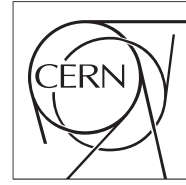**The Compact Muon Solenoid Experiment**

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

13 November 2014 (v2, 17 November 2014)

Data-driven background measurements in CMS

Matti Kortelainen for the CMS Collaboration

Abstract

The data-driven background measurement methods employed in the searches for a charged Higgs boson in the CMS experiment are reviewed. In the τ_h+jets final state analysis these backgrounds include the multijet events, and the EWK+$t\bar{t}$ events where the selected τ_h originates from a hadronically decaying τ lepton. In the $\mu\tau_h$ final state analysis the backgrounds from events with misidentified τ_h, and from Drell-Yan events with one genuine τ_h and the other τ decaying to a muon are measured from data.

Presented at *Charged2014 Prospects for Charged Higgs Discovery at Colliders*

Data-driven background measurements in CMS

MATTI KORTELAINEN* **on behalf of the CMS Collaboration**
Helsinki Institute of Physics, Finland
E-mail: matti.kortelainen@cern.ch

The data-driven background measurement methods employed in the searches for a charged Higgs boson in the CMS experiment are reviewed. In the τ_h+jets final state analysis these backgrounds include the multijet events, and the EWK+$t\bar{t}$ events where the selected τ_h originates from a hadronically decaying τ lepton. In the $\mu\tau_h$ final state analysis the backgrounds from events with misidentified τ_h, and from Drell-Yan events with one genuine τ_h and the other τ decaying to a muon are measured from data.

---

*Speaker.

1. Introduction

The CMS and ATLAS collaborations have observed a boson [1, 2] with a mass of approximately $m_H = 125\,\text{GeV}$ and properties consistent with the predicted standard model (SM) Higgs boson [3, 4, 5, 6]. The observation makes models with an extended Higgs sector more interesting, although the measured properties also constrain the models. A popular extension on the SM Higgs sector is an addition of a second complex doublet leading to a prediction of the existence of a charged Higgs boson.

CMS [7] has conducted the following searches for a light ($m_{H^+} < m_{\text{top}} - m_b$) charged Higgs boson in the top quark decays (τ_h denotes the hadronic decay of the τ lepton):

- τ_h+jets final state for the $H^+ \to \tau\nu_\tau$ decay channel [8, 9], and

- μ+jets final state for the $H^+ \to c\bar{s}$ decay channel [10, 11],

and the following searches for a heavy ($m_{H^+} > m_{\text{top}} - m_b$) charged Higgs boson

- τ_h+jets final state for the $H^+ \to \tau\nu_\tau$ decay channel [8, 9], and

- $\mu + \tau_h$ and ee/eμ/$\mu\mu$ final states for the $H^+ \to \tau\nu_\tau$ and $H^+ \to t\bar{b}$ decay channels [12, 13]

using the proton-proton collision data at $\sqrt{s} = 8$ TeV recorded in 2012 corresponding to an integrated luminosity of $19.7\,\text{fb}^{-1}$. Important parts of the analyses of τ_h+jets and $\mu + \tau_h$ final states are the data-driven background measurements, which are described in this note.

In the τ_h+jets final state analysis the background is divided into multijet events, and events from various electroweak processes, like W+jets, Z+jets, diboson (WW, WZ, ZZ) as well as from the SM $t\bar{t}$ and tW production. The latter background is further divided into two parts: the first one labelled "EWK+$t\bar{t}$ with τ_h" consists of events where the selected τ_h originates from a hadronically decaying τ lepton, and the second one labelled "EWK+$t\bar{t}$ no τ_h" consists of events where an electron, a muon, or a jet is misidentified as the τ_h. Of these backgrounds the multijet and "EWK+$t\bar{t}$ with τ_h" backgrounds are measured from data as described in Sections 2.1 and 2.2, respectively. The "EWK+$t\bar{t}$ no τ_h" background is estimated using simulation.

In the $\mu + \tau_h$ final state analysis the background arises from two sources: misidentified τ_h, and genuine τ_h. The misidentified τ_h background is measured from data, as is also the $Z \to \tau\tau \to \tau_h\mu$ component of the genuine τ_h background. The measurements are described in Sections 3.1 and 3.2, respectively. The remaining genuine τ_h background is estimated using simulation.

The backgrounds in the remaining analyses, i.e. of μ+jets and ee/eμ/$\mu\mu$ final states, are estimated using simulation. Common themes in the background measurements are the fake-rate technique for jets misidentified as the τ_h, and the embedding technique for events with genuine τ_h.

2. Background measurements in the τ_h+jets final state analysis

2.1 Measurement of the multijet background

The multijet background is measured with a "τ_h fake-rate" technique. First, the probability of the τ_h candidate to pass the nominal and inverted τ_h isolation criteria is measured in bins of

τ_{h} transverse momentum in an event sample obtained prior applying the $E_{\mathrm{T}}^{\mathrm{miss}}$, b tagging, and $\vec{p}_{\mathrm{T}}^{\tau_{\mathrm{h}}}$–$\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ back-to-back parts of the event selection. The event sample that passes the nominal τ_{h} isolation selection contains a non-negligible contribution from EWK+$t\bar{t}$ events. Therefore the numbers of multijet and EWK+$t\bar{t}$ events are determined by maximum likelihood fit of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. One such fit of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is performed per $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ bin. The $E_{\mathrm{T}}^{\mathrm{miss}}$ templates are obtained for multijet events from the control sample by subtracting simulated EWK+$t\bar{t}$ events from data, and for EWK+$t\bar{t}$ events from simulation in the nominal region. Figures 1 (a–g) show the sum of the multijet template and EWK+$t\bar{t}$ template after the fit to data in the nominal region for all $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ bins. The measured fake-rate probabilities are shown in Fig. 1 (h). The measured fake-rate probabilities are applied as weights to multijet events, obtained by subtracting simulated EWK+$t\bar{t}$ events from data passing all selection criteria of the signal region, except that the τ_{h} isolation criteria is inverted.

As a cross-check, the transverse mass (m_{T}) distribution obtained with the τ_{h} fake-rate technique is compared to the m_{T} distribution obtained with the nominal τ_{h} isolation criteria in a control region dominated by multijet background, defined by inverting the b-tagging requirement. The comparison is depicted in Fig. 2. The transverse mass distribution for multijet events obtained with the τ_{h} fake-rate method is found to agree with the m_{T} spectrum obtained with nominal τ_{h} isolation criteria in this control region.
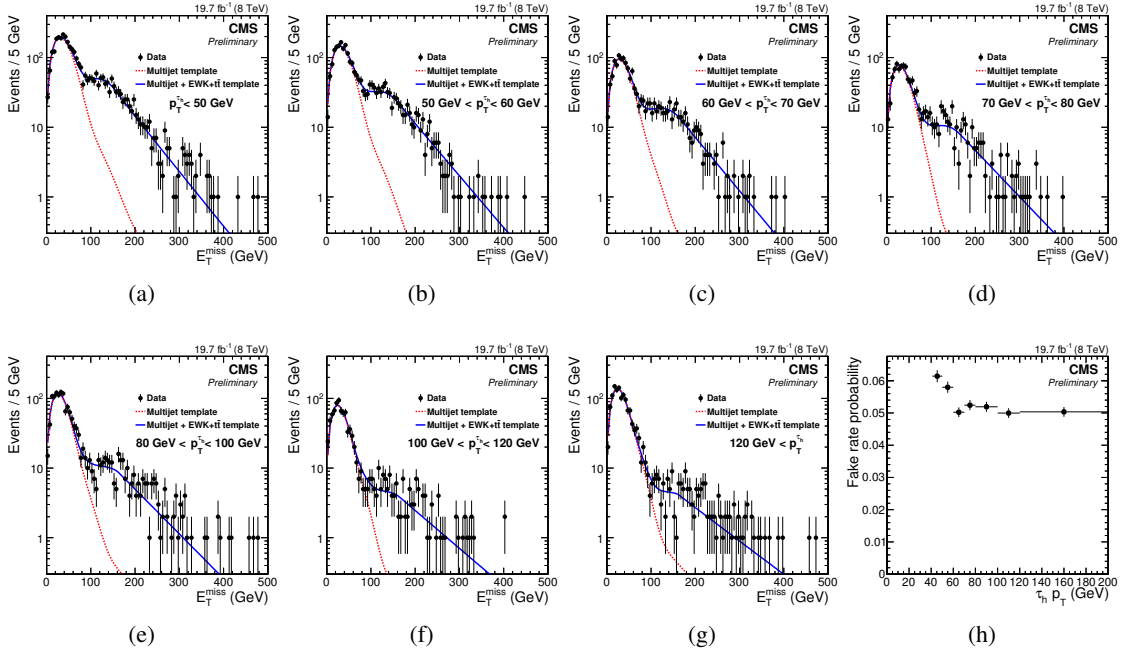

Figure 1: (a–g) The multijet template (red dashed line), the EWK+$t\bar{t}$ template on top of the multijet template (blue solid line), and the data distribution in the nominal region (black dots) for deriving the number of multijet events in the nominal region in bins of $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$. (h) The fake-rate probability in $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ bins. The vertical bars represent statistical uncertainty, and the horizontal bars denote the binning.
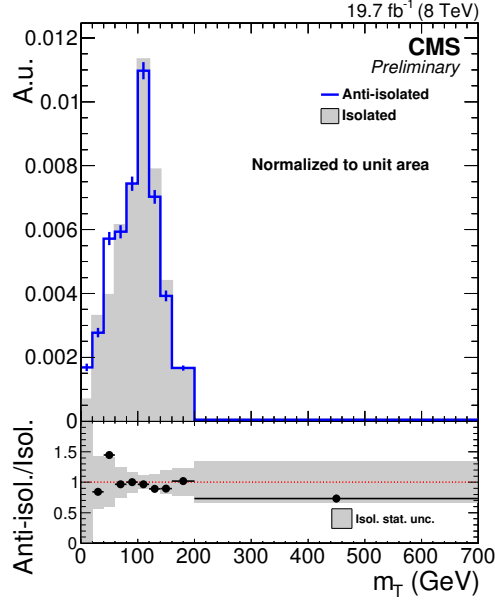
Figure 2: Comparison of shapes of the transverse mass distributions from the isolated (filled histogram) and anti-isolated (solid line) τ_h samples in the inverted b tagging control region. The vertical bars and light grey area represent statistical uncertainty, and the horizontal bars denote the binning.

2.2 Measurement of the "EWK+$t\bar{t}$ with τ_h" background

The m_T distribution for the "EWK+$t\bar{t}$ with τ_h" background is modelled via the so-called τ embedding technique. A control sample of $\mu + \geq 3$ jets is selected by requiring one muon ($p_T > 41$ GeV, $|\eta| < 2.1$ and at least three jets ($p_T > 30$ GeV, $|\eta| < 2.4$). Events with additional muons or electrons are rejected. The selected μ is replaced by a simulated τ lepton decay, where the τ lepton momentum is set to be the same as the momentum of the μ. The reconstructed τ decay products are merged with the original μ+jets event, from which the reconstructed muon is removed. These hybrid events are then subjected to the same event selection as the the τ_h+jets events, i.e. τ_h identification, b tagging, E_T^{miss} requirement, and the azimuthal angle selections.

The normalization of the "EWK+$t\bar{t}$ with τ_h" background needs to be corrected for the muon trigger and identification efficiencies, and for the $\tau+E_T^{miss}$ trigger efficiency. The efficiency of the τ part of the trigger is taken into account by weighting the events, while the E_T^{miss} part is included by approximating the E_T^{miss} in the trigger with a hybrid calorimetric E_T^{miss} constructed from the original event and the simulated τ lepton decay. In a fraction of the selected μ+jets events the μ originates from a decay of a τ lepton, leading to an overestimation of the background by a few percent. The bias is corrected for using a p_T^{μ}-dependent event weights derived from simulated $t\bar{t}$ events. A difference is seen in the transverse mass distribution between non-embedded τ_h+jets and embedded μ+jets events in simulated $t\bar{t}$ events, which constitutes about 85 % of the "EWK+$t\bar{t}$ with τ_h" background. The difference is corrected for with m_T-dependent event weights derived from simulated $t\bar{t}$ events.

To validate the method, simulated μ+jets events are passed through the embedding procedure and compared to the sum of the expected events from the simulation with the τ_h+jets selection and genuine τ_h. The comparison is shown in Fig. 3, where the m_T distribution from the embedded events is found to be in agreement with the same distribution from the τ_h+jets events with genuine τ_h. In this comparison only the uncertainties that affect each event sample independently of the other are considered.
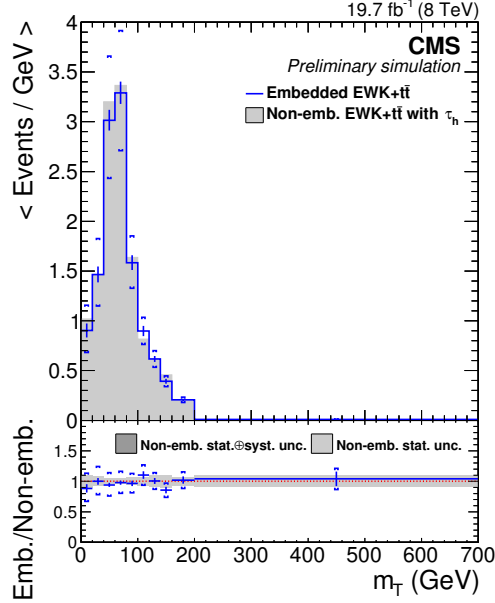

Figure 3: The m_T distribution for the "EWK+t$\bar{\text{t}}$ with τ_h" background for validating the embedding procedure. The simulated μ+jets events are passed through the embedding procedure (solid line) and compared to the simulated τ_h+jets with a genuine τ_h (filled histogram). Only the systematic uncertainties that affect only one of them are shown here. The hatched area, light grey area, and vertical black bars depict the statistical uncertainty, while the dark grey area and the brackets at the end of the vertical black lines represent the total uncertainty. The horizontal bars denote the binning.

3. Background measurements in the μ+τ_h final state analysis

3.1 Measurement of the misidentified τ_h background

The background from events where a jet is misidentified as the τ_h is measured with a fake rate technique. A probability for a jet to be misidentified as a τ_h (jet-to-τ_h probability) is measured from data separately for W+jets events and multijet events. The jets in the former are primarily quark jets, while in the latter the jets are dominantly gluon jets.

The W+jets events are triggered with a single isolated muon trigger ($p_T > 20\,\text{GeV}$ and $|\eta| < 2.1$), and the events are required to have at least one jet ($p_T > 20\,\text{GeV}$ and $|\eta| < 2.4$) and the transverse mass reconstructed from the muon and E_T^{miss} is required to be greater than $50\,\text{GeV}$. The
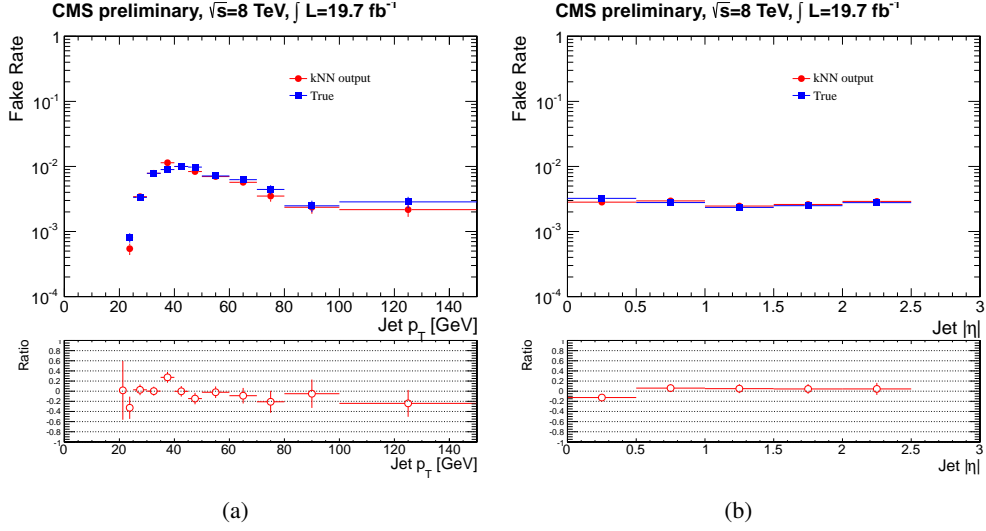
Figure 4: The probability of a jet to be misidentified as a τ_h as a function of the jet (a) p_T and (b) η evaluated from W+jets type events. The estimated fake rate is shown as red dots, and the fake rate from simulation truth is shown as blue squares. The horizontal bars denote the binning.

multijet events are triggered with a single jet trigger ($p_T > 30\,\text{GeV}$), and are required to have at least two jets ($p_T > 20\,\text{GeV}$, $|\eta| < 2.4$). All jets except the one firing the trigger are used in the misidentification rate calculation in order to avoid bias from the trigger. However, if two or more jets fire the trigger, all jets are used to calculate the misidentification rate. The jet-to-τ_h probability was parameterized as a function of the jet p_T, η, and radius $R = \sqrt{\sigma_{\eta\eta}^2 + \sigma_{\phi\phi}^2}$ using the k-Nearest Neighbour (kNN) regression [14]. The misidentification rates are shown in Fig. 4 for W+jets events and multijet events.

The misidentified τ_h background is then estimated as follows. A set of $\mu + \geq 3$ jets events is selected by requiring 1 isolated muon, E_T^{miss}, at least three jets, and at least one b-tagged jet. The selection thresholds are the same as in the signal event selection [12, 13]. This sample is dominated by W+jets and $t\bar{t} \to \mu\nu + q\bar{q}'$ events. The events are weighted by applying the jet-to-τ_h probability to each jet in the events. The selected events contain a small contribution from genuine τ_h events. The amount of these events is estimated using simulation and subtracted. The raw estimates are weighted to take into account the quark and gluon jet composition in these $\mu + \geq 3$ jets events.

3.2 Measurement of the $Z \to \tau\tau \to \tau_h\mu$ background

The background from events with $Z \to \tau\tau \to \tau_h\mu$ decays, where the hadronically decaying tau is identified as the τ_h, and the tau decaying to μ is identified as the μ, is measured with tau embedding technique. The shape of the b-tagged jet multiplicity distribution is taken from $Z \to \mu\mu$ events in data where both muons have been replaced with reconstructed particles from simulated τ lepton decays. The normalization of the distribution is taken from simulation of $Z \to \tau\tau$ events.

4. Summary

The data-driven background measurement methods employed in the searches for a charged Higgs boson in the CMS experiment were reviewed. The fake-rate technique is used for measuring the multijet background in the τ_h+jets final state analysis, and for measuring the background from jets misidentified as the τ_h in the μ+τ_h final state analysis. The tau embedding technique is employed for the measurement of the "EWK+$t\bar{t}$ with τ_h" background in the τ_h+jets final state analysis, and for measuring the background from $Z \to \tau\tau \to \tau_h\mu$ events in the μ+τ_h final state analysis.

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) 30–61.

[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) 1–29.

[3] CMS Collaboration, *Observation of a new boson with mass near 125 GeV in pp collision at $\sqrt{s} = 7$ and 8 TeV*, *JHEP* **06** (2013) 081.

[4] CMS Collaboration, *Study of the Mass and Spin-Parity of the Higgs Boson Candidate via its Decay to Z Boson Pairs*, *Phys. Rev. Lett.* **110** (2013) 081803.

[5] ATLAS Collaboration, *Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC*, *Phys. Lett. B* **726** (2013) 88.

[6] CMS Collaboration, *Measurement of the properties of a Higgs boson in the four-lepton final state*, *Phys. Rev. D* **88** (2014) 092007.

[7] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[8] CMS Collaboration, *Search for charged Higgs bosons with the $H^+ \to \tau^+\nu_\tau$ decay channel in the fully hadronic final state at $\sqrt{s} = 8$ TeV*, *CMS Physics Analysis Summary* **CMS-HIG-14-020**, 2014.

[9] L. Wendland, *CMS H^+ searches, part 2*, these proceedings.

[10] CMS Collaboration, *Search for a light charged Higgs boson in the $H^+ \to c\bar{s}$ channel at CMS*, *CMS Physics Analysis Summary* **CMS-HIG-13-035**, 2014.

[11] G. Kole, *$H^+ \to cs$ in CMS*, these proceedings.

[12] CMS Collaboration, *Search for a heavy charged Higgs boson in proton-proton collisions at $\sqrt{s} = 8$ TeV with the CMS detector*, *CMS Physics Analysis Summary* **CMS-HIG-13-026**, 2014.

[13] P. Vischia, *CMS H^+ searches, part 1*, these proceedings.

[14] A. Hoecker et al., *TMVA Users Guide*, arXiv:physics/0703039, 2009.