IL NUOVO CIMENTO Vol. 36 C, N. 4 Luglio-Agosto 2013
DOI 10.1393/ncc/i2013-11537-y

Communications: SIF Congress 2012

Search for the standard model Higgs boson in associated WH production in the $e\mu\tau$ and $\mu\mu\tau$ final states

M. T. Grippo(*) on behalf of the CMS Collaboration
Università di Siena, and INFN, Sezione di Pisa - Italy

ricevuto il 28 Dicembre 2012

Summary. — The Standard Model (SM) Higgs boson is mainly produced from gluon-gluon and vector boson fusion at LHC. The associated production with vector bosons, although with a lower cross section, can be also considered a sensitive channel because a significant background rejection can be achieved using the presence of highly energetic charged leptons coming from the decays of W/Z. In the light-mass region, the SM Higgs boson decay into τ-lepton pairs has the second highest branching ratio, after the decay into $b\bar{b}$. For these reasons, a search for WH process is performed, in which the W boson decays into muon or electron, and the Higgs boson into a τ-pair, in which one τ decays into leptons and the other hadronically (τ_h). The analysis is based on data collected with the CMS detector during 2011, corresponding to an integrated luminosity of $4.7\,\text{fb}^{-1}$ at a center-of-mass energy of 7 TeV at LHC. The results are consistent with the expected SM background, so upper limits are set at 95% CL for the SM Higgs boson production cross section. The method used to estimate the background from data, is based on the fake-rate technique (CMS Physics Analysis Summary, CMS PAS HIG-12-006 (2012)).

PACS 07.05.Hd – Data acquisition: hardware and software.
PACS 14.80.Bn – Standard-model Higgs bosons.
PACS 29.85.Fj – Data analysis.

1. – Description of the CMS detector

The Compact Muon Solenoid (CMS) is a multi-purpose detector, present at the Large Hadron Collider (LHC), built at CERN (European Organization for Nuclear Research), near Geneva.

CMS is designed in order to investigate a wide scenery of new Physics; it has a cylindrical symmetry with respect to the beam and is composed of a central part, called *barrel*

(*) E-mail: grippomariateresa@gmail.com

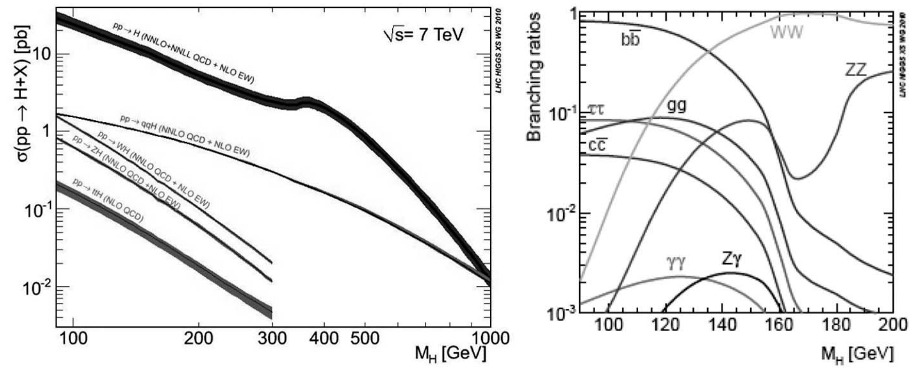

Fig. 1. – The SM Higgs boson production cross section (left) and branching ratios (right) as a function of the Higgs boson mass for different processes.

and of two disks, called *endcaps*. As suggested by the name, its two most important features are the superconducting solenoid, which provides a nominal magnetic field of 3.8 T, and a compact system of chambers to reveal μ. CMS is composed of different subdetectors: starting from the interaction point, it can be found the inner tracking system (silicon microstrip detector and silicon pixel detector), the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and muon detectors (MD), made of Resistive Plate Chamber (RPC) and Drift Tube (DT) in barrel and of Cathode Strip Chamber (CSC) and RPC in endcap.

CMS granularity and great performances provide a good efficiency to detect and reconstruct particles created during pp collisions. A more detailed description of CMS can be found in [1].

2. – Physics overview

One of the most important question in high-energy physics is the origin of masses of the particles, explained by the Standard Model introducing the existence of a new boson, called Higgs boson, which provides the masses through the spontaneous breaking of the electroweak symmetry [2].

The SM Higgs boson is produced in different processes at LHC (fig. 1, left): the process with the highest cross section is gluon-gluon fusion, followed by vector boson fusion. The associated production with vector bosons W/Z has a lower cross section, but can be considered a sensitive channel thanks to the presence of high energetic leptons, coming from W/Z [3]. Moreover the SM Higgs boson decays into τ-lepton pairs, in the region $M_H < 140\,\text{GeV}$, is the second highest branching ratio, approximately equal to 8% (fig. 1, right), after the decay $H \to b\bar{b}$, which is more difficult to reconstruct due to the predominant QCD background [3].

For these reasons, a search is performed for a SM Higgs boson, produced in association with a W, which decays into leptons, to suppress the QCD background, while the Higgs boson decays into a τ-pair, in which one τ decays into leptons and the other hadronically (τ_h). This analysis is also sensitive to Higgs decays into W bosons pairs, where all three W bosons decay to leptons.

TABLE I. – *Kinematic selections applied to the final state objects in the different channels.*

Channel	Object	p_T	η
	Leading μ	$> 20\,\text{GeV}$	< 2.1
$\mu\tau\tau$	Sub-leading μ	$> 10\,\text{GeV}$	< 2.1
	τ_h	$> 20\,\text{GeV}$	< 2.3
	μ	$> 20\,\text{GeV}$	< 2.1
$e\tau\tau$	e	$> 10\,\text{GeV}$	< 2.5
	τ_h	$> 20\,\text{GeV}$	< 2.3

3. – Event reconstruction and selection

This analysis has used all CMS data collected in 2011 at 7 TeV center-of-mass energy, corresponding to an integrated luminosity of $4.7\,\text{fb}^{-1}$ and has considered two trilepton channels: events with an e, μ of the same charge and a τ_h oppositely charged in the final state ($e\mu\tau$ channel), and events with two like-signed μ and an oppositely charged τ_h ($\mu\mu\tau$ channel) [4].

After the online trigger selection, candidate events are identified applying suitable selections to all channels, which match the expected final state.

The trigger requirements applied to each channel depend on the increasing istantaneous luminosity, in fact the thresholds change for different run periods, and they require two muons for $\mu\mu\tau$ channel and one high-p_T muon and one electron for $e\mu\tau$ channel.

The identification methods for muon, electron and τ_h leptons are identical to those used for the analysis $H \to \tau\tau$, described in [5]. The algorithm used to reconstruct objects, like muons and electrons, is the Particle Flow technique which combines the information from all CMS sub-detectors to identify and reconstruct individual particles in the event [6]. To avoid that a jet is incorrectly identified as a muon or electron, a Particle-Flow–based isolation is required for muons and electrons, using an isolation cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.4$ about the lepton candidate, corrected for pile-up contamination using the $\Delta\beta$ technique [5].

Hadronic τ are reconstructed through their decay products; the algorithm used in CMS is called "Hadron Plus Strip" (HPS) [7, 8], which reconstructs the decay mode of τ_h using the electromagnetic and hadronic information. Since the jets produced by taus are highly collimated, τ_h candidates are required to fulfill an isolation criterion [5]; the isolation is corrected for pile-up effects using the $\Delta\beta$ method. Moreover some criteria are required to avoid that an electron or muon is identified as a τ_h.

The kinematic requirements for light leptons, μ and e, in each channel are fixed by the trigger thresholds. The τ_h candidate has always a transverse momentum greater than 20 GeV. The detail of the kinematic selection are reported in table I.

The probability for a quark or gluon jet to pass the hadronic tau identification and isolation ("fake" jet$_\tau$) is higher than the probability for the same jet to pass the electron or muon identification and isolation requirements (10 to 100 times greater). In order to remove the background due to fake jet$_\tau$, in this analysis the absolute sum of the charges of the three final state objects is required to be one. That is, in the $\mu\mu\tau$ channel the two muons must have the same charge, which removes the large background $Z/\gamma \to \mu^+\mu^-$ + fake jet$_\tau$; in the $e\mu\tau$ channel the muon and the electron must have the same charge, removing $t\bar{t}$ and $Z \to \tau_e\tau_\mu$ + fake jet$_\tau$ backgrounds, where τ_e and τ_μ are the leptonic decays of τ in e and μ, respectively.

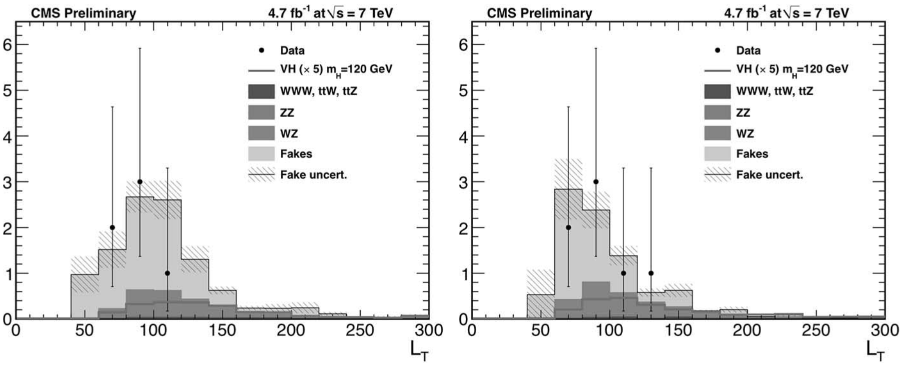

Fig. 2. – Distribution of L_T observable for $e\mu\tau$ (left) and $\mu\mu\tau$ (right) channels after all cuts except the L_T selection. The fake background is estimated using the data-driven technique, described in sect. **4**.

A new variable is introduced called L_T, which is the scalar sum of the transverse energy of the three lepton candidates in the event. It is required in this analysis $L_T > 80\,\text{GeV}$, in order to suppress the background, which has a softer p_T spectrum compared to the signal one. The distribution for both channels, after all other selections, is shown in fig. 2. Finally events with additional identified and isolated electrons, muons and b-jets are rejected, in order to reduce the contribution of ZZ and $t\bar{t}$ backgrounds.

4. – Background estimation

The background contribution is due to three different types of processes. The largest irreducible background comes from diboson events (WZ and ZZ), which contain three isolated leptons, e, μ or τ_h, in the final state. The reducible background comes from W, Z, $t\bar{t}$ and QCD events, in which a quark or gluon jet is incorrectly identified as an isolated e or μ ("fake" background). Finally, there is a contribution from the rare SM processes WWW and $t\bar{t}W$, which contain isolated leptons in the final state.

The irreducible backgrounds, ZZ and WZ, are estimated using Monte Carlo (MC) simulations, respectively based on PYTHIA [9] and MADGRAPH [10] event generators. They are normalized using the NLO theoretical prediction [11], confirmed by the 2011 CMS measurements of WZ and ZZ cross sections [12]. Also the rare processes WWW and $t\bar{t}W$ are estimated using events simulated by MADGRAPH, using the leading order prediction of the cross section.

The fake backgrounds are estimated using the "fake rate method", characterized by incorrectly identified or non-isolated leptons which pass all the final selection cuts. First of all the method measures the probability $f(p_T)$ for a lepton "inside" the jet, like a μ or e, which passes a *loose* selection, to pass a *tight* selection. The *loose* selection contains all relevant cuts, used in the analysis, with the exception of the final identification and the PF isolation of that object. The *tight* selection is the final analysis selection.

This probability, parametrized by jet p_T, is measured in selected background-enriched control regions, in order to be exclusive of the signal region, to be as close as possible to the signal selections to avoid any possible bias and to have a low contamination from processes with real isolated leptons. The three background enriched control regions are: $W \to \mu\nu + jet$, $Z \to \mu\mu + jet$ and QCD events (enriched of anti-isolated muons), selected

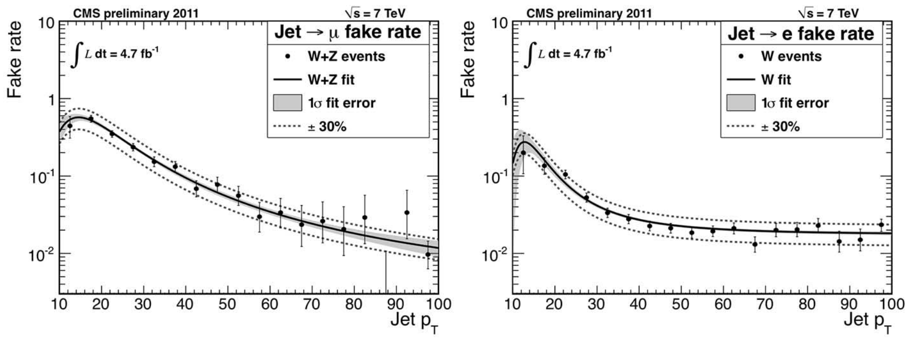

Fig. 3. – Jet to muon fake rate (left) and jet to electron fake rate (right) measured in W/Z events *versus* jet p_T.

with the same triggers used in the analysis and with a particular selection to enrich the sample of "fake objects". The fake rates for the muon and electron are measured in the three control regions; two examples are reported in fig. 3.

The fake background contribution, in the isolated signal region, can be estimated considering the events which pass the *loose* selection, but fail the identification and isolation cuts, called "anti-isolated" events and dominated by fake objects. In particular the expected background is evaluated by weighting the anti-isolated events by the corrected probability $p = f(p_T)/(1 - f(p_T))$. In general, the fake rate is different for electroweak and QCD processes and this is also considered in the analysis.

The charge requirement ($e^\pm\mu^\pm\tau_h^\mp$ and $\mu^\pm\mu^\pm\tau_h^\mp$ respectively for $e\mu\tau$ and $\mu\mu\tau$ channels) determines which are the "fakeable objects": in the $e\mu\tau$ channel, the electron and muon; in the $\mu\mu\tau$ channel the two muons, categorized as leading and sub-leading with respect to the p_T.

In each channel there are different exclusive categories of backgrounds. For example in the $e\mu\tau$ channel there are backgrounds with only one fake object and two isolated leptons, like $Z \to \tau\tau \to e\tau_h + jet_\mu$ and $Z \to \tau\tau \to \mu\tau_h + jet_e$, where jet_μ and jet_e represent the possibility that a jet is incorrectly identified, respectively, as a muon and an electron. These backgrounds are estimated by anti-isolating the muon for the first and the electron for the second and weighting the events by the measured muon/electron fake rate; all contributions are summed together to estimate this kind of fake backgrounds.

Moreover there are backgrounds with both a fake electron and a fake muon, like QCD and $W \to \tau_h\nu + jet_\mu + jet_e$, which are estimated by both the electron and muon fake contribution. Summing together the two estimations there are "double counted" events, which are corrected by using the e and μ simultaneously as "fakeable objects" and applying the electron and muon weights. In order to have a correct estimation of the backgrounds the double fake contribution is subtracted from the electron and muon sum.

Finally there are backgrounds, like QCD, in which all three leptons are fake. This contribution is estimated using the region where all three leptons are simultaneously anti-isolated, and extrapolating under the signal region with the fake-rate probability.

The same procedure is used in the $\mu\mu\tau$ channel, replacing the "electron" fake with the "sub-leading muon" fake.

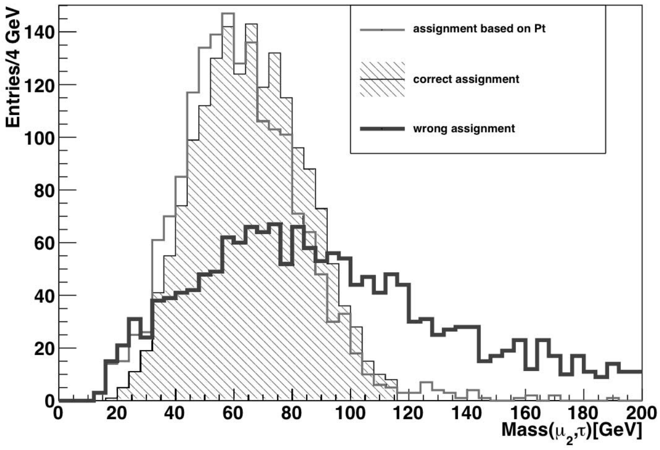

Fig. 4. – Invariant mass of the sub-leading lepton and τ_h for studies with *truth matching* technique.

5. – Signal WH studies with "truth matching" technique(1)

The selection described in sect. **3**, is designed for the signal WH topology, thinking that the leading lepton (greater p_T) comes from W decay, while the light lepton coming from the Higgs boson decay is expected to have lower p_T, as it is associated with more neutrinos. The validity of this assignment is studied with Monte Carlo events using the *truth matching* technique, in which reconstructed objects, that pass all selections, are compared with those reconstructed at generator level.

The first step is to identify at generator level the three leptons and then verify if there is a match among these objects and the reconstructed leptons passing all the selection, using the variable ΔR. The aim of this procedure is to understand how many times this assignment is correct and if this can influence the invariant mass distribution of the sub-leading lepton and τ_h, which is expected to be correlated with the invariant mass of the Higgs boson candidate.

As can be seen in fig. 4, the invariant mass of the sub-leading lepton and hadronic tau (called "assignment based on p_T") is almost identical to the true visible invariant mass of the Higgs boson (called "correct assignment"), confirming the hypothesis chosen for the analysis.

6. – Data and Monte Carlo corrections

MC samples are corrected for the effect due to the multiple proton-proton interactions per bunch-crossing, called "pile-up", reweighting the simulated events with the distribution of the real data. Moreover the MC samples are scaled using correction factors corresponding to the difference between the measured and simulated efficiencies for electron and muon reconstruction, identification, isolation and trigger, evaluated using the tag-and-probe method. Finally corrections to jet energy scale are also applied in data and MC samples.

(1) This is an extract of the master thesis of the author.

TABLE II. – *Number of events for each process in the $\mu\tau\tau$ channel and in the $e\tau\tau$ channel.*

Channel	$\mu\tau\tau$	$e\tau\tau$
Fakes	3.09 ± 1.03	5.64 ± 1.80
WZ	2.13 ± 0.35	2.12 ± 0.35
ZZ	0.17 ± 0.07	0.18 ± 0.07
Triboson	0.20 ± 0.20	0.20 ± 0.20
Backgrounds	5.60 ± 1.11	8.15 ± 1.84
Observed	5	4
$VH(120)$	0.40	0.38

7. – Results and systematic uncertainties

After all selections five events are observed in the $\mu\tau\tau$ channel. The expected results indicate that the background is dominated by $W + jets$, $Z/\gamma + jets$ and $t\bar{t}$ events, in agreement with observed data. In the $e\mu\tau$ channel a total of four events are observed; the simulated background is dominated by $t\bar{t}$ events, in agreement with real data.

The observed and expected results for each channel are reported in table II; the invariant mass of the sub-leading lepton and τ_h after all selection cuts is shown for both channels in fig. 5.

Limits on the Standard Model Higgs production are computed using the expected number of background events, the number of observed events and the expected number of signal events, performing a maximum-likelihood fit in the two channels.

The systematic uncertainties, used in the limits as nuisance parameters, are enumerated in table III.

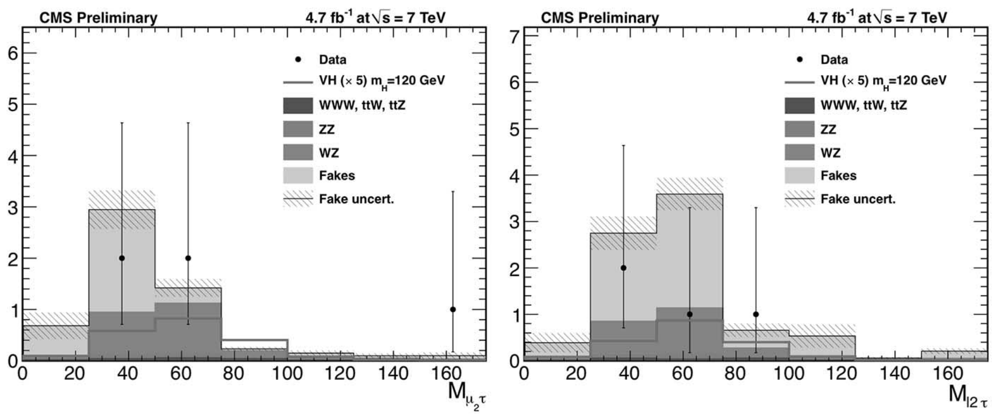

Fig. 5. – Invariant-mass distribution of the sub-leading lepton and τ_h in the $\mu\mu\tau$ channel (left), and the $e\mu\tau$ channel (right), after all selections. The plots show the combined fake rate estimate and its error; the sum of all backgrounds (diboson and fake) is given in the black line.

TABLE III. – *Systematic uncertainties applied in the analysis.*

Channel	Value
Luminosity [13]	4.5%
σ_{WZ} [12]	16.6%
σ_{ZZ} [12]	40%
$\sigma_{ttW,WWW}$	100%
σ_H(PDF) [3]	4.5%
MC Eff. L_T	5%
MC Stat.	5%–30%
Fakes	30% + stat. uncertainty
Tau energy scale [8]	1–2.5%
Tau ID [8]	6%
Muon ID + Iso	1%
Electron ID + Iso	2%

An uncertainty on the total integrated luminosity equal to 4.5% [13] is applied to MC samples. The uncertainties on the diboson backgrounds are taken from 2011 CMS cross section measurements [12]; for the rare SM process $t\bar{t}W$ and WWW a 100% uncertainty is applied on the cross section. An uncertainty of 4.5% due to the QCD scale and parton distribution function (PDF) [3] is applied on the signal cross section.

On signal and diboson events a 5% uncertainty is applied, related to the efficiency of the L_T selection. The uncertainties on the electron and muon identification, isolation and trigger efficiencies correspond to the uncertainties on the measured data - simulation correction factors, briefly described in sect. **6**. For the fake errors two contribution should be considered: the first is a 30% uncertainty, due to a possible error in the fit of the fake rate function and to a different estimation of the fake rate in the three background-enriched control regions; the second is an uncorrelated uncertainty applied to the fake rate, corresponding to limited statistics in the anti-isolated region.

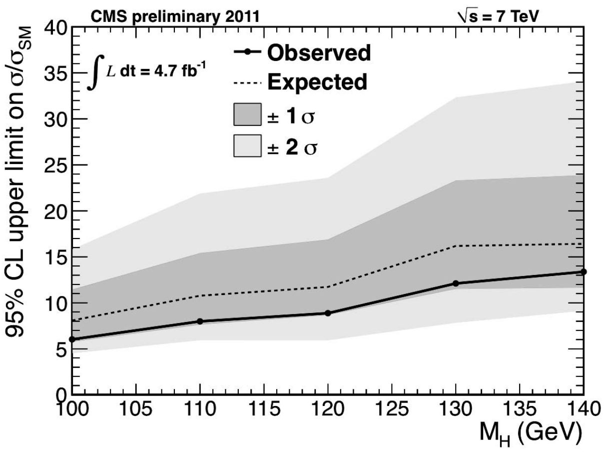

Fig. 6. – Observed and expected range of 95% CL upper limits on SM Higgs boson production using the combination of the two channels.

The τ_h identification uncertainty, equal to 6%, is estimated using the tag-and-probe method in $Z \to \tau\tau$ events [8]. Varying the τ_h energy scale within its 3% uncertainty, it is found to have a 1–2.5% effect on the total yield, depending on the simulated sample [8]. The electron energy scale uncertainty (1% in the barrel, 2.5% in the end caps) is found to not affect the analysis, so it is not used to compute the limit.

For the calculation of the limits is used a modified frequentist method, called CL_s method [14, 15], in which the expected number of signal and background events are used for each channel and the respective systematic uncertainties. The contribution from $H \to WW$ decays for $M_H > 120\,\mathrm{GeV}$ is also included. The limit is expressed in units of σ_{SM}, the expected SM Higgs boson cross section.

The observed events show no evidence for the presence of a Higgs boson signal and 95% CL upper limits are set on the Higgs boson cross section. In fig. 6 the limit on SM WH production *versus* Higgs boson mass is shown for the combination of the two channels.

8. – Conclusion

A search for Standard Model Higgs boson production in association with a W boson is performed using all data collected by CMS experiment in 2011, corresponding to an integrated luminosity of $4.7\,\mathrm{fb}^{-1}$ in pp collisions at $\sqrt{s} = 7\,\mathrm{TeV}$.

Two final states are explored: events with an e, a μ and a τ_h, and events with two muons and a τ_h, requiring the same charge for the two light leptons. The expected signal in the low-mass range considered ($M_H < 140\,\mathrm{GeV}$) is sensitive to $H \to \tau\tau$ and $H \to WW$ decays. A total of nine events are observed, compatible with the SM expected background. The background with fake objects is estimated using a data-driven method, called "fake rate" technique. The irreducible background due to diboson events is evaluated by CMS simulations. As no excess is observed, upper limits are set at 95% CL for the product of the SM Higgs production cross section and decay branching fraction in the mass range $100 < M_H < 140\,\mathrm{GeV}$. At $M_H = 125\,\mathrm{GeV}$ the observed limit is approximately ten times the predicted SM Higgs production cross section.

* * *

I would like to thank Giuseppe Bagliesi for his precious teachings and for his continuos help, also for this article; all group from INFN Bari and Pisa and SAHA Institute with whom I worked for this analysis and with whom I started a new experience.

REFERENCES

[1] CMS Collaboration, *JINST*, **3** (2008) S08004.
[2] Higgs P. W., *Phys. Lett.*, **12** (1964) 132; *Phys. Rev. Lett.*, **13** (1964) 508; *Phys. Rev.*, **145** (1966) 1156.
[3] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs cross sections: 1. Inclusive observables, CERN Report*, CERN-2011-002 (2011).
[4] CMS Collaboration, *Search for the standard model Higgs boson in associated WH production in the $e\mu\tau$ and $\mu\mu\tau$ final states*, *CMS Physics Analysis Summary*, **CMS PAS HIG-12-006** (2012) .
[5] CMS Collaboration, *Search for Neutral Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV, CMS Physics Analysis Summary*, CMS-PAS-HIG-11-029 (2011).

[6] CMS Collaboration, *Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and* E_T^{miss}, *CMS Physics Analysis Summary*, CMS-PAS-PFT-09-001 (2009); CMS Collaboration, *Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV, CMS Physics Analysis Summary*, CMS-PAS-PFT-10-002 (2010).

[7] CMS Collaboration, *Tau Identification in CMS, CMS Physics Analysis Summary*, CMS-PAS-TAU-11-001 (2011).

[8] CMS Collaboration, *Performance of tau lepton reconstruction and identification in CMS, JINST*, arXiv:hep-ex/1109.6034 (2011).

[9] Mrenna S. *et al.*, *JHEP*, **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026.

[10] Alwall J. *et al.*, *JHEP*, **09** (2007) 028, arXiv:0706.2334; doi:10.1088/1126-6708/2007/09/028.

[11] Williams C., Campbell J. and Ellis K., *MCFM - Monte Carlo for FeMtobarn processes*, `mcfm.fnal.gov`.

[12] CMS Collaboration, *Measurement of WW and observation of WZ and ZZ in leptonic modes, CMS Physics Analysis Summary*, CMS-PAS-EWK-11-010 (2011).

[13] CMS Collaboration, *Measurement of CMS Luminosity, CMS Physics Analysis Summary*, CMS-PAS-EWK-10-004 (2010).

[14] Junk T., *Nucl. Instrum. Methods A*, **434** (1999) 435443, doi:10.1016/S0168-9002(99)00498-2.

[15] Read A. L., *Modified frequentist analysis of search results (the CLs method), CERN Yellow Report*, CERN-2000-005 (2000) 81.