Study of $\mu\tau_h$ final state using 2017 data with the CMS experiment at the LHC

Author: Oleg Filatov*

Supervised by: Valeria Botta†, Elisabetta Gallo†, Alexei Raspereza†

Abstract

The goal of the Summer Student project reported here is to analyze the first 2017 data collected by the CMS experiment at the LHC in proton-proton collisions at a center-of-mass energy of 13 TeV. The analysed data corresponds to an integrated luminosity of 3.78 fb^{-1}. The study focuses on ditau events in the $\mu\tau_h$ final state, including the comparison of several key variables distributions in 2016 and 2017 data as well as the comparison of 2017 data to the latest available simulation (2016). For the latter, dedicated muon efficiency corrections were derived using a tag-and-probe technique. As a result, overall good agreement between the 2017 and 2016 data distributions is observed, also when comparing 2017 data with the simulation. In addition, a new RooFit method of fitting mass distributions for the tag-and-probe method has been implemented. The efficiencies curves obtained with this new technique are consistent with the previous one, except for some slight discrepancies in the low p_T region, which requires further study.

*Moscow Institute of Physics and Technology (MIPT), Moscow, Russia

†Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany

Contents

1 Introduction

In 2012 a scalar boson particle was discovered by the ATLAS and CMS experiments at the CERN Large Hadron Collider (LHC) in the ZZ, $\gamma\gamma$ and W^+W^- decay channels, analysing data of proton-proton collision collected in 2011 and 2012 at center-of-mass energies of $\sqrt{s} = 7$ and 8 TeV respectively [1–3]. The properties of such particle are so far compatible with the neutral scalar particle predicted by the Brout-Englert-Higgs mechanism of electroweak symmetry breaking [4–9], commonly known as the Higgs boson (H).

After the discovery of the Higgs boson, its properties and couplings to Standard Model particles are being studied [10] with increasing precision, as more data are becoming available from the LHC. For instance, in order to establish the direct coupling of H to fermions, which is essential to understanding of the mass generation mechanism, the most promising channel is the Higgs to $\tau\tau$ one, that has the best signal-to-background ratio for a Higgs mass of 125 GeV. Recently, CMS published the results of $\mathrm{H} \to \tau\tau$ analysis including 2016 data [11], presenting the first observation of this decay by a single experiment.

Currently, CMS has just restarted taking data in June 2017 after an extended technical stop period during which the pixel detector has been completely replaced with a new upgraded one [15]. The aim of this Summer Student project is to explore the features of the very first 2017 data collected with the CMS Detector in the context of the aforementioned Higgs to $\tau\tau$ analysis, focusing on the $\mu\tau_{\mathrm{h}}$ final state. The amount of 2017 data studied in this work corresponds to an integrated luminosity of 3.785 fb^{-1}

In particular, we measured muon identification and trigger efficiencies in 2017, compared important kinematic distributions in 2017 and 2016 data and finally performed data/Monte Carlo (MC) comparison. As a dedicated MC for 2017 is not yet available, the one produced for 2016 analysis has been used, corrected with the needed scale factors.

In addition to that, we improved the code used for fitting Z boson invariant mass distributions, needed to measure the muon efficiencies, with implementation of the RooFit package [22].

The report will proceed as follows. In Section 2 a brief description of the CMS detector is given. Section 3 is dedicated to the Monte Carlo scale factor corrections calculation. In Section 4 we describe the process of $\mu\tau_{\mathrm{h}}$ final state reconstruction and compare 2017 data with 2016 data/2016 MC. Additionally, the idea of mass distributions fitting improvement is proposed in Section 5. Lastly, we summarize our results in Section 6.

2 The CMS detector

The CMS (Compact Muon Solenoid) detector (Fig. 1) is a general purpose detector in the form of a cylinder with 22 m in length, 15 m in diameter and total weight of 12.5 t. It is comprised of a silicon tracking system (pixel + strip detector), a lead tungstate crystal electromagnetic homogeneous calorimeter (ECAL), a brass and scintillator sampling hadron calorimeter (HCAL) and gas-ionization muon chambers

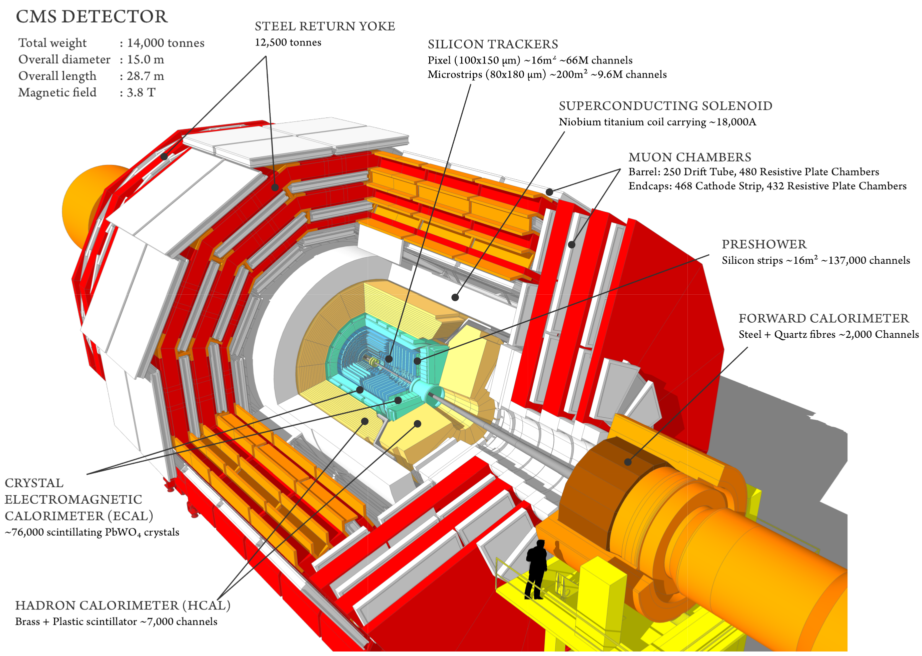

Figure 1: The CMS detector at the LHC [12]

embedded in the steel flux-return yoke outside of the solenoid. The superconductive solenoid provides a magnetic field of 3.8 T.

Events of interest are selected using a two-tiered trigger system. The first level (L1), composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a time interval of less than 4 μs. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to about 1 kHz before data storage.

A complete description of the CMS detector can be found elsewhere [19].

Now the CMS detector is intensively being prepared for the High Luminosity LHC upgrade [13] in order to improve the overall efficiency of the detector under increased instantaneous luminosity. This is planned to be achieved in two Phases. The Phase-I Upgrade [14] has been almost finished and involved modification of the Pixel Tracker [15], HCAL [16] and Level-1 Trigger systems [17]. The Phase II Upgrade [18] will be installed from the beginning of 2024 and is expected to be largely completed in 2026.

3 Monte Carlo corrections

One of the important milestones in this analysis is to obtain a good description of the data through the MC simulation. In order to achieve this, several corrections, referred further as scale factors (SF), need to be derived and applied to the MC.

Tag&Probe method

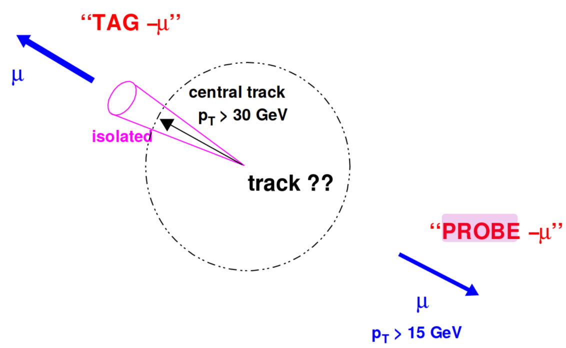

Figure 2: Schematic illustration of the T&P method (arXiv:1003.0521)

To derive these MC corrections, one firstly has to calculate muon identification and isolation as well as trigger efficiencies in data and MC. To do this, we use the Tag&Probe (T&P) method. The idea behind it, e.g. in case of muons, is to use resonances decaying into muon pairs such as J/ψ, $\psi(2S)$, Z boson depending on the p_T of interest. For this study we select events with two opposite sign muons with invariant mass consistent with the Z boson decay. Then we apply strict cuts on one of the muons in order to "tag" the Z boson (further referred as "tag" muon) and loose cuts on the other one, on which we will "probe" our selection criteria (further referred as "probe" muon). Fig. 2 provides a schematic illustration of the T&P method, while the cuts applied to the tag and probe muons are listed in Table 1.

The efficiency of the selection criteria, like identification, isolation and trigger requirements, can be estimated from the number of probe muons passing the selection criteria, w.r.t the total number of probe muons. The efficiencies are measured by fitting the dimuon invariant mass distributions in bins of p_T and η. As a parameter of the fit we obtain the number of passing and failing probes and eventually can calculate the efficiency for that particular p_T and η bin as

$$\varepsilon_{\mathrm{Data/MC}}(\mathrm{p_T}, \eta) = \frac{N_{\mathrm{pass}}(\mathrm{p_T}, \eta)}{N_{\mathrm{pass}}(\mathrm{p_T}, \eta) + N_{\mathrm{fail}}(\mathrm{p_T}, \eta)} \tag{1}$$

Tag	Probe
HLT IsoMu24eta2p1, Medium ID muon, $p_T > 25$ GeV/c, $\lvert\eta\rvert < 2.1$, $\lvert d_{xy}\rvert < 0.045$ cm, $\lvert d_z\rvert < 0.2$ cm, Iso < 0.15 (in the cone $\Delta R \equiv \sqrt{(\Delta \mathrm{p_T})^2 + (\Delta\eta)^2} = 0.4$)	$p_T > 5$ GeV/c, $\lvert\eta\rvert < 2.4$ $\lvert d_{xy}\rvert < 0.2$ cm $\lvert d_z\rvert < 0.5$ cm, ID+Iso (trigger only)
$m(\mu^+\mu^-) > 50$ GeV, $\Delta R(\mu^+, \mu^-) > 0.5$, $\Delta R = 0.5$ trigger matching	

Table 1: Cuts applied for tag and probe muon in T&P method for muon ID+Iso and trigger efficiency estimation

The measurement is performed separately in data and in MC to derive $\varepsilon_{\mathrm{data}}(\mathrm{p_T}, \eta)$ and $\varepsilon_{\mathrm{MC}}(\mathrm{p_T}, \eta)$. After we calculate these efficiencies we derive scale factors as a function of p_T and η using the ratio of efficiencies:

$$\mathrm{SF}(\mathrm{p_T}, \eta) = \frac{\varepsilon_{\mathrm{data}}(\mathrm{p_T}, \eta)}{\varepsilon_{\mathrm{MC}}(\mathrm{p_T}, \eta)} \tag{2}$$

Then while doing our analysis we weight our MC events with these SFs, based on the p_T and η of muons in the event.

Muon Efficiencies & Scale Factors

The efficiencies and scale factors for muon identification and isolation criterion (ID+Iso), requiring Medium Muon ID, $|d_{xy} < 0.045$ cm$|$, $|d_z < 0.2$ cm$|$ and relative muon isolation < 0.15 (in $\Delta R = 0.4$ cone)), as well as the trigger efficiencies of the single muon trigger HLT_IsoMu24 have been measured for 2017 data and 2016 MC. Throughout this analysis we use the SingleMuon/Run2017B-PromptReco-v1/MINIAOD dataset with total integrated luminosity (referred as "2017 data") of 3.785 fb^{-1} with corresponding JSON file Cert_294927-297723_13TeV_PromptReco_Collisions17_JSON.txt. As MC simulation for the efficiency calculation we use the 2016 MC Drell-Yan (DY) Z/γ^*+jets samples (referred as "2016 MC").

Corresponding efficiency plots have been obtained (Fig. 3, 4). The scale factors in the form of 2D histograms are presented in Fig. 5. In general, the 2017 ID+Iso efficiency is the same or higher compared to the one of the 2016 data. The best improvement takes place in the last $|\eta| > 2.1$ bin. One possible explanation for this might be the upgrade of the pixel detector, particularly the installation of an additional endcap layer on each side.

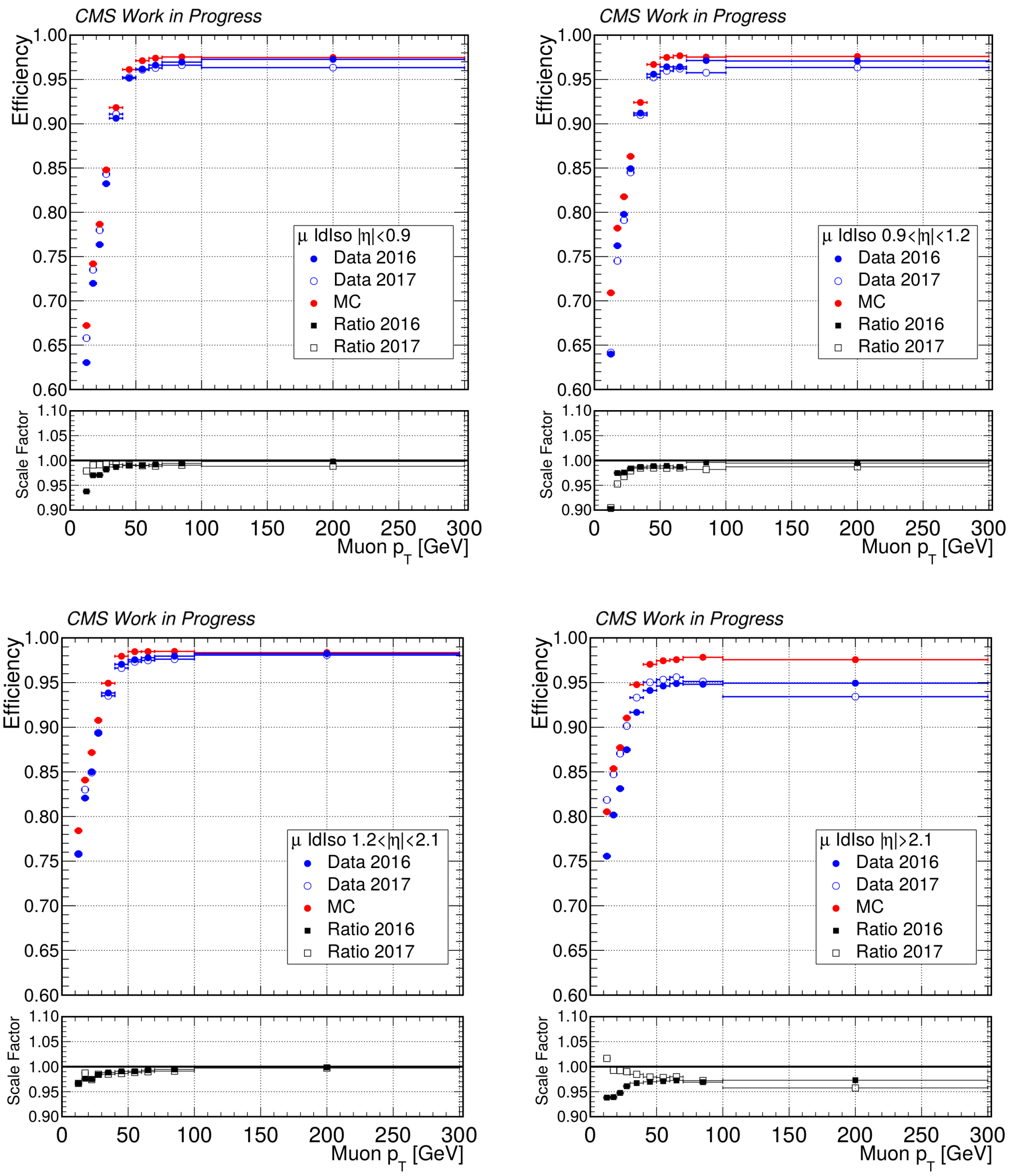

Figure 3: Efficiency plots for ID+Iso criteria for 2017 data, 2016 data and 2016 MC.

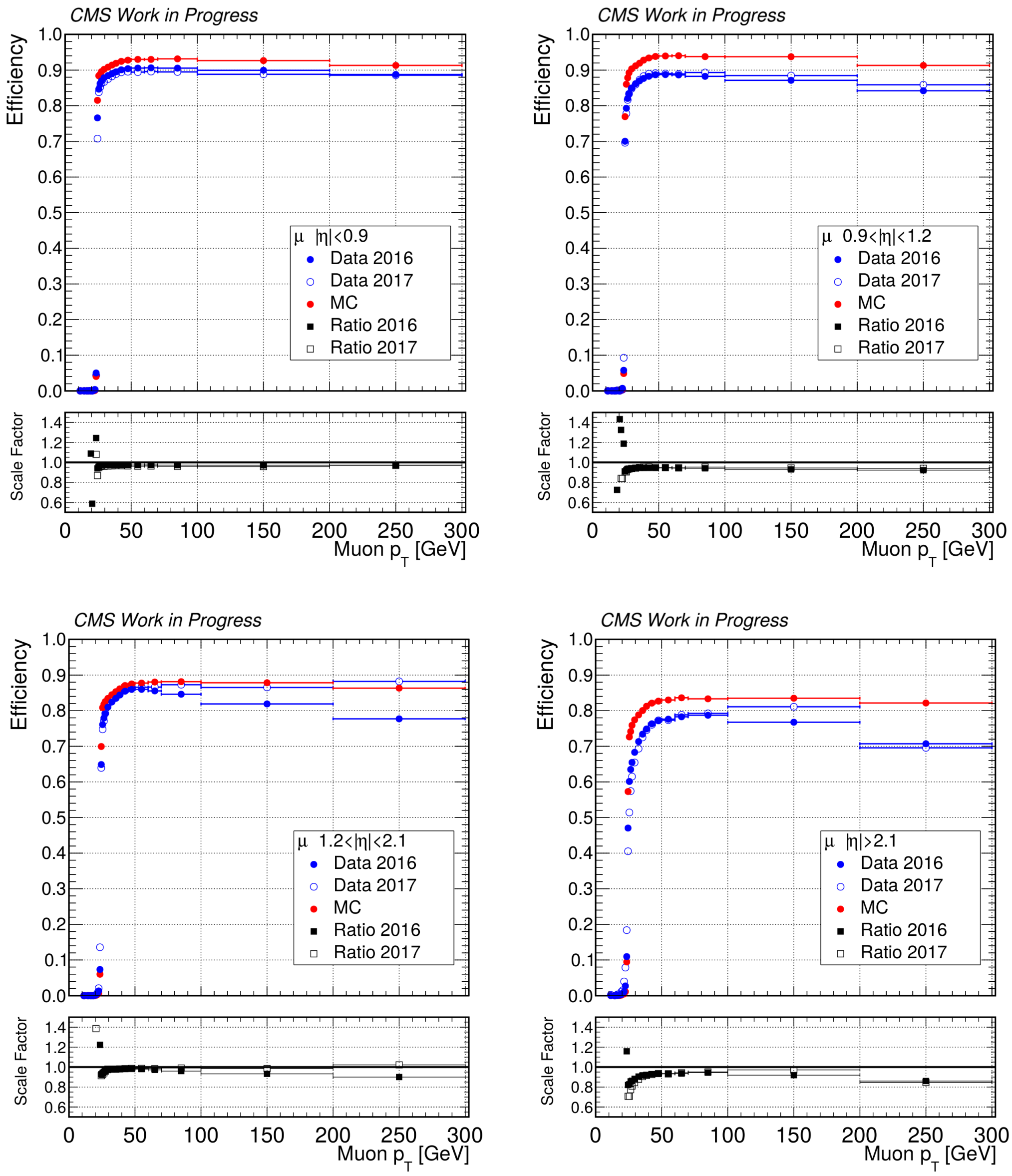

Figure 4: Efficiency plots for HLT_IsoMu24 trigger selection for 2017 data, 2016 data and 2016 MC.

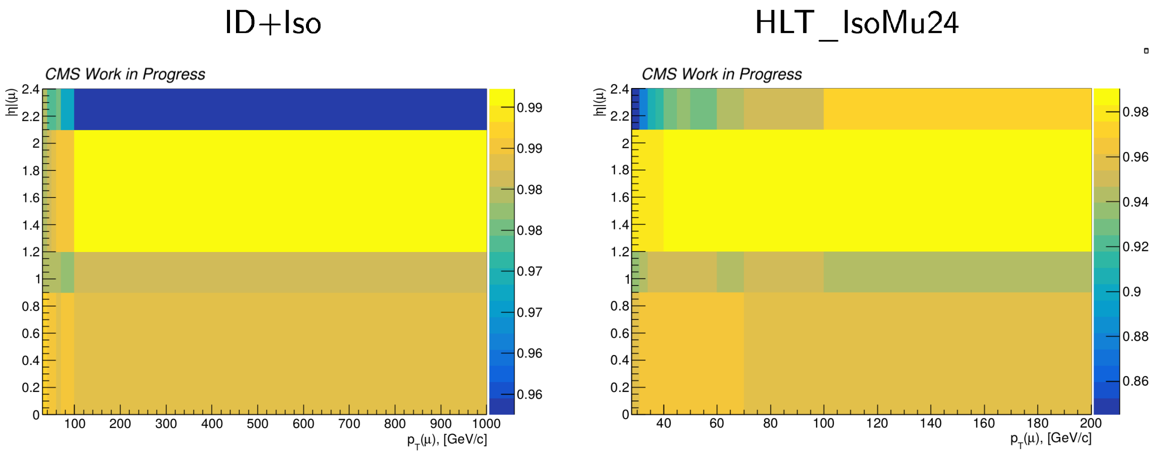

Figure 5: Scale factor corrections for 2016 MC as a bin function of p_T and η

4 $\mu\tau_h$ final state

Having muon ID+Iso SF corrections been updated we reconstructed the $\mu\tau_h$ final state, where μ stands for a τ lepton decaying into μ channel and τ_h for a hadronically decaying tau [20]. Such final state accounts for $\approx 20\%$ of the $\tau\tau$ decay modes.

We select events with one isolated muon matched to the trigger (HLT_IsoMu24) and one hadronically decaying τ passing the tight identification requirement. In order to suppress the backgrounds, events with additional isolated leptons are vetoed and the transverse mass of the MET (MET stands for the missing energy transverse, the imbalance in the transverse momentum of all visible particles) + muon system is required to be < 50 GeV. The full selection is listed in Table 2.

2017 data/ 2016 data comparison

As first step, the comparison of 2017 data distributions with the previous 2016 data in the $\mu\tau_h$ final state has been made. For 2016 data we use SingleMuon__Run2016B-H datasets with 35.9 fb^{-1} of integrated luminosity. The corresponding JSON file is Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt. In order to compare 2017 and 2016 distributions, the latter are scaled with the ratio of integrated luminosities(3.785/35.9).

The comparison of some key variables in 2017 and 2016 data is shown in Fig. 6. After applying the same selection cuts and luminosity scaling we obtain approximately the same number of events (31920 and 32506 for 2016 and 2017, respectively) and in general nicely matching distributions, which is a promising start. However, there are some discrepancies to be explained. Firstly, there is a small excess of events in the high-η region for 2017 data. It might be because of the presence of additional layers in the new pixel tracker [15] or because of the higher tau fake rate that can be observed from the visible mass plot as excess of events around the Z boson mass. Secondly, complete difference in ϕ(MET) distributions is due to using uncorrected MET in 2017 data and corrected in 2016 data. Then, the wider distribution of τ impact parameter in 2017 might be explained with not yet accurate tracker alignment. Finally, from the hadronic tau MVA discriminants (byTightIsolationMVArun2v1DBoldDMwLT) distributions one can observe the need to retrain such MVA classifier.

μ	τ_h
HLT_IsoMu24 matching, $p_T > 25$ GeV/c, $\|\eta\| < 2.4$, $\|d_{xy}\| < 0.045$ cm, $\|d_z\| < 0.2$ cm, Iso < 0.15, $m_T < 50$ GeV	$p_T > 20$ GeV/c, $\|\eta\| < 2.3$, $\|d_z\| < 0.2$ cm, againstElectronVLooseMVA6 > 0.5, againstMuonTight3 > 0.5, byTightIsolationMVArun2v1DBoldDMwLT > 0.5
opposite sign, dilepton_veto = 0, extraelec_veto = 0, extramuon_veto = 0	

Table 2: Cuts applied on μ and τ_h in the reconstruction of $\mu\tau_h$ final state

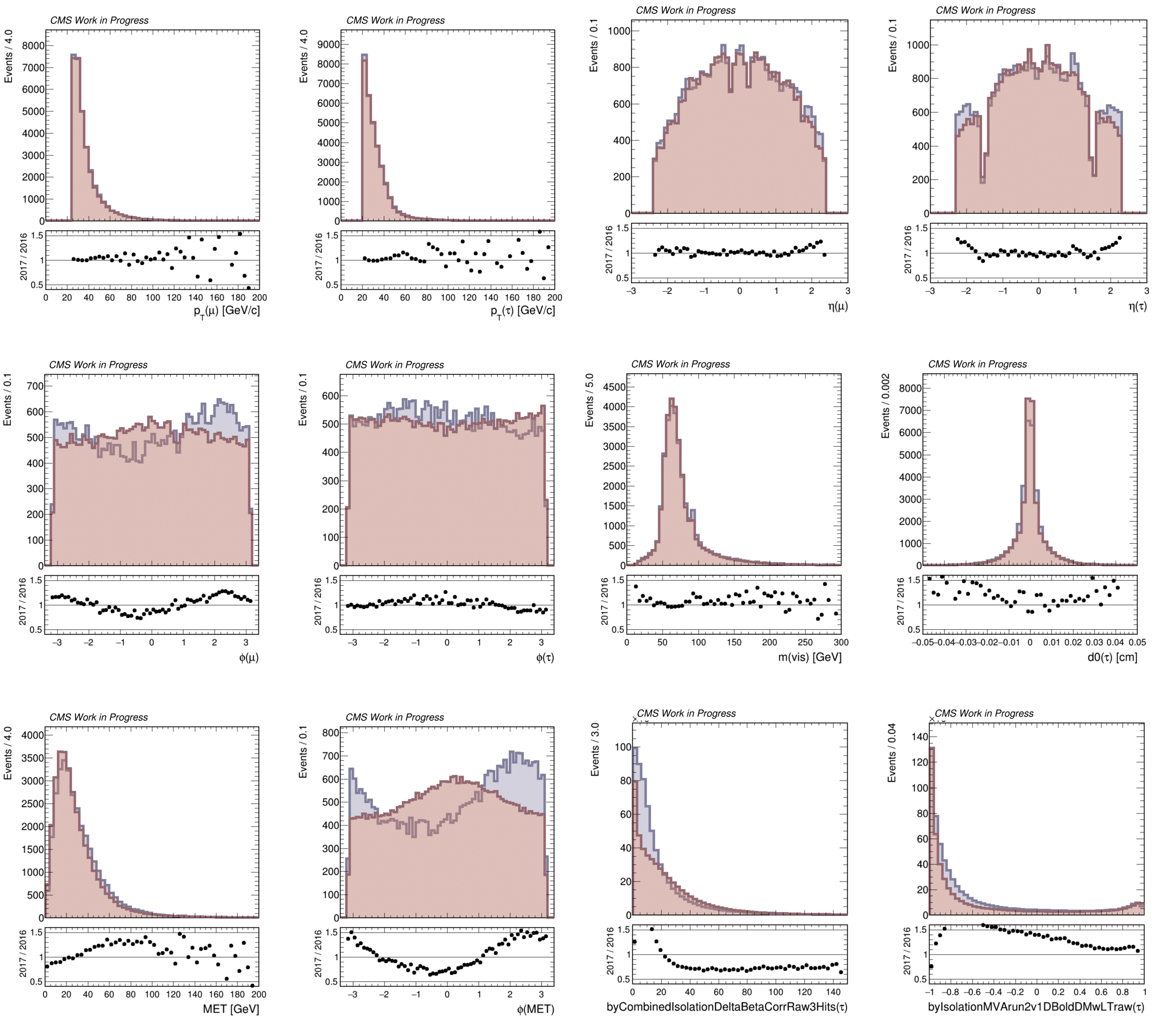

Figure 6: Distributions of several variables for $\mu\tau_h$ final state 2017 data/2016 data comparison. The brown filled histogram corresponds to 2016 data, the grayish to 2017 data.

2017 data/2016 MC comparison

After comparing 2017 data with 2016 data we considered checking of 2017 data/2016 MC matching, keeping in mind the SFs previously derived. As for the simulation samples used, signal H(125)$\rightarrow \tau^+\tau^-$ events are modeled for gluon-gluon fusion (ggH), vector boson fusion (VBF) and associated W/Z boson production (WH or ZH) processes. The following processes are used to describe background events: single top, ZZ, WW, WZ, $t\bar{t}$, W+jets, DY $Z/\gamma^* +$jets. Using the same cuts as for the 2017 data/2016 data comparison (Section 4) (except for the tighter $p_T(\tau_h) > 30$ GeV/c cut) and applying the following weights to correct the MC events the same plots as in the previous section were obtained (Fig. 8,9):

- MC generator weight
- pile-up weight (as in 2017)
- muon ID+Iso and trigger scale factors (2017 data/ 2016 MC)
- τ ID scale factor (measured in 2016)
- μ tracking efficiency (measured in 2016)
- Z p_T reweighting (measured in 2016, applied to DY only)
- $\mu - \tau$ fake rate scale factor (measured in 2016)
- top p_T reweighting (measured in 2016, applied to $t\bar{t}$ only)

In general, good agreement of data and simulation is observed, except for the already discussed MET and MVA discriminant distributions. Furthermore, weighting events with the scale factors indeed improves data/MC ratio. Most noticeably it can be seen in the m_{vis} variable, where in the first three bins the data/MC ratio = 1 is achieved when applying muon SFs. Also, applying scale factors an improvement is seen in the $p_T(\mu)$ distribution. The other variables are not significantly affected.

5 Fitting improvement

As final step of the project, the idea of fitting procedure improvement was suggested. Fitting of mass distributions is the essential step in the muon efficiency calculation explained in Section 3. Previously, the standard TH1::Fit method of ROOT software [21] was used. Instead, the usage of the dedicated RooFit package [22] methods has been made available.

In RooFit the construction of fitting models is implemented in a very intuitive way, making the fitting process simple, compared to ROOT. Also, RooFit allows to perform a fit of datasets obtained directly from a TTree's branch, while in ROOT only histograms might be fitted, which gives us statistically more "trustworthy" results. Finally, it is not trivial to extract the number of events in ROOT after the fit, while in RooFit the number of events is a parameter and can be obtained directly.

	ROOT	RooFit
Signal (fail)	2-sided double (right), single (left) Gaussian + FSR	BW $\otimes$ Gaussian (Voigtian) + FSR
Signal (pass)	2-sided double Gaussian	BW $\otimes$ Crystal Ball
Background	Exponential	
FSR	Gaussian	
RooFit Comments	Failing: fit every p_T bin with signal (fail) + FSR; Passing: force to fit with signal (fail) + FSR first five p_T bins (< 40 GeV/c)	

Table 3: Description of the models used to fit mass distributions in failing and passing T&P categories before (with ROOT) and after (with RooFit)

The models used in the previous ROOT fitting method and in the new RooFit method are briefly described in Table 3. In the failing category (that is when a probe muon fails selection criteria) as a signal model we use a sum of the Breit-Wigner function convoluted with Gaussian and FSR (final state radiation) component, which is used to model the decay of Z boson when one (or both) muon radiates a photon and thus reduces the overall invariant mass of dimuon system. Before the combination of 3 half-gaussians (one right-sided and two left-sided) with 3 different sigmas and common mean was used. In the passing category in RooFit we use the convolution of Breit-Wigner and Crystal Ball functions [23], which properly describes the Z peak tail. Before we used 4 half-gaussians (two right-sided and two left-sided) with 4 different sigmas and common mean which is, in fact, a slightly artificial model. Also, in order to describe low-p_T mass distributions properly we change this model to the failing signal model. In both categories we use exponential function as the background model. Examples of the fit are illustrated on Fig.10.

The efficiency and scale factor curves comparison for both techniques have been extracted and the corresponding results are shown in Fig. 11 and 12. As can be observed, both methods give the same results in all η and p_T bins but the first two. This leads to the assumption that we don't understand correctly the background behavior in the low-p_T region, especially in data. One of the possible solutions might be fixing the mass distribution template from MC as a signal model and fitting data distribution with this template and additional background model (exponential function, as first approximation).

6 Summary

As the result of this research, the first look at 2017 data has been made. Starting from the calculation of muon ID+Iso criteria and trigger efficiencies, we have derived corresponding scale factor correction to the 2016 year MC samples. Then, the $\mu\tau_{\mathrm{h}}$ final state has been reconstructed. Firstly, we have compared 2016 data with 2017 data distributions. Secondly, keeping in mind the contribution of Higgs boson decaying into a τ lepton pair in this final state and weighting MC events with scale factors previously estimated, we have compared 2016 MC with 2017 data. In both cases, good data/data (data/MC) agreement is observed. Lastly, the improvement of fitting procedure in the muon efficiency calculation with introducing RooFit package has been done. Scale factors calculated with the new method are consistent with the previous method, except for the first two $\mathrm{p_T}$ bins, which indicates the need to continue investigating the background's behavior in this region.

Apart from the mentioned results and what is more important, the basic knowledge of performing analysis in the Higgs physics at CMS has been acquired, from physical processes behind this topic to learning new software details and programming skills.

7 Acknowledgements

First and foremost, the author would like to greatly thank his supervisors Valeria Botta, Elisabetta Gallo and Alexei Raspereza for the profoundly interesting project to study, for giving (almost) complete carte blanche to explore the topic and yet for providing support and help where needed (that is almost everywhere).

Then, all the Summer Student Programme organizing team, especially Olaf Behnke, for creating the productive, enjoyable and ideally work-life balanced environment.

Next, all the summies, with whom the author managed to meet and make friends. As a matter of fact, students here turned out to be of the most all-round interests, from music to literature and philosophy, which impressed the author a lot and provided him with many new thoughts and ideas.

And last, but definitely not the least, the city of Hamburg and, moreover, Europe, deserves a very word of acknowledgement. Hamburg - for the harmoniously matching with the author's mood weather (also promoting productive working hours). And both of them - for the huge impression, inspiration and vast number of discoveries which deeply influenced the author's state of mind.

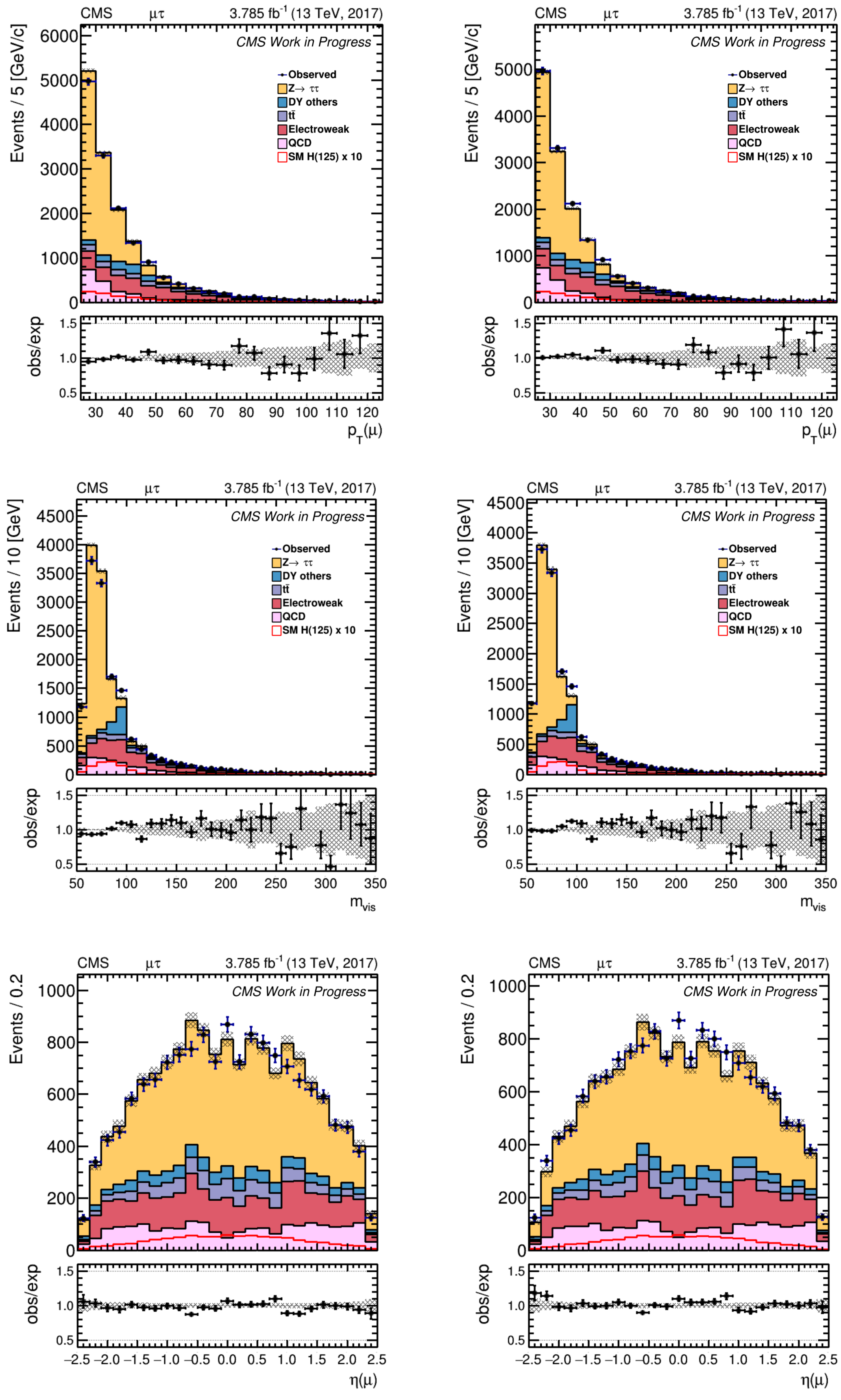

Figure 7: Comparison of variables distributions before (left) and after (right) applying SFs for $\mu\tau_{\mathrm{h}}$ final state 2017 data/2016 MC comparison.

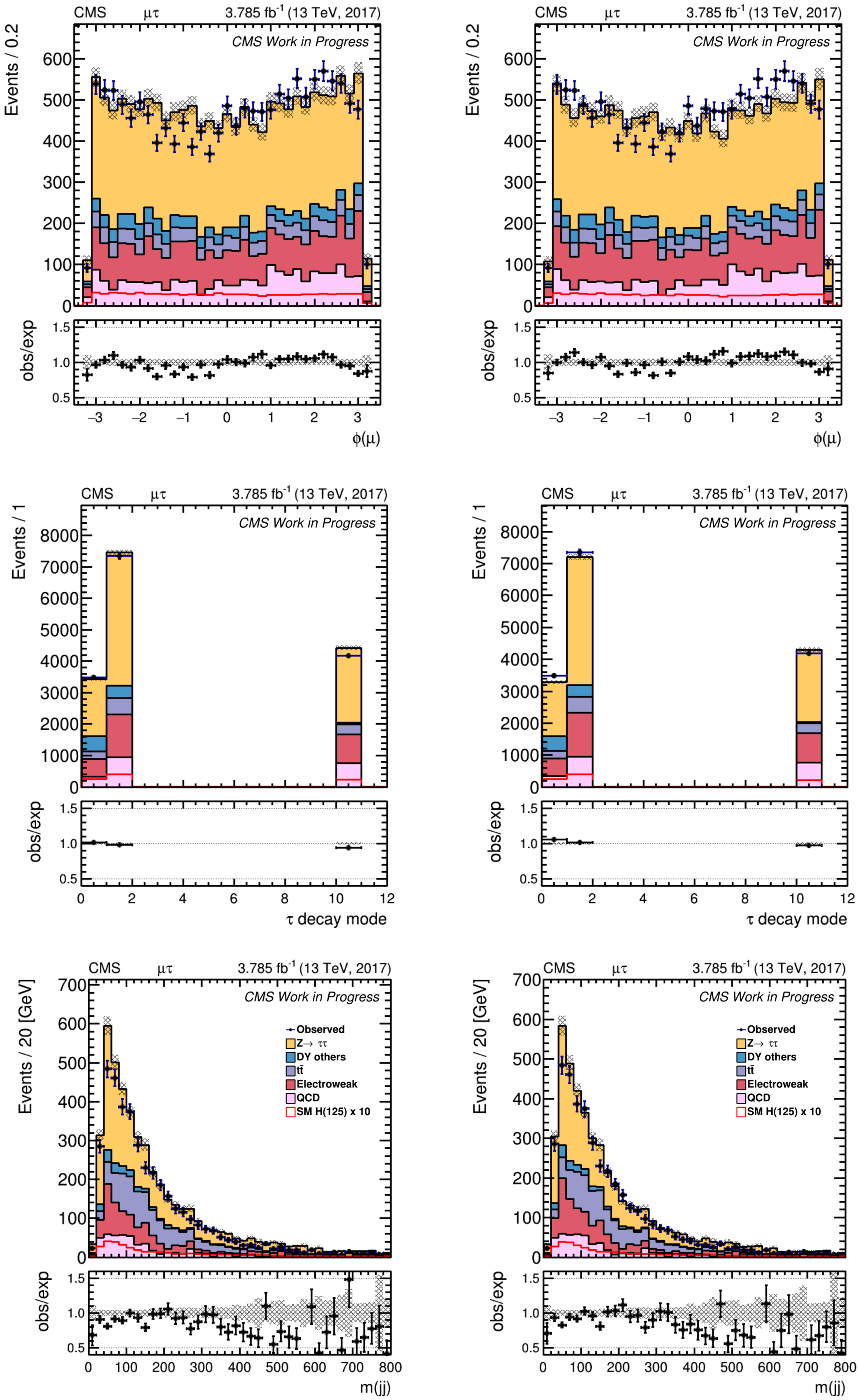

Figure 8: (continue) Comparison of variables distributions before (left) and after (right) applying SFs for $\mu\tau_{\rm h}$ final state 2017 data/2016 MC comparison.

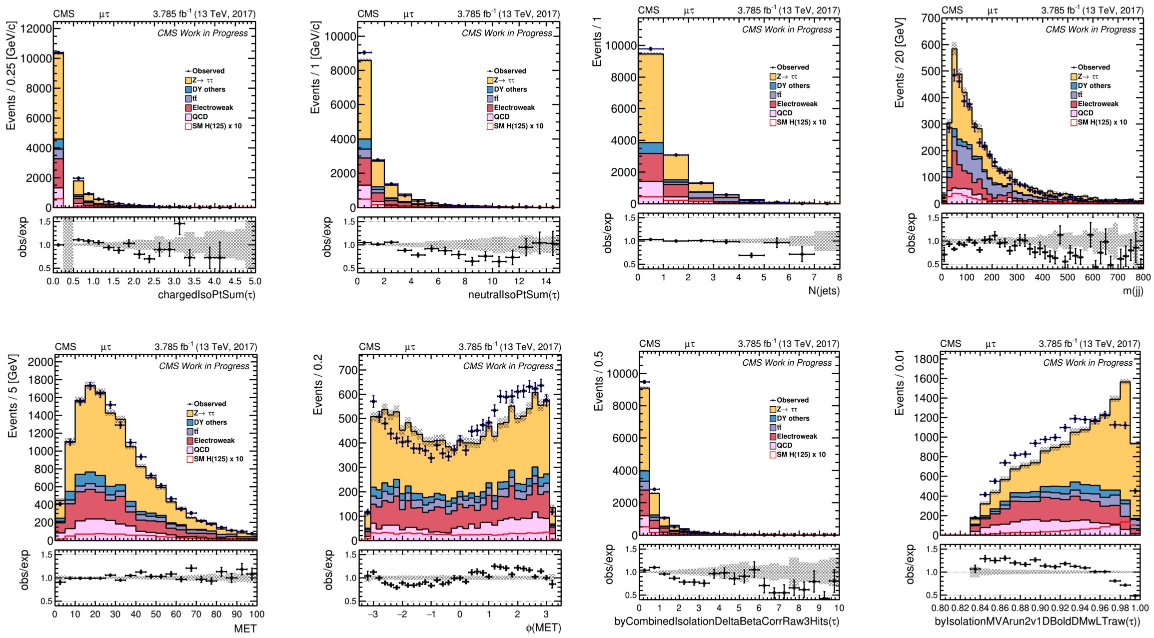

Figure 9: Other distributions for $\mu\tau_{\mathrm{h}}$ final state 2017 data/2016 MC comparison.

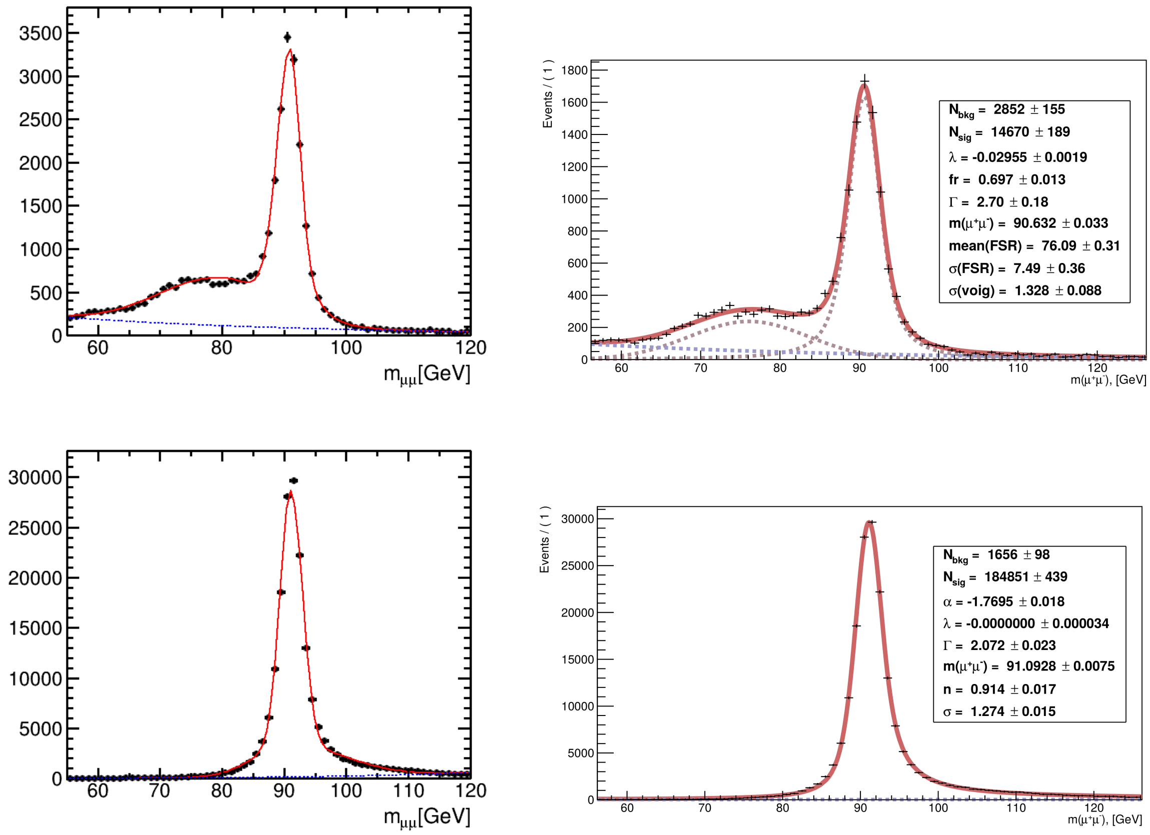

Figure 10: Examples of the fit in the failing (first line) and passing (second line) categories with ROOT (left column) and RooFit (right column) methods for some selected bin of p_{T} and η.

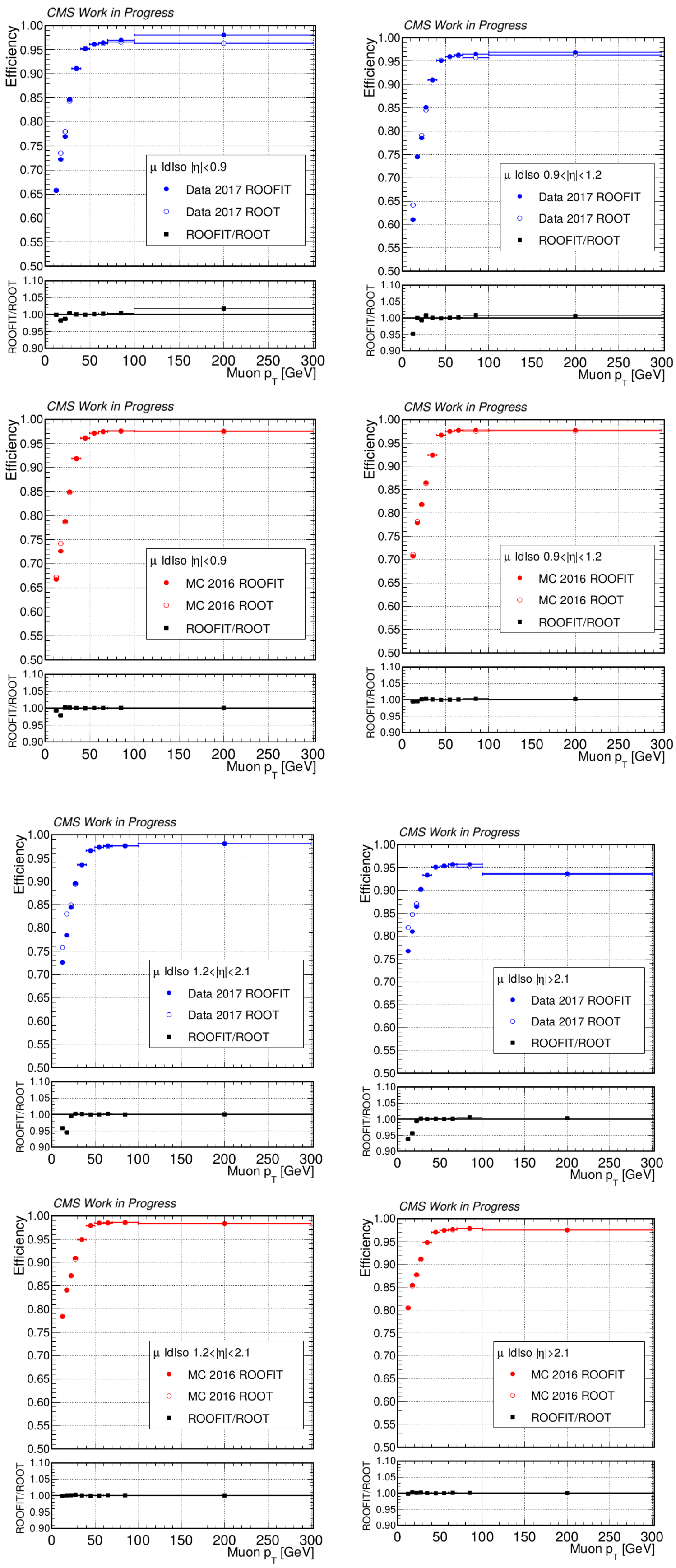

Figure 11: Comparison of muon ID+Iso criteria efficiencies on data (blue plots) and MC (red plots) obtained by fitting with ROOT and RooFit methods

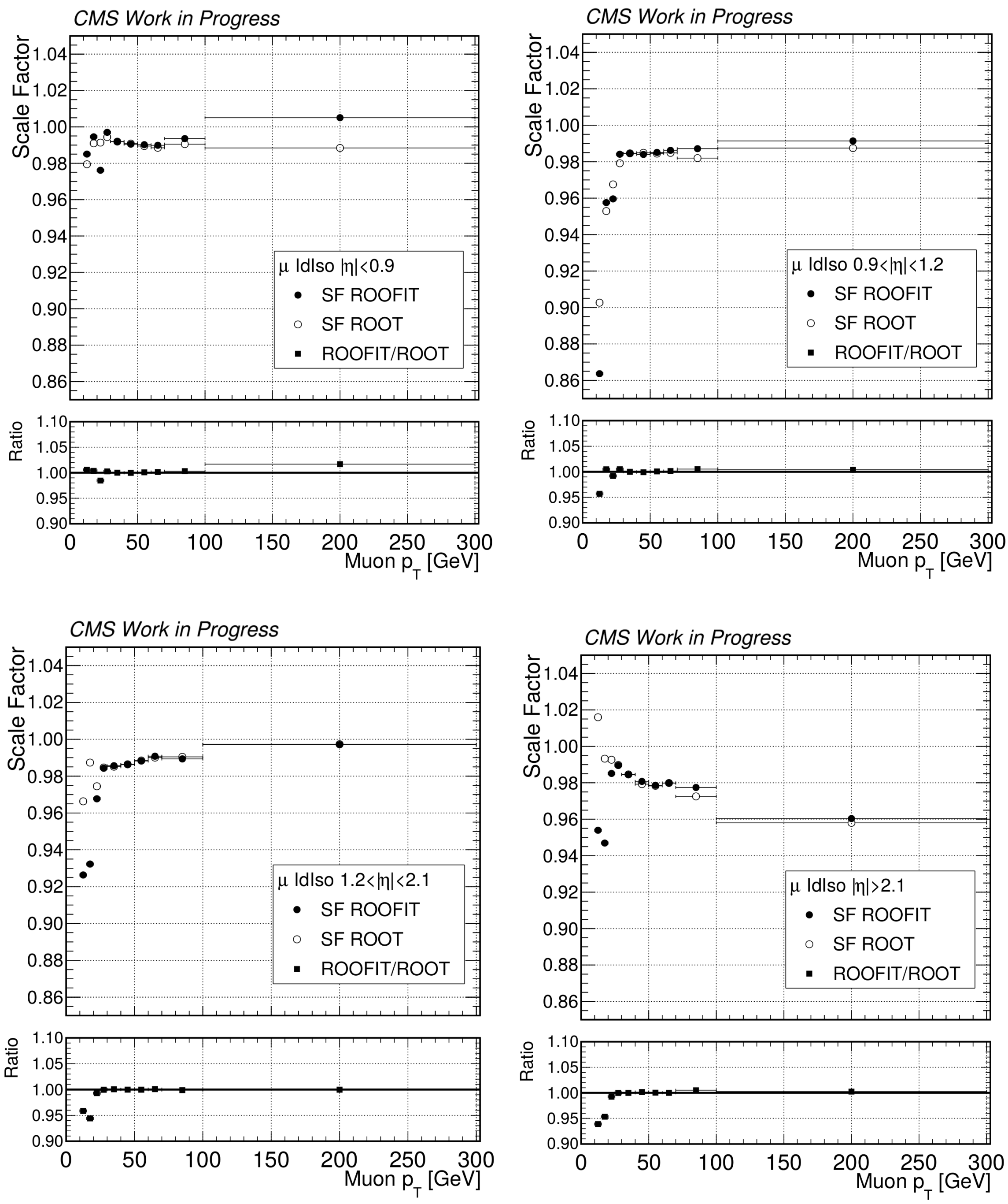

Figure 12: Comparison of the scale factors obtained with ROOT and RooFit methods.

References

[1] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", Phys. Lett. B 716 (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[2] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", Phys. Lett. B 716 (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[3] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV", JHEP 06 (2013) 081, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.

[4] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", Phys. Rev. Lett. 13 (1964) 321, doi:10.1103/PhysRevLett.13.321.

[5] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", Phys. Lett. 12 (1964) 132, doi:10.1016/0031-9163(64)91136-9.

[6] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", Phys. Rev. Lett. 13 (1964) 508, doi:10.1103/PhysRevLett.13.508.

[7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles", Phys. Rev. Lett. 13 (1964) 585, doi:10.1103/PhysRevLett.13.585.

[8] P. W. Higgs, "Spontaneous symmetry breakdown without massless bosons", Phys. Rev. 145 (1966) 1156, doi:10.1103/PhysRev.145.1156.

[9] T. W. B. Kibble, "Symmetry Breaking in Non-Abelian Gauge Theories", Phys. Rev. 155 (1967) 1554, doi:10.1103/PhysRev.155.1554.

[10] Aad, Georges, et al. "Combined Measurement of the Higgs Boson Mass in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments." Physical review letters 114.19 (2015): 191803.

[11] CMS Collaboration, "Observation of the Higgs boson decay to a pair of τ leptons" arXiv:1708.00373

[12] Sakuma, Tai, and Thomas McCauley. "Detector and event visualization with sketchup at the CMS experiment." Journal of Physics: Conference Series. Vol. 513. No. 2. IOP Publishing, 2014.

[13] Rossi, Lucio, and Oliver Brning. High luminosity large hadron collider a description for the European Strategy Preparatory Group. No. CERN-ATS-2012-236. 2012.

[14] CMS collaboration. Technical proposal for the upgrade of the CMS detector through 2020. No. CMS-UG-TP-1. 2011.

[15] CMS collaboration. "CMS technical design report for the pixel detector upgrade." CMS Technical Design Report CERN-LHCC-2012-016, CMS-TDR-11 (2012).

[16] Mans, J. CMS technical design report for the phase 1 upgrade of the hadron calorimeter. No. CERN-LHCC-2012-015. CMS-TDR-010, 2012.

[17] CMS collaboration. "CMS technical design report for the level-1 trigger upgrade." CMS Technical Design Report CERN-LHCC-2013-011, CMS-TDR-12 (2013).

[18] CMS collaboration. "Technical proposal for the phase-II upgrade of the CMS detector." CERN, CERN-LHCC-2015-010. LHCC-P-008 (2015).

[19] CMS Collaboration, "The CMS experiment at the CERN LHC." JINST 3 (2008): S08004.

[20] CMS Collaboration. "Reconstruction and identification of τ lepton decays to hadrons and $\nu\tau$ at CMS." (2016).

[21] ROOT Data Analysis Framework https://root.cern.ch

[22] The Toolkit for Data Modeling with ROOT https://root.cern.ch/roofit

[23] J. E. Gaiser, Appendix-F Charmonium Spectroscopy from Radiative Decays of the J/Psi and Psi-Prime, Ph.D. Thesis, SLAC-R-255 (1982)