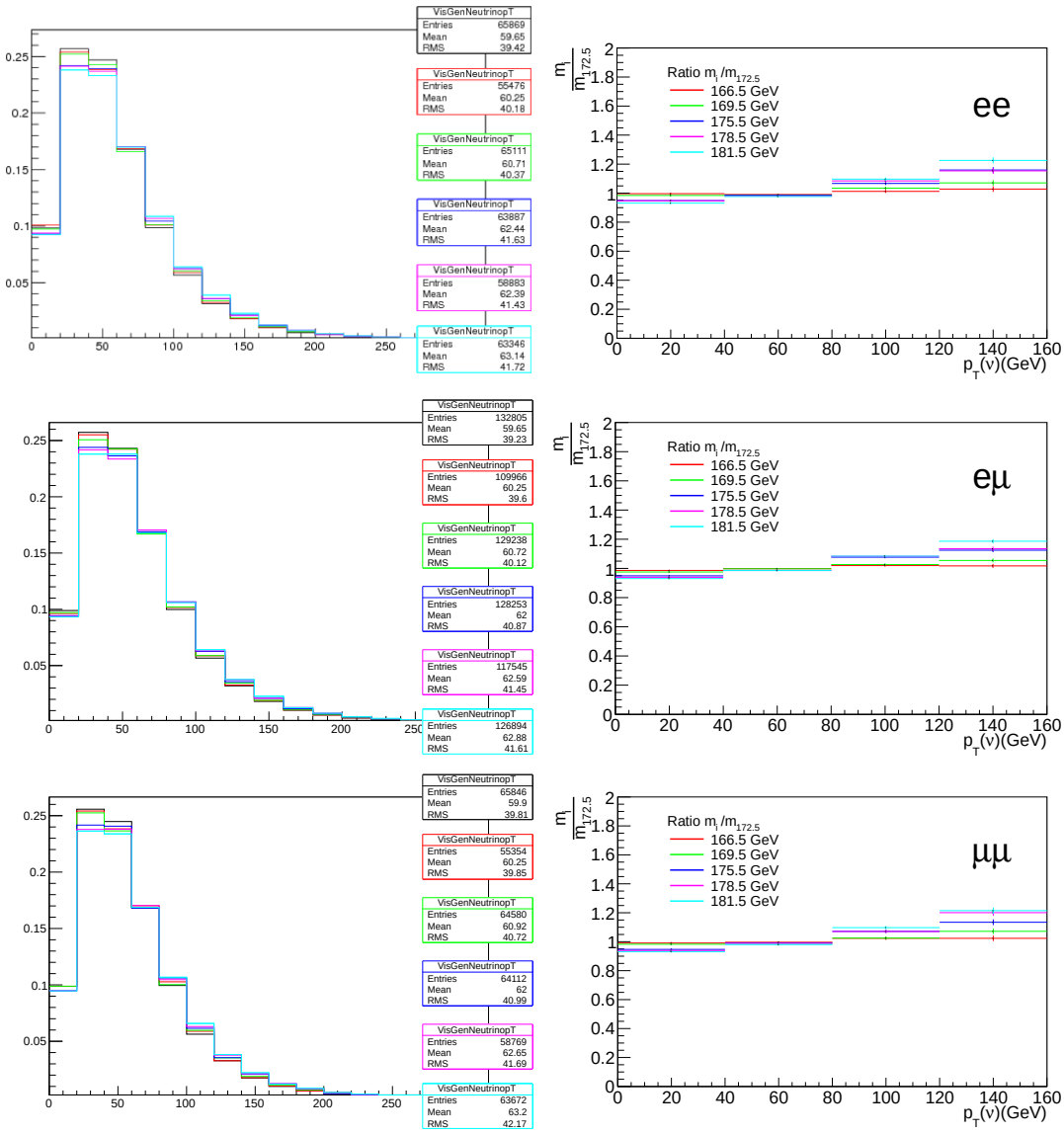

Figure 10: On the left side generated neutrino p_T spectra are shown for each channel. Each spectrum was generated with a variated top quark mass (from 163.5 to 181.5 GeV in stepts of 3 GeV) in order to study possible differences between them. The ratios between each sample obtained with a variated mass and the one genereated with the reference value ($m_t^{nom} = 172.5$ GeV) are shown in the right side. Results for the ee, $e\mu$ and $\mu\mu$ channels are shown in each row.

References

[1] Simone Alioli, Patricia Fernandez, Juan Fuster, Adrian Irles, Sven-Olaf Moch, et al. A new observable to measure the top-quark mass at hadron colliders. *Eur.Phys.J.*, C73:2438, 2013.

[2] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 03:S08004, 2008.

[3] The CMS Collaboration. Measurement of the Jet Multiplicity in dileptonic Top Quark Pair Events at $\sqrt{s}$ =8 TeV. *CMS Physics Analysis Summary*, TOP-12-041, 2013.