Di-Higgs Boson Fake Lepton Study with a Signature of Two Same Electric Charge Light Leptons Associated with a Tau – Summer Student Report

Ahmed Hamidaddeen

CERN Summer Student 2023

Supervisor: André Sopczak

IEAP CTU in Prague

Abstract

The search for di-Higgs boson production HH probes the tri-linear coupling HHH predicted by the Higgs mechanism. In this search with two same-sign electrically charged light leptons and one hadronically decaying tau ($2\ell SS1\tau$ channel), the $t\bar{t}$ process is the dominant source of fake (non-prompt) leptons. These fake leptons result predominantly from heavy flavour (HF) reactions, and the normalisation factors for fake electrons and fake muons are determined by a data-driven method. This study updates and extends the 2022 summer student study with the addition of an optimized computing method and full systematics. For the data-driven background determination, the Template Fit method is used. The resulting normalisation factors from HF decays are 0.87 ± 0.10 for electrons and 0.75 ± 0.06 for muons, including systematic uncertainties. The dominant uncertainty is statistical.

1 Introduction

The data-driven fake rate determination in the HH analysis was studied with the Matrix Method 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. Technical details of the analysis are given the internal ATLAS note [3].

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

In this search with two same-sign electrically charged light leptons and one hadronically decaying tau (2ℓSS1τ channel), the di-vector boson (VV) and $t\bar{t}$ background reactions are the dominant sources of the fake leptons in the signal region (SR). The study shows that the Template Fit method is well suited for the data-driven background determination. Expected systematic uncertainties are also studied by varying the light lepton definitions.

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 muon 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.

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. For this purpose, the light lepton definitions were varied and the effect on the normalization factors were studied.

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, a template method has been developed to estimate these backgrounds.

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.

The main contribution to non-prompt lepton background comes from $t\bar{t}$, 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). ($\gamma* \to \ell\ell$). It originates mainly from $t\bar{t}\gamma$ and $V\gamma$ events.

Following the study with the two normalization factors, sensitivity was only obtained for the HF factors because of event statistics. Therefore, the TF method was then only used for the HF normalization factors.

While the analysis is blinded, the templates are fitted to the data using control regions only.

After unblinding the normalization factors are derived in one simultaneous fit including the signal region.

Two 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 control regions that are orthogonal to the signal region and that contain all the variety of non-prompt leptons.

The following distributions are exploited to best discriminate among the NFs in the simultaneous template fit:

- ΔR_{ll} in $\mu e + ee$ channel, to estimate NF_e^{HF}

- H_T in $e\mu + \mu\mu$ channel, to estimate NF_μ^{HF}

The main uncertainty corresponds to the statistical uncertainty on the NFs that varies from 6% (muons) to 10% (electrons).

2 Results From Template Fit

The event selection is prepared for two types of fakes (HF_e and HF_μ). The selection is the same except for the di-lepton requirement (ee, eμ, $\mu\mu$). Compared to the Signal Region (SR), the number of b-jet requirement is inverted, thus requiring at least one b-jet. The details of the event selection are in the ATLAS internal note.Figure 1 shows the obtained normalisation factors, and the ΔR_{ll} distribution for the CR is shown in Fig. 2, and the H_T for the CR is shown in Fig. 3.

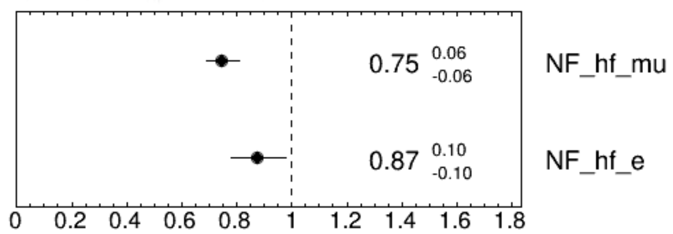
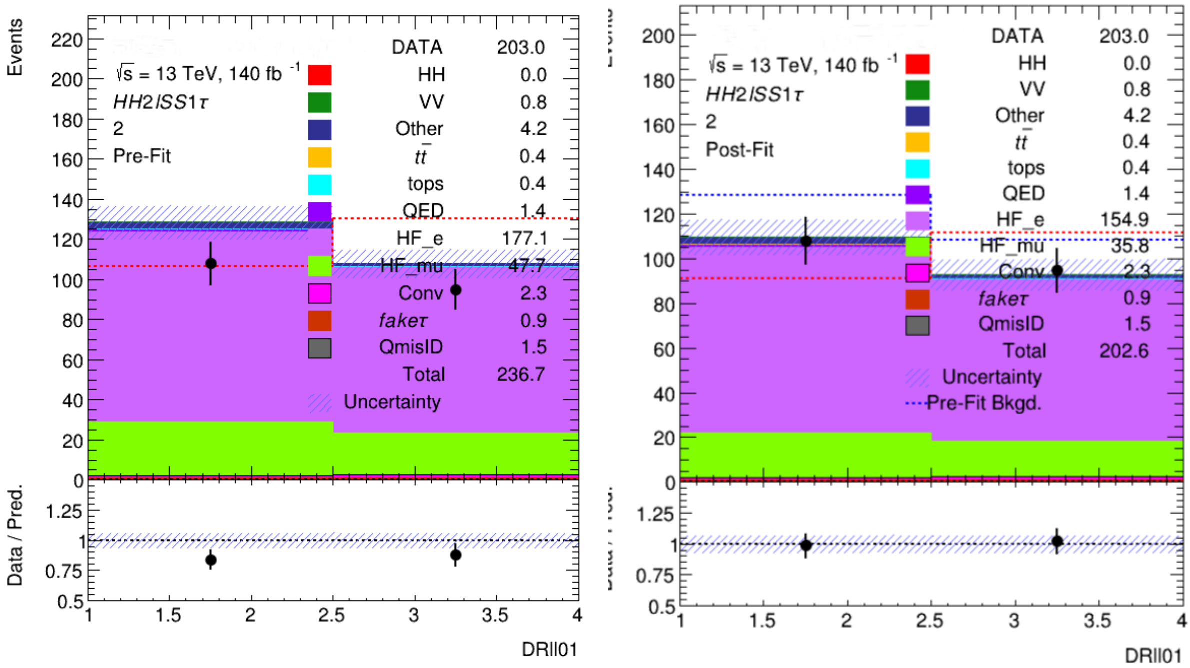
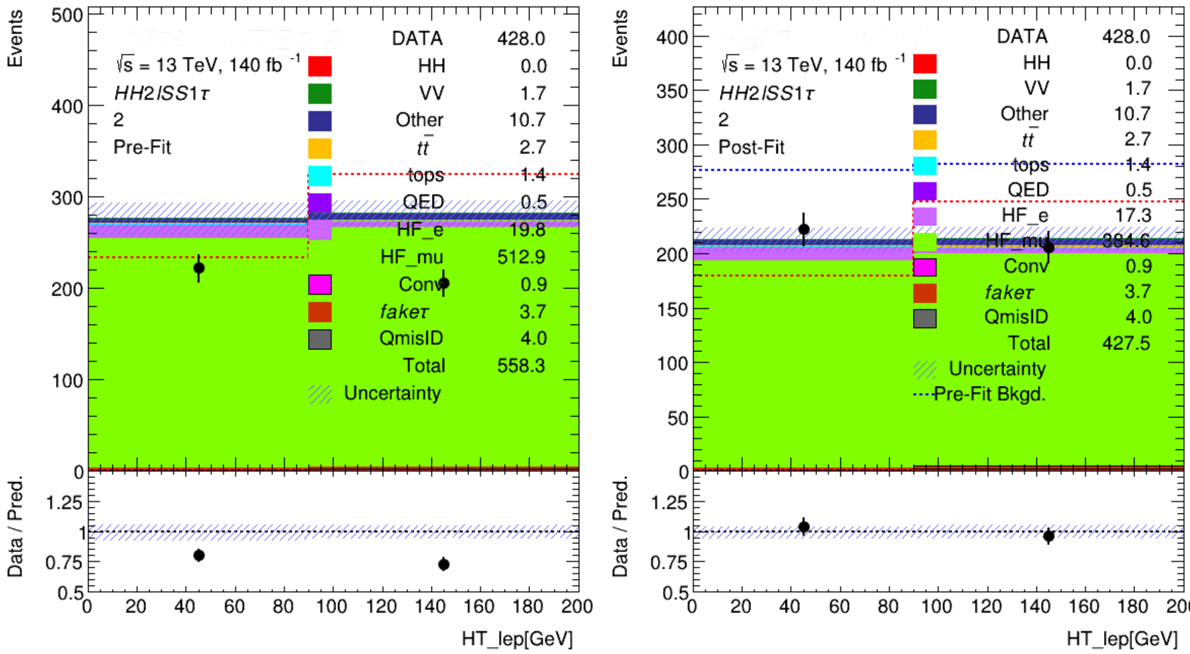


Figure 1: NormFactors for HF_e and HF_mu including systematics



(a) Pre-fit (b) Post-fit

Figure 2: DRll01 distribution restricted to μe and ee.



(a) Pre-fit (b) Post-fit

Figure 3: Sum of transverse momentum of leptons restricted to $e\mu$ and $\mu\mu$.

3 Systematics

Several systematic uncertainties were investigated and ranking plots were created which order the importance of the systematic uncertainties for the electron and muon related uncertainties separately. The variations of light lepton definitions on the heavy flavour normalisation factors are found to be most important. The systematic uncertainties were included in the fits for

- DRll01, and
- HT_lep.

Details of the systematic uncertainty study are given the internal ATLAS note [3].

4 Conclusion

In this summer student project, a contribution to a challenging analysis of the di-Higgs boson search was made. The specific analysis channel contains two same-sign electrically charged light leptons and one hadronically decaying tau (2ℓSS1τ). Two dedicated selection regions were used for the heavy flavour normalisation factors determination with the requirement of at least one b-tagged jet. Moreover, this extends a study carried by a previous summer student. In this study, full systematic uncertainties were applied to determine muon and electron fake normalisation factors. The uncertainties on the obtained normalisation factors are 0.06 of muons and 0.10 for electrons including statistical and systematic uncertainties. A dedicated study showed that the statistical uncertainties are 0.04 for muons and 0.08 for electrons.

Acknowledgments

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

References

1. Robl, J. B. 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. CERN-STUDENTS-Note-2021-233. `https://cds.cern.ch/record/2789878` (2021).
2. Fleischmann, T. M. Light Non-prompt (Fake) Lepton and Di-boson (VV) Background Estimation with a Template Fit Method for HH Production with a Signature of Two Same-sign Electrically Charged Light Leptons Associated with a Tau Lepton. CERN-STUDENTS-Note-2022-226. `https://cds.cern.ch/record/2845134` (2022).
3. Hamidaddeen, A. K. A. & Sopczak, A. *Fake Lepton Study in the Di-Higgs Multilepton Final State with a Signature of Two Same Electric Charge Light Leptons and a Hadronically Decaying Tau* tech. rep. (CERN, Geneva, 2023). `https://cds.cern.ch/record/2880006`.