Determination of Fake Electrons, Muons and Taus in the tbH^+ Search for the 2lSS1tau Channel Using Template Fit Method – Summer Student Report

Amnah Alkandari

CERN Summer Student 2023

Supervisor: André Sopczak

IEAP CTU in Prague

Abstract

This study contributes to the search for the charged Higgs boson tbH^+. In this search with two same-sign electrically charged light leptons and one hadronically decaying tau (2ℓSS1τ channel), the $t\bar{t}$, $t\bar{t}W$, and $t\bar{t}Z$ backgrounds are dominant for fake lepton contributions. The estimation of fake events for these backgrounds by a data-driven method is particularly important. For the data-driven background determination, the Template Fit method is used. The resulting normalization factors including statistical and systematic uncertainties, for fake light lepton events from heavy flavour decays are 0.84±0.15 for electrons, 0.94±0.08 for muons, and 1.05±0.02 for taus.

1 Introduction

The search for the charged Higgs boson $tbH^+(H^+ \to Wh \to W\tau\tau\ )$ with performance optimisation for signal and background separation using machine learning with ATLAS data was studied previously [1] [2]. In this project, the Template Fit method is applied for determining the fake rate normalization factors based on the simulated shapes and recorded data. It is a contribution to the tbH^+ analysis based on the thesis by Martin Rames (Search for tbH^+ with Performance Optimisation for Signal and Background Separation Using Machine Learning with ATLAS Data) [2]. Technical details of this study are given in [3].

The estimate of fake light leptons is important for the analysis channel requiring two same-sign light leptons and one hadronically decaying tau. The fake rate determination cannot rely on Monte-Carlo (MC) simulation alone, therefore the Template Fit as a data-driven method is chosen [4].

In this search with two same-sign electrically charged light leptons and one hadronically decaying tau (2ℓSS1τ channel [5]), the $t\bar{t}W$, $t\bar{t}$ and $t\bar{t}Z$ background reactions are dominant sources of the fake leptons in the signal region (SR). For the data-driven background determination, the Template Fit method is used with the implementation of ATLAS Isolation and Fake Forum (IFF) tools [6]. For the display of signal, background and data, the TRExFitter package is used [7]. The study shows that the Template Fit method is well suited for the data-driven background determination. Systematic uncertainties are also studied.

A description is given of the estimation of the light non-prompt lepton background (also referred to as fakes) in the ee, $e\mu$ and $\mu\mu$ same-sign plus one tau-hadron channel using the Template Fit (TF) method.

Fake light lepton events are defined as events where at least one of the light leptons is not produced in decays of electroweak bosons but rather result from another source. These are heavy-flavor hadron decays, photon conversions, misreconstruction of particles such as hadronic jets, or prompt light leptons whose charge is misidentified. Fake leptons are often also referred to as electrons or muons which are not originating from the primary Interaction Point (IP). Most likely they are resulting from semi-leptonic B-meson decays. The B-meson travels a distance before it decays, and the b-quark converts into a c-quark and a lepton and corresponding neutrino. Thus, these leptons appear to originate not from the primary IP. The dominant origin of these fake lepton events is $t\bar{t}$ production as each t-quark decays into a b-quark. Further aspects of fake lepton events are discussed in Babar Ali's thesis [8].

In this analysis, the signal region (SR) is defined as having exactly two light leptons with the same electric charge and one hadronically-decaying tau lepton (2ℓSS1τ).

An important step is to verify the performance of the Template Fit method and to determine the related systematic uncertainty.

1.1 Template Fit

Since the non-prompt light lepton background in the 2ℓSS1τ channel is a mixture of leptons from semi-leptonic heavy-flavour (HF) decays and photon conversions. The HF contribution is dominant and a template method has been developed to estimate the background.

Within the template fit method, the normalization of the different "fakes" contribution templates, as given by the Monte Carlo of all processes contributing to non-prompt lepton background, are left free-floating in a fit to data, and these normalization factors are used to correct the fakes Monte Carlo estimates. The template fit method is a semi-data-driven method, i.e. it relies on the truth information from $t\bar{t}$, $Z{+}jets$, $W{+}jets$, $V{+}\gamma$, and single top Monte Carlo simulation to define different types of fake/non-prompt leptons, and on the general description of fakes kinematics by the simulation (shape of the distibutions).

The main contribution to non-prompt lepton background comes from $t\bar{t}$, ttW and $V + jets$ processes. Based on the truth classification of events containing a non-prompt lepton, the following main contributions are distinguished, and free-floating normalization factors (NF) are assigned to each of them:

- NF_e^{HF}: normalization factor applied to events with one non-prompt electron from B decay, C decay or light hadron (dominated by B decay).
- NF_μ^{HF}: normalization factor applied to events with one non-prompt muon from B decay, C decay or light hadron (dominated by B decay).
- NF_τ^{HF}: normalization factor applied to events with one non-prompt tau from B decay, C decay or light hadron (dominated by B decay).

With the blinded analysis, the templates are fitted to the data using control regions only. These control regions are orthogonal to the signal region. The normalization factors will be derived in one simultaneous fit including the signal region.

Six dedicated control regions have been developed in order to improve the constraint on the NFs of the internal and external conversions, and decrease the correlation between these two mainly in the electron channels. Events with same-sign di-leptons and at least one b-tagged jets are used to define the control regions.

The sub-leading lepton transverse momentum distribution is exploited to best discriminate among the NFs in the simultaneous template fit

2 Results from Template Fit

The preselection yields of the TF are compared with those of M. Rames [2], and good agreement is obtained for the preselection.

2.1 Heavy Flavour Fake Lepton

The event selection is prepared for three types of fakes (HF_e, HF_μ and HF_τ). The selection is the same except for the di-lepton requirement (ee, eμ, $\mu\mu$). The details of the event selection are given in Appendix **??**. The following regions (HFall, HFel, and HFmu) have a pre-selection of two leptons same sign, one hadronically decaying tau, having exactly two to three jets, and at least one b-jet.

2.2 HFall Region

The HFALL TT region requires leading and sub-leading lepton tight with no flavor requirements on the light lepton.The HFALL TT DF region has an additional requirement with different flavor leading and sub-leading leptons. This means if the leading lepton was an electron, the sub-leading lepton should be a muon and vice versa.. Whereas the HFALL TT SF region requires the same flavor leading and sub-leading lepton. This means if the leading lepton was an electron, the sub-leading lepton should also be an electron. The pre-fit and post-fits were based on the event distribution for the tau width over the tau transverse momentum Tables 1 and 2 show the change in the simulated events for the fakes and the decrease in the total background events after fitting.

Event	Pre-fit	Post-fit
$fake\ \tau$	10.5	11.1
$fake\ \mu$	2.7	2.6
$fake\ e$	13.1	12.0
$fake\ l + \tau$	1.2	1.2
$Total\ background$	45.1	43.7

Table 1: Comparison of simulated events before and after fitting in HFall TT DF region

Event	Pre-fit	Post-fit
$fake\ \tau$	10.1	10.6
$fake\ \mu$	4.1	3.9
$fake\ e$	2.7	2.5
$fake\ l + \tau$	1.3	1.3
$Total\ background$	43.3	42.7

Table 2: Comparison of simulated events before and after fitting in HFall TT SF region

2.3 HFel Region

The HFel region is an electron enhanced region which requires leading and sub-leading lepton to be an electron. The HFel MT region requires leading lepton medium and sub-leading lepton tight. The pre-fit and post-fits were based on the number of events. HFel TM region requires leading lepton tight and sub-leading lepton medium. The fit was made based on the distribution of events with the tau width over the tau transverse momentum. Tables 3 and 4 show the change in the simulated events for the fakes and the decrease in the total background events after fitting. Since it is an electron enhanced region, Table 4 shows the largest decrease in events occurred for the fake electron.

Event	Pre-fit	Post-fit
$fake\ \tau$	0.5	0.6
$fake\ e$	0.2	0.2
$fake\ l + \tau$	4.3	4.5
$Total\ background$	6.6	6.3

Table 3: Comparison of simulated events before and after fitting in HFel MT region

2.4 HFmu Region

The HFmu region is a muon enhanced region which requires leading and sub-leading lepton to be a muon. The HFmu MT region requires leading lepton medium and sub-leading lepton tight. The fit was made based on the number of events. The HFmu TM region requires a leading lepton tight and a sub-leading lepton medium. The data is fitted on the distribution of events

Event	Pre-fit	Post-fit
$fake\ \tau$	1.1	1.1
$fake\ e$	26.2	23.3
$fake\ l+\tau$	16.3	15.2
$Total\ background$	46.7	42.7

Table 4: Comparison of simulated events before and after fitting in HFel TM region

with the tau width over the tau transverse momentum. Tables 5 and 6 show the change in the simulated events for the fakes and the decrease in the total background events after fitting. Since it is a muon enhanced region, both tables show the largest decrease in events occurred for the fake muon.

Event	Pre-fit	Post-fit
$fake\ \tau$	0.5	0.5
$fake\ \mu$	18.1	17.4
$fake\ e$	0.1	0.1
$fake\ l+\tau$	11.6	11.7
$Total\ background$	32	31.4

Table 5: Comparison of simulated events before and after fitting in HFmu MT region

Event	Pre-fit	Post-fit
$fake\ \tau$	2.7	2.8
$fake\ e$	0.5	0.5
$fake\ l+\tau$	47.6	47
$Total\ background$	135.8	130.3

Table 6: Comparison of simulated events before and after fitting in HFel TM region

2.5 Fit Results

Figure 1 shows the obtained normalization factors before systematics are applied.

2.6 Signal Region

The pre-selection in the signal region requires two leptons same sign, one hadronically decaying tau, at least four jets, and at least one b-jet. A cross section of 0.001 pb is applied with a blinding threshold of 15%. Figure 2 shows the distribution of events with the tau transverse momentum. A comparison between this study and M. Rames' study [2] is shown in Table 7. The difference in events is due to the inclusion of fakes in this study, for example the ttW is significantly lower in this study due to it being one of the main contributors to the fakes.

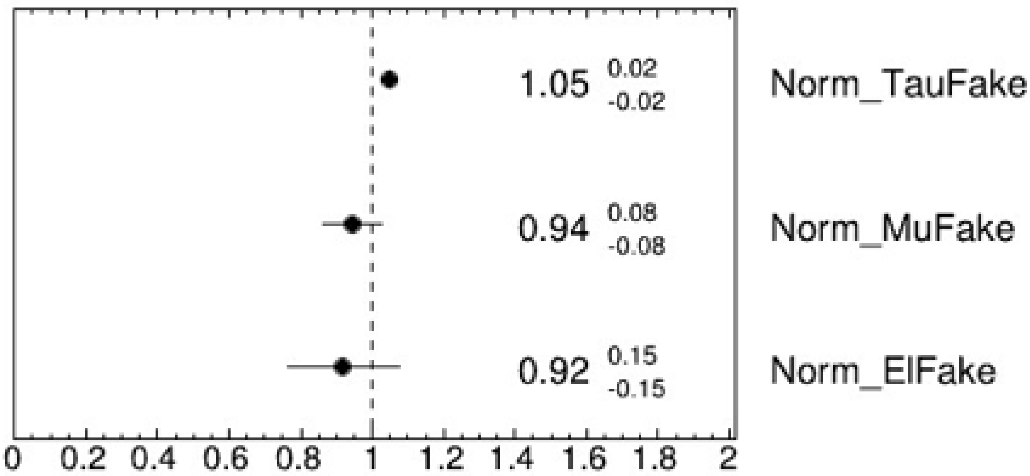

Figure 1: NormFactors for HF_e, HF_mu and HF_tau

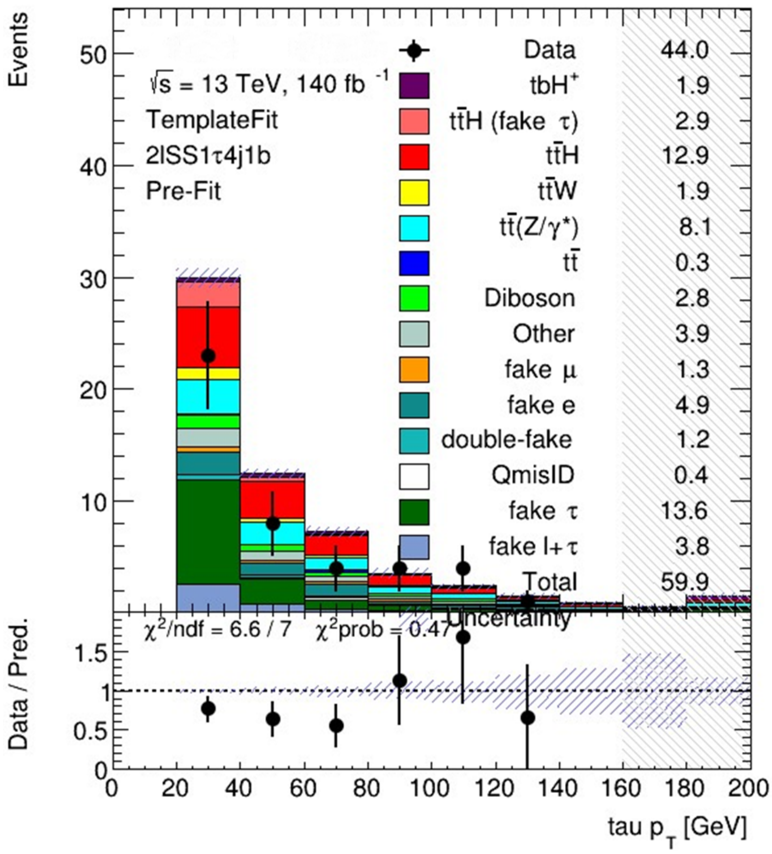

Figure 2: Signal region for cross section of 0.001 pb

Event	This Study	Martins Study
tbH^+	1869.4	1852.6
$t\bar{t}H$	12.9	12.8
ttW	1.9	8.8
$t\bar{t}$	0.3	7.4
$t\bar{t}Zd$	8.2	12.6
$fake\ \tau$	20.3	—
$fake\ \mu$	1.3	—
$fake\ e$	4.9	—
$fake\ l+\tau$	3.8	—
$Total\ Events$	1927.5	1903.1
$Total\ Background$	58	50.5

Table 7: Comparison of simulated events for this study and M. Rames study

3 Systematics

Figure 3 shows the obtained normalization factors after systematics are applied. There are no significant changes of the given uncertainties, however, for the electron normalization factor, the mean value changes within its uncertainty. This could be due to the fact that with systematics the yields have larger uncertainties, and thus the fit could converge to another minima.

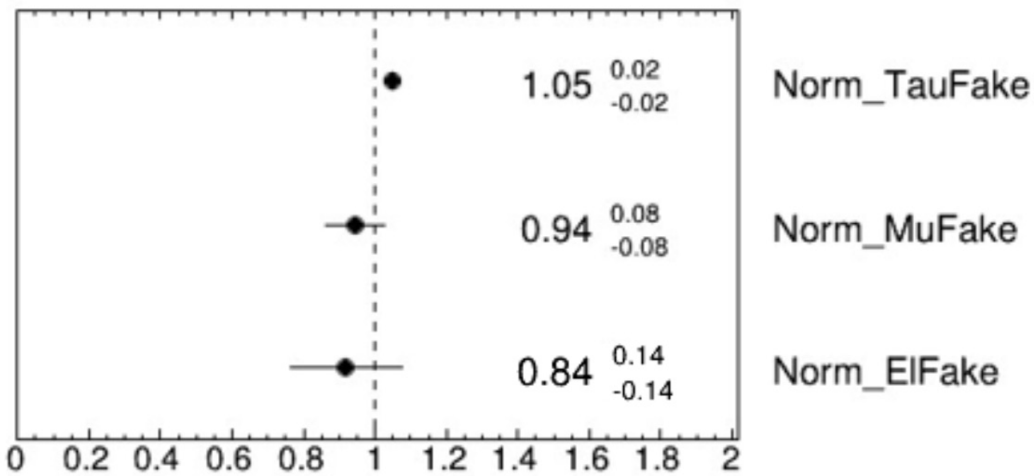

Figure 3: NormFactors for HF_e, HF_mu and HF_tau including systematics

3.1 Signal Region

Figure 4 shows the signal region with a tbH^+ cross section of 0.001 pb. The grey hatched region is blinded which indicates the last three bins are blinded. The blue hatched region indicates the statistical and systematic uncertainties. With the condition of 15% blinding threshold.

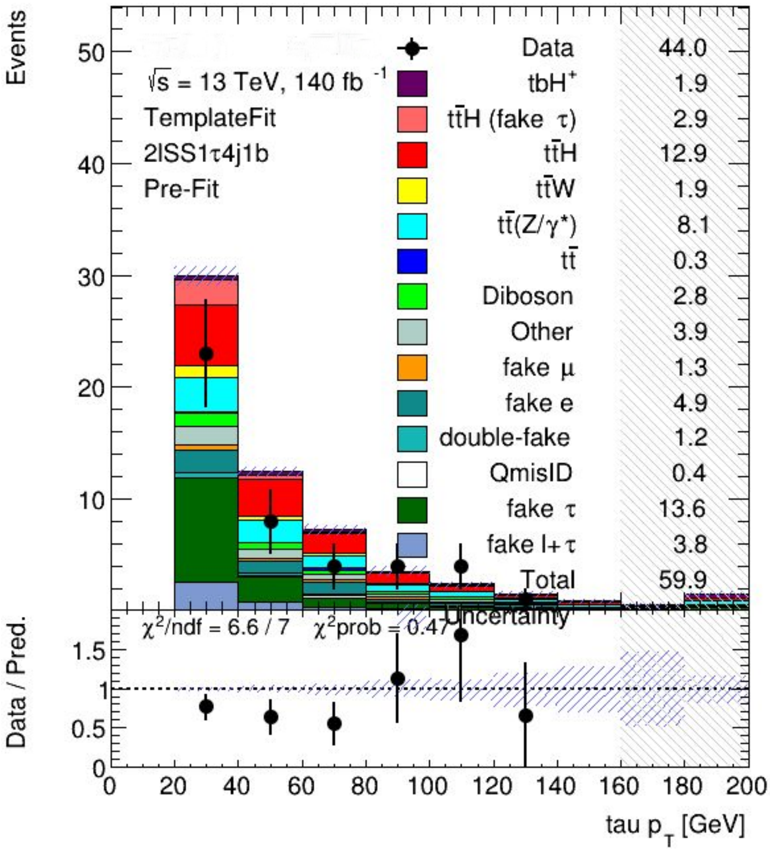

Figure 4: Signal region for tbH^+ cross section of 0.001 pb including systematics

4 Conclusion

In this summer student project, a contribution to a challenging search for the charged Higgs boson particle was made. The specific analysis channel contains two same-sign electrically charged light leptons and one hadronically decaying tau ($2\ell SS1\tau$). Three dedicated selection regions were used for the heavy flavour normalization factors determination with the requirement of at least one b-tagged jet. A systematic uncertainty study was performed. The total uncertainties are 15% for electron, 8% for muons and 2% for taus.

Acknowledgements

The author would like to thank the supervisor André Sopczak and CERN for making this summer student project possible. The discussions, advice, help and support from Luca Martinelli and Vladyslav Yazykov is warmly acknowledged.

References

1. Pospisil, J. *Application of Machine Learning for the Charged Higgs Boson Search Using ATLAS Data* CERN-THESIS-2022-065. Geneva, 2022. `https://cds.cern.ch/record/2812372`.
2. Rames, M. *Search for tbH^+ with Performance Optimisation for Signal and Background Separation Using Machine Learning with ATLAS Data* CERN-THESIS-2023-076. Geneva, 2023. `http://cds.cern.ch/record/2862250`.
3. Alkandari, A. & Sopczak, A. *Determination of Fake Electrons, Muons and Taus in the tbH^+ Search for the 2lSS1tau Channel Using Template Fit Method - Summer Student Project* ATL-COM-PHYS-2023-1074. Geneva, 2023. `https://cds.cern.ch/record/2882137`.
4. *Fake Lepton Backgrounds* `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/FakeLeptonBackgrounds` (2021).
5. Agaras, M. N. *et al. DiHiggs HH multilepton full Run2* ATL-COM-PHYS-2020-229. Geneva, 2020. `https://cds.cern.ch/record/2713728`.
6. Robl, J. B. & Sopczak, A. *Light Non-prompt (Fake) Lepton Background Estimation with the Matrix Method for HH Production with a Signature of Two Same Electric Charge Light Leptons Associated with a Tau - Summer Student Project* ATL-COM-PHYS-2021-974. Geneva, 2021. `https://cds.cern.ch/record/2789839`.
7. TRExFitter. `https://trexfitter-docs.web.cern.ch/trexfitter-docs`.
8. Ali, B. *Search for the Standard Model Higgs boson produced in association with a top quark pair in multilepton final states at $\sqrt{s}$ = 13 TeV with the ATLAS detector* Presented 28 Feb 2020 (Dec. 2019). `https://cds.cern.ch/record/2774418`.