

# Contents

<b>1</b>	<b>Searches for New Physics in <math>\tau^+\tau^-</math> Final States</b>	<b>2</b>
1.1	Signal Modelling . . . . .	3
1.1.1	Additional Higgs Bosons . . . . .	3
1.1.2	Vector Leptoquarks . . . . .	6
1.2	Event Selection . . . . .	7
1.2.1	Trigger Requirements . . . . .	9
1.2.2	Offline Requirements . . . . .	9
1.3	Signal Extraction . . . . .	11
1.4	Background Modelling Overview . . . . .	12
1.5	QCD Estimation in the $e\mu$ Channel . . . . .	15
1.6	Embedding Method . . . . .	15
1.7	Fake Factor Method . . . . .	16
1.7.1	Determination Regions . . . . .	16
1.7.2	Parametrisation . . . . .	19
1.7.3	Corrections . . . . .	19
1.7.4	Application Region Fractions . . . . .	19
1.7.5	Applying Fake Factors . . . . .	19
1.8	Uncertainty Model . . . . .	19
1.9	Postfit Plots . . . . .	25
1.10	MC Corrections . . . . .	25
1.11	Model Independent Results . . . . .	25
1.11.1	Limit Setting . . . . .	25
1.11.2	Significance and Compatibility . . . . .	25
1.11.3	2D Likelihood Scans . . . . .	25
1.12	Model Dependent Limits . . . . .	25
1.12.1	Limit Setting . . . . .	25

# Chapter 1

## Searches for New Physics in $\tau^+\tau^-$ Final States

The  $\tau^+\tau^-$  final states are a powerful tool to search for new physics at collider experiments. As the heaviest lepton, they are sensitive to resonant production of new neutral particles where the couplings have mass hierarchy. They are also sensitive to non-resonant effects from new physics mediators. This chapter will detail the searches for two such areas of new physics: additional Higgs bosons and vector leptoquarks. These searches are split up into three sections:

- i) A model independent search for single narrow spin-0 resonance,  $\phi$ , produced via gluon fusion ( $gg\phi$ ) or in association with a bottom quark ( $bb\phi$ ). The SM Higgs boson is treated as a background. The Yukawa couplings that contribute to the gluon fusion loop are set to SM values.
- ii) A search for the MSSM Higgs sector, in a number of benchmark scenarios. The benchmark scenarios are defined in Section 1.1.1. The production of SM Higgs boson is also used to constrain the available phase space.
- iii) A search for the t-channel exchange of a  $U_1$  vector leptoquark. Two scenarios are taken, based of the best fit to the b anomalies. These scenarios are detailed in Section 1.1.2.

These searches are performed with the full run-2 dataset ( $138\text{ fb}^{-1}$ ) collected by the CMS experiment. The search for additional Higgs bosons had previously been performed with data collected in 2016 ( $39\text{ fb}^{-1}$ ) and results were consistent with the SM background prediction.

## 1.1 Signal Modelling

### 1.1.1 Additional Higgs Bosons

Extended Higgs sectors, such as that of the MSSM, can be probed by direct searches for the additional bosons and further precise measurements of the Standard Model Higgs boson. This search for an extended Higgs sector is motivated by Type II 2HDMs, such as the MSSM. In these models  $\tan\beta$  enhances couplings of additional Higgs bosons to bottom-like quarks and leptons, whilst top-like couplings are suppressed. This narrows down the most important production modes of the Higgs boson into two categories: Gluon fusion and production in association with a bottom quark. Examples of these are shown in Figure 1.1.

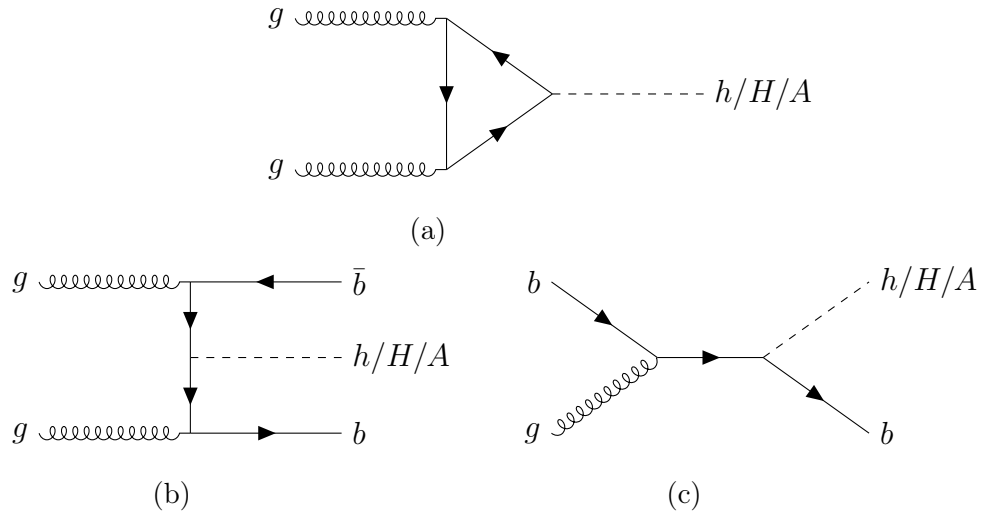


Figure 1.1: Diagram (a) shows the production of neutral Higgs bosons from gluon fusion. The dominant loop contributions to this diagrams are from top-only, bottom-only and top-bottom interference. Diagrams (b) and (c) show production in association with b quarks.

With the  $\tan\beta$  enhancement, the decays of additional Higgs bosons to tau leptons and bottom quarks are most likely. Tau leptons are identified with a higher purity than bottom quarks at the CMS detector. It is also easier to separate  $\tau^+\tau^-$  from the large QCD multijet background produced from the high energy proton-proton collisions. This hypothesis was tested with the 2016 dataset and although no deviations were observed, the strongest limits on the MSSM phase space was placed by the  $\tau^+\tau^-$  final states.

For this analysis, the production of additional Higgs bosons over a mass range of 60 GeV to 3.5 TeV are generated. Gluon fusion is simulated at NLO precision

using the 2HDM implementation of POWHEG 2.0. The kinematic properties are highly dependent on the contributions to the loop, which vary dependent on the specific signal model. To account for the different loop contributions at the NLO plus parton shower prediction, weights based off the  $p_T$  spectra are calculated to split the contributions from the t quark only, b quark only, and tb-interference. Once individual templates have been determined for each contribution to the loop, the 2HDM samples can be scaled to the correct specific MSSM scenario prediction by the following formula.

$$\begin{aligned} \frac{d\sigma_{\text{MSSM}}}{dp_T} = & \left( \frac{Y_{t,\text{MSSM}}}{Y_{t,2\text{HDM}}} \right)^2 \frac{d\sigma_{2\text{HDM}}^t(Q_t)}{dp_T} + \left( \frac{Y_{b,\text{MSSM}}}{Y_{b,2\text{HDM}}} \right)^2 \frac{d\sigma_{2\text{HDM}}^b(Q_b)}{dp_T} + \\ & \left( \frac{Y_{t,\text{MSSM}}}{Y_{t,2\text{HDM}}} \frac{Y_{b,\text{MSSM}}}{Y_{b,2\text{HDM}}} \right) \left\{ \frac{d\sigma_{2\text{HDM}}^{t+b}(Q_{tb})}{dp_T} - \frac{d\sigma_{2\text{HDM}}^t(Q_{tb})}{dp_T} - \frac{d\sigma_{2\text{HDM}}^b(Q_{tb})}{dp_T} \right\}, \end{aligned}$$

where  $Q_i$  are resummation scales that depend on the mass of the additional Higgs boson. Further contributions from any Supersymmetric partners have been checked and account for less than a few percent and so are neglected. This is also done separately for the scalar and pseudoscalar additional Higgs bosons, as the  $p_T$  distributions can differ. The MSSM benchmark scenarios considered are detailed in Ref. [1]. The scenarios provide the relative Yukawa couplings (to calculate the cross sections) and branching fractions of the MSSM Higgs bosons. An example of the changes to gluon fusion production, in the MSSM  $M_h^{125}$  scenario with  $m_A = 1600$  GeV and  $\tan\beta$  varying is shown in Figure 1.2. The distributions peak at a higher  $p_T$  for the top quark loop, therefore at smaller  $\tan\beta$ , where the top quark contribution is dominant, an additional Higgs boson would be more boosted.

Production in association with bottom quarks is simulated at NLO precision using the corresponding POWHEG 2.0 implementation in the four-flavour scheme. All additional Higgs boson signal generation is performed using the parton distribution function (PDF) NNPDF3.1. Tau lepton decay, parton showering and hadronisation are all modelled with the PYTHIA event generator where the PU profile is matched to data. All events generated are passed through a GEANT4-based simulation of the CMS detector and reconstructed in the same way as data.

The model dependent search for the MSSM also looks to find differences from the observed SM Higgs boson and the predicted MSSM SM-like Higgs boson. In each MSSM benchmark scenario, an uncertainty of  $\pm 3$  GeV is given on the prediction for the SM Higgs boson mass. This uncertainty is to reflect the contribution from any

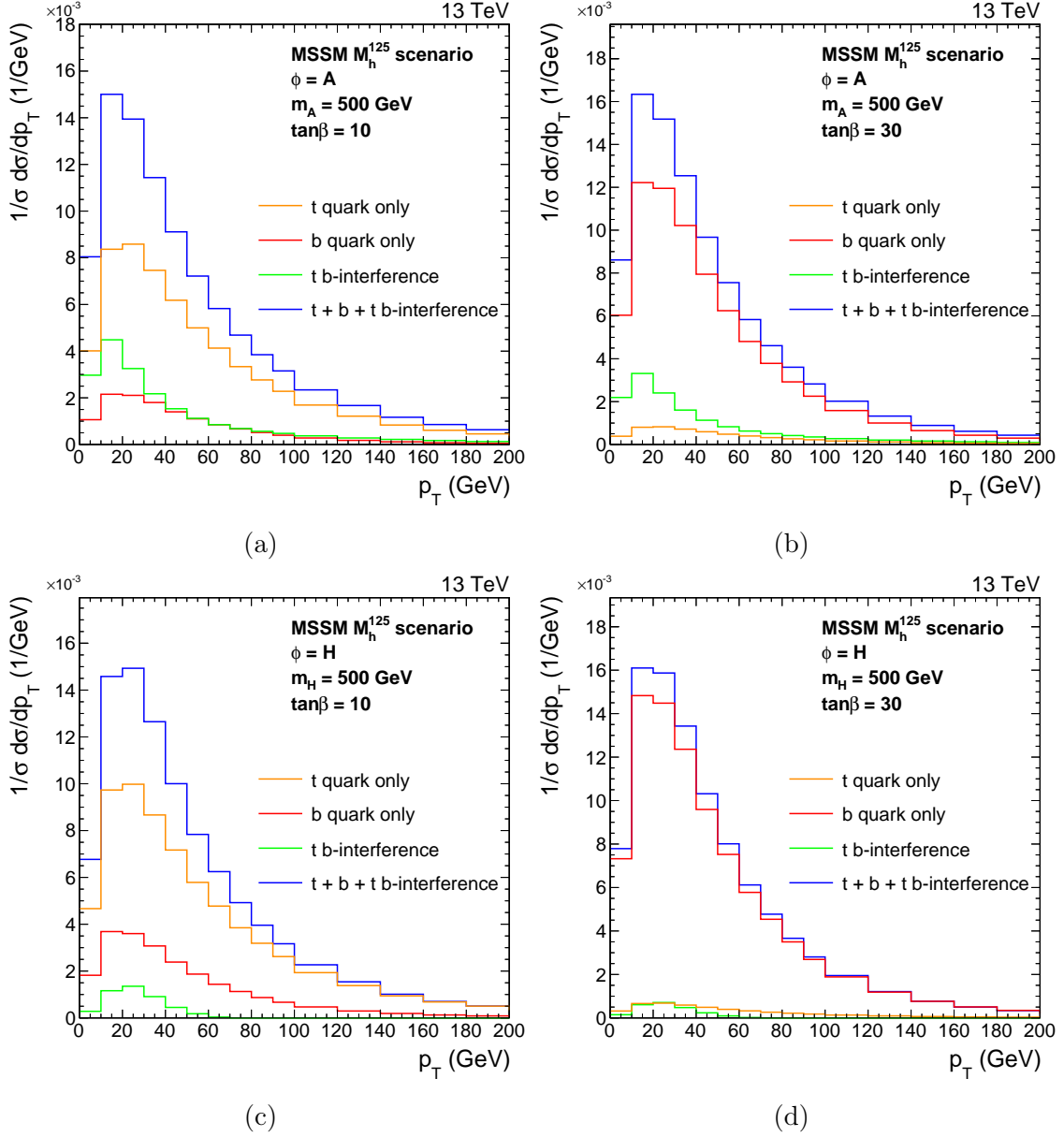


Figure 1.2:  $p_T$  density distributions of the A (top) and H (bottom) boson, with contributions to the gluon fusion loop displayed individually and summed. These are shown for  $\tan\beta$  values of 10 (left) and 30 (right) where  $m_A = 500$  GeV in the MSSM  $M_h^{125}$  scenario.

unknown higher-order corrections. The value of the mass is allowed to vary within this window, however the Yukawa couplings are rescaled the observed mass.

### 1.1.2 Vector Leptoquarks

The best fit in the vector leptoquark phase space to the B anomalies yielded large bottom quark and tau lepton couplings to the  $U_1$  particle. The possible production modes of a  $\tau^+\tau^-$  final state are shown in Figure 1.3.

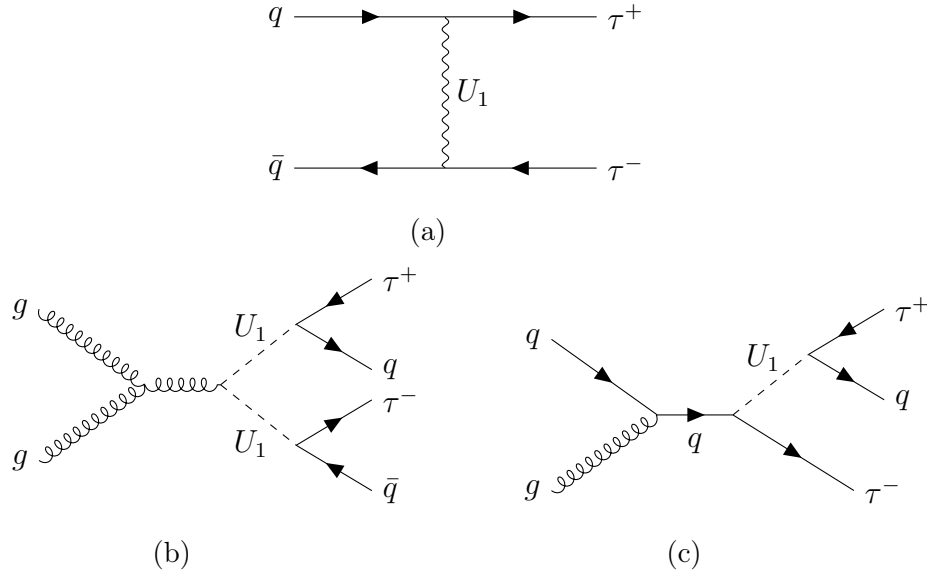


Figure 1.3: Feynman diagrams showing the contribution from  $U_1$  vector leptoquarks to the final state with a pair of oppositely charged tau leptons. Diagram (a) shows t-channel, (b) pair and (c) single production of a vector leptoquark.

Pair and single production of a vector leptoquark is dependent on its strong coupling, which is highly model dependent. For large mass,  $m_U$ , the probability of producing an on-shell  $U_1$  singlet or pair is heavily suppressed due to the momentum of the initial partons. These production processes are not discussed further in this search. Further studies have been used to search for single and pair production at the CMS experiment and no statistically significance derivation was observed.

The t-channel process contain two vertices with a  $U_1$  vector leptoquark, a quark and a tau lepton, and hence the cross section will scale with  $g_U^4$ . From the best fit to B anomalies this vertex will be dominated by the b quark and hence the initial state will be mostly from  $b\bar{b}$ , with sub-dominant contributions from  $b\bar{s}$ ,  $s\bar{b}$  and  $s\bar{s}$ . Although there are no additional b quarks in the final state in the LO process, initial state radiation can lead to additional b quarks in the final state. In this search

the two scenarios discussed in Section ?? are considered. The only non negligible parameter for  $\tau^+\tau^-$  final states from the fit in the  $m_U$ - $g_U$  phase space is the  $\beta_L^{s\tau}$  parameter. This is set to the best fit value.

The signal process of the  $U_1$  t-channel exchange is simulated in the five-flavour scheme (5FS) at LO precision using the MADGRAPH5\_aMC@NLO event generator, v2.6.5. Events are generated with one or fewer outgoing partons from the matrix element and the MLM prescription is then used for matching, with a scale set to 40 GeV. Negligible dependence of the  $U_1$  decay width ( $\Lambda$ ) is observed, for simulation this is chosen to approximately match the value predicted by the B anomaly fit. Samples with a mass between 1 and 5 TeV at  $g_U = 1$  are generated.

The interference between the  $U_1$  signal and  $Z/\gamma^* \rightarrow \tau\tau$  production was checked. A large destructive affect is observed, with the magnitude dependent on  $g_U$ . To account for this, separate samples are produced for this interference, generated in the same way as the t-channel exchange. The interference samples are then split into two with a di-tau mass split in order to have a sufficient number of events in the high di-tau mass regions. The cross section of these interference samples scale with  $g_U^2$ . Examples of the generator level di-tau mass distributions are shown in Figure 1.4.

The t-channel signal produces a broad distribution in  $m_{\tau\tau}$  due to its non-resonant nature. The interference is mostly a destructive effect (except for at small  $m_{\tau\tau}$ ), with the yield becoming less negative at higher  $m_{\tau\tau}$ . The interference peaks negatively between 100 and 200 GeV and in this region the combined yield can be negative. Due to the difference in scaling of the two effects, at small  $g_U$  the interference is more dominant than the signal and hence the yield of the combined result is reduced.

## 1.2 Event Selection

The possible decays of two tau leptons and their branching fractions, where the tau decay is grouped into three categories e,  $\mu$  and  $\tau_h$  as defined in Section ??, are shown in Table 1.1. For this search the four largest branching fraction channels used:  $\tau_h\tau_h$ ,  $e\tau_h$ ,  $\mu\tau_h$  and  $e\mu$ . This accounts for approximately 94% of di-tau events. The two same lepton channels are neglected due to small branching ratio and the dominating  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$  backgrounds.

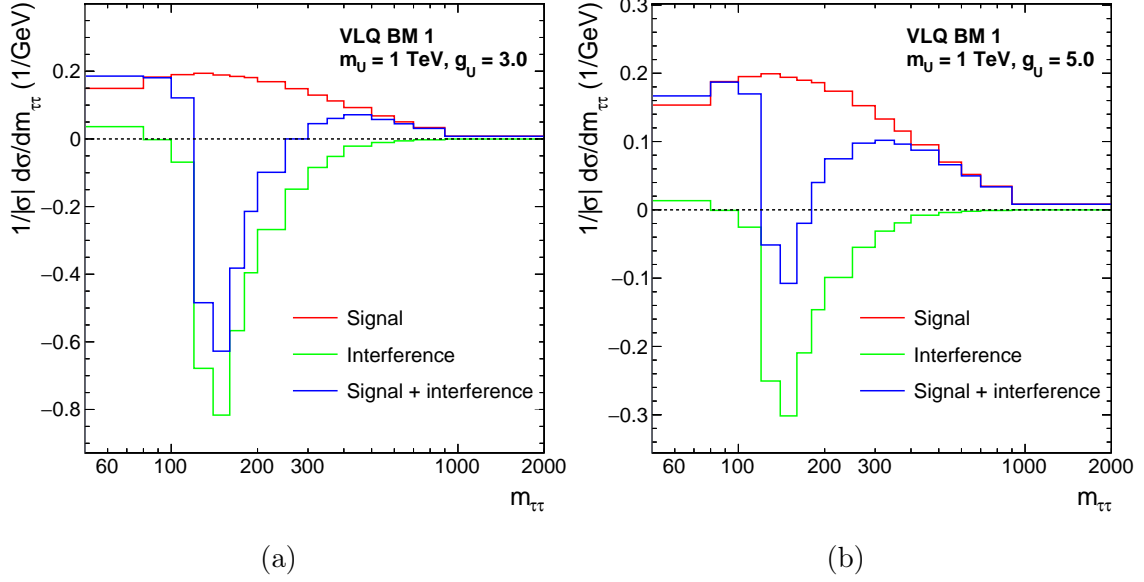


Figure 1.4: The generator level  $m_{\tau\tau}$  density distributions of the t-channel vector leptoquark signal and the interference with Drell-Yan. This is shown in the VLQ BM 1 scenario for a leptoquark of mass 1 TeV for coupling strengths of  $g_U = 3$  (a) and  $g_U = 5$  (b).

Channel	Branching Fraction
$\tau_h\tau_h$	42.0%
$e\tau_h$	23.1%
$\mu\tau_h$	22.6%
$e\mu$	6.2%
$ee$	3.2%
$\mu\mu$	3.0%

Table 1.1: Branching fractions of the decays of two tau leptons.



### 1.2.1 Trigger Requirements

In the four final state pairs a number of different online trigger requirements are needed. In the  $\tau_h\tau_h$  channel, two possible triggers are available: the double- $\tau_h$  and single- $\tau_h$  triggers. The single- $\tau_h$  trigger has a high  $p_T$  threshold at 120 (180) GeV for events recorded in 2016 (2017-2018), whilst the double- $\tau_h$  has a  $p_T$  threshold at 40 GeV. Therefore, the double- $\tau_h$  trigger is used individually where the  $\tau_h$  has  $p_T$  is below the single- $\tau_h$  threshold and the union of single- $\tau_h$  and double- $\tau_h$  triggered events are taken above the threshold.

In the  $e\tau_h$  and  $\mu\tau_h$  channels, there are three possible triggers available: the single- $e/\mu$ , single- $\tau_h$  and the  $e/\mu$ - $\tau_h$  cross-trigger. The cross-trigger is used for events where the light lepton has  $p_T$  between the thresholds for the cross-trigger and single- $e/\mu$  shown in Table 1.2. The light lepton used in the cross-trigger is required to be in the central barrel of the detector within  $|\eta| < 2.1$ . Above these light lepton  $p_T$  thresholds the single- $e/\mu$  trigger is used, where here it is required that the  $\tau_h$  has  $p_T > 30$  GeV. At  $\tau_h$   $p_T$  above the single- $\tau_h$  thresholds, the single- $\tau_h$  trigger is used in combination with the single- $e/\mu$  trigger.

Year/ Trigger	$e\tau_h$ cross-trigger	single- $e$	$\mu\tau_h$ cross-trigger	single- $\mu$
2016	23	26	20	23
2017	25	28	20	25
2018	25	33	21	25

Table 1.2: Lower trigger light lepton thresholds  $p_T$  in GeV for the  $e\tau_h$  and  $\mu\tau_h$  channels.

In the  $e\mu$  channel, there are three possible triggers available: the single- $e$ , single- $\mu$  and the  $e$ - $\mu$  cross-trigger. However, only the cross-trigger is used in this analysis, due to the larger efficiencies of correctly selecting light leptons. The  $e$  and  $\mu$  are required to have  $p_T > 15$  GeV and  $|\eta| < 2.4$ .

### 1.2.2 Offline Requirements

All offline selections stated are in addition the object selection discussed in Section ???. In this analysis, hadronic tau candidates are required to pass the **Medium**  $D_{\text{jet}}^{\text{WP}}$ .  $D_e^{\text{WP}}$  and  $D_\mu^{\text{WP}}$  are dependent on the channel. The **V Loose**, **Tight**, **V Loose**  $D_e^{\text{WP}}$  and the **V Loose**, **V Loose** and **Tight**  $D_\mu^{\text{WP}}$  are used in the  $\tau_h\tau_h$ ,  $e\tau_h$  and  $\mu\tau_h$

channels respectively. The tighter working point for the same light lepton discrimination as tagged in the event is used to remove light leptons faking hadronic taus from the  $Z \rightarrow ll$  process. The light lepton isolation requirement is  $I_{\text{rel}}^{e/\mu} < 0.15$  except for in the  $e\mu$  channel where the muon is required to have  $I_{\text{rel}}^\mu < 0.2$ .

The selected  $\tau$  lepton decay candidates are required to have opposite charge and to be separated by more than  $\Delta R > 0.5$  in all channels except  $e\mu$  where  $\Delta R > 0.3$ . In events where the numbers of an object in the event is greater than the required number of objects in the  $\tau\tau$  decay channel, the objects are sorted by the maximum  $D_{\text{jet}}^{\text{score}}$  if  $\tau_h$  or minimum  $I_{\text{rel}}$  if a light lepton and the leading objects are chosen. In order to maintain orthogonality between channels, events with additional light leptons passing looser selections than the nominal requirements, are rejected from the selection. The looser selections help to suppress the  $Z \rightarrow ll$  background process further.

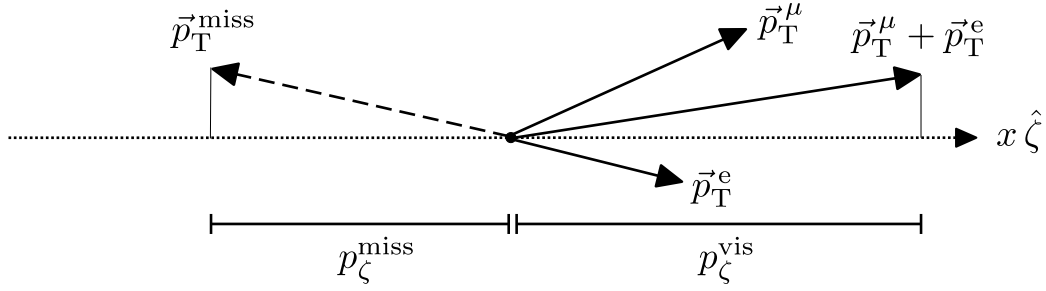
In the  $e\tau_h$  and  $\mu\tau_h$  channels, a cut is placed at 70 GeV on the transverse mass between the light lepton  $\vec{p}_T$  and the missing  $\vec{p}_T$ , where the transverse momentum is defined as,

$$m_T(\vec{p}_T^i, \vec{p}_T^j) = \sqrt{2p_T^i p_T^j (1 - \cos \Delta\phi)}, \quad (1.1)$$

where  $\Delta\phi$  to the azimuthal angle between  $\vec{p}_T^i$  and  $\vec{p}_T^j$ . The variable is used to remove  $W + \text{jets}$  background events, where a jet fakes a hadronic tau and the MET and light lepton from the  $W$  decay are aligned and hence the event has a large  $m_T(\vec{p}_T^{e/\mu}, \vec{p}_T^{\text{MET}})$ . In the  $e\mu$  channel an additional cut is placed on a variable named  $D_\zeta$ , which is defined as,

$$D_\zeta = p_\zeta^{\text{miss}} - 0.85p_\zeta^{\text{vis}}; \quad p_\zeta^{\text{miss}} = \vec{p}_T^{\text{miss}} \cdot \hat{\zeta}; \quad p_\zeta^{\text{vis}} = (\vec{p}_T^e + \vec{p}_T^\mu) \cdot \hat{\zeta} \quad (1.2)$$

where  $\vec{p}_T^{e/\mu}$  corresponds to the transverse momentum vector of the electron or muon and  $\hat{\zeta}$  to the bisectonal direction between the electron and the muon in the transverse plane [?]. A diagram of the inputs is shown Figure 1.5. The linear combination is optimised for genuine di-tau events to peak around  $D_\zeta = 0$  GeV. It is motivated by the expectation that in di-tau decays from a resonance, the visible and missing (from tau neutrinos) momentums are roughly aligned and of similar magnitudes. In  $W + \text{jets}$  and  $t\bar{t}$  events the directions of the visible and missing products are expected to be more randomly distributed and lead to a non-peaking  $D_\zeta$ . Therefore only events with  $D_\zeta > -35$  GeV are considered for signal events. No b tagged events with this cut are vetoed and b tagged events with this cut are used for a  $t\bar{t}$  control region and discussed further in Section 1.4.

Figure 1.5: Diagram of inputs to the  $D_\zeta$  variable.

### 1.3 Signal Extraction

The optimisation of the signal extraction depends on which of the three scenarios, set out at beginning of this section, is searched for. The components of the optimisation are named the high mass, low mass and SM Higgs optimisation procedures. For the model independent search (i) the high or low mass optimisation procedures are used depending on whether the mass of the resonance is greater or less than 250 GeV. The search for the MSSM Higgs sector (ii) uses the high mass or the SM Higgs optimisation procedures depending on whether the reconstructed di-tau mass  $m_{\tau\tau}$  is greater or less than 250 GeV. Finally, the search for vector leptoquarks (iii) uses only the high mass optimisation procedure. The procedures are discussed in detail below.

The high mass optimisation procedure follows what was done in Ref. [1]. Firstly each category is split into categories with no b tagged and with one or more b tagged events. This firstly helps target the additional Higgs boson production modes gluon fusion and b associated production respectively. Secondly, the initial state radiation of a t-channel vector leptoquark signal dominated by initial states of b quarks, can lead to additional b jets in the final state. The reduced backgrounds in b tagged events allows for a more sensitive vector leptoquark search in this category.

The  $e\tau_h$  and  $\mu\tau_h$  channels are further subdivided into categories

The corresponding categories in  $e\tau_h$  and  $\mu\tau_h$  channels are defined by

- **Tight- $m_T$** :  $m_T^{e/\mu}[\text{GeV}] < 40$ ;
- **Loose- $m_T$** :  $40 \leq m_T^{e/\mu}[\text{GeV}] < 70$ .

The bulk of the signal events, particularly for low mass hypotheses, lies in the **Tight- $m_T$**  sub-category. The **Loose- $m_T$**  category has been added to increase the signal acceptance for mass hypotheses of  $m_{A,H} > 700$  GeV.

In the  $e\mu$  final state, three sub-categories are defined by

- **Low- $D_\zeta$** :  $-35 \leq D_\zeta[\text{GeV}] < -10$ ;
- **Medium- $D_\zeta$** :  $-10 \leq D_\zeta[\text{GeV}] < 30$ ;
- **High- $D_\zeta$** :  $D_\zeta[\text{GeV}] \geq 30$ .

In this way sub-categories with different signal purities and  $t\bar{t}$  fractions can be exploited during the statistical inference for the signal. The expected signal, for all masses tested, is mostly located in the **Medium- $D_\zeta$**  sub-category.

	No b tag			b tag		
$e\mu$	Low- $D_\zeta$	Medium- $D_\zeta$	High- $D_\zeta$	Low- $D_\zeta$	Medium- $D_\zeta$	High- $D_\zeta$
$e\tau_h$	Loose- $m_T$		Tight- $m_T$	Loose- $m_T$		Tight- $m_T$
$\mu\tau_h$	Loose- $m_T$		Tight- $m_T$	Loose- $m_T$		Tight- $m_T$
$\tau_h\tau_h$						
$t\bar{t}(e\mu)$				$D_\zeta < -35 \text{ GeV}$		
	Signal region (SR)					
	Control region					

Figure 1.6: Overview of the categories used for the extraction of the signal in the high mass optimisation procedure.

$$m_T^{\text{tot}} = \sqrt{m_T(\vec{p}_T^{\tau_1}, \vec{p}_T^{\text{miss}})^2 + m_T(\vec{p}_T^{\tau_2}, \vec{p}_T^{\text{miss}})^2 + m_T(\vec{p}_T^{\tau_1}, \vec{p}_T^{\tau_2})^2} \quad (1.3)$$

## 1.4 Background Modelling Overview

The analysis considers several backgrounds including Drell-Yan,  $t\bar{t}$ ,  $W$ +jets, QCD, di-boson, single-top, and electroweak  $W$  and  $Z$  bosons production. These are split into a five categories:

- Events containing only genuine tau leptons.
- Events with a jet misidentified as a hadronic tau ( $\text{jet} \rightarrow \tau_h$ ) in the  $e\tau_h$ ,  $\mu\tau_h$  or  $\tau_h\tau_h$  channels.

No b tag		b tag		
$e\mu$	Medium- $D_\zeta$ $p_{\text{T}}^{\tau\tau} < 50 \text{ GeV}$ $50 < p_{\text{T}}^{\tau\tau} < 100 \text{ GeV}$ $100 < p_{\text{T}}^{\tau\tau} < 200 \text{ GeV}$ $p_{\text{T}}^{\tau\tau} > 200 \text{ GeV}$	High- $D_\zeta$ $p_{\text{T}}^{\tau\tau} < 50 \text{ GeV}$ $50 < p_{\text{T}}^{\tau\tau} < 100 \text{ GeV}$ $100 < p_{\text{T}}^{\tau\tau} < 200 \text{ GeV}$ $p_{\text{T}}^{\tau\tau} > 200 \text{ GeV}$	Medium- $D_\zeta$	High- $D_\zeta$
	Tight- $m_{\text{T}}$ $p_{\text{T}}^{\tau\tau} < 50 \text{ GeV}$ $50 < p_{\text{T}}^{\tau\tau} < 100 \text{ GeV}$ $100 < p_{\text{T}}^{\tau\tau} < 200 \text{ GeV}$ $p_{\text{T}}^{\tau\tau} > 200 \text{ GeV}$		Tight- $m_{\text{T}}$	
$e\tau_{\text{h}}$	Tight- $m_{\text{T}}$ $p_{\text{T}}^{\tau\tau} < 50 \text{ GeV}$ $50 < p_{\text{T}}^{\tau\tau} < 100 \text{ GeV}$ $100 < p_{\text{T}}^{\tau\tau} < 200 \text{ GeV}$ $p_{\text{T}}^{\tau\tau} > 200 \text{ GeV}$		Tight- $m_{\text{T}}$	
$\mu\tau_{\text{h}}$	Tight- $m_{\text{T}}$ $p_{\text{T}}^{\tau\tau} < 50 \text{ GeV}$ $50 < p_{\text{T}}^{\tau\tau} < 100 \text{ GeV}$ $100 < p_{\text{T}}^{\tau\tau} < 200 \text{ GeV}$ $p_{\text{T}}^{\tau\tau} > 200 \text{ GeV}$		Tight- $m_{\text{T}}$	
$\tau_{\text{h}}\tau_{\text{h}}$	$p_{\text{T}}^{\tau\tau} < 50 \text{ GeV}$ $50 < p_{\text{T}}^{\tau\tau} < 100 \text{ GeV}$ $100 < p_{\text{T}}^{\tau\tau} < 200 \text{ GeV}$ $p_{\text{T}}^{\tau\tau} > 200 \text{ GeV}$			
$t\bar{t}(e\mu)$			$D_\zeta < -35 \text{ GeV}$	
<div></div> Signal region (SR)		<div></div> Control region		

Figure 1.7: Overview of the categories used for the extraction of the signal in the low mass optimisation procedure.

- iii) Events with jets faking both light leptons ( $\text{jet} \rightarrow l$ ) in the  $e\mu$  channel.
- iv) Events from  $t\bar{t}$  with a prompt light lepton ( $e$  or  $\mu$  not from a  $\tau$  decay) and the other object (if there are not two prompt light lepton) is from a genuine tau leptons.
- v) Other events. This is a small contribution and hence why it is grouped.
  - Non  $t\bar{t}$  events with a prompt light lepton ( $e$  or  $\mu$  not from a  $\tau$  decay) and the other object (if there are not two prompt light lepton) is from a genuine tau leptons.
  - Events with a light lepton faking a hadronic tau and the other object (if there are not two light leptons faking a hadronic tau) are reconstructed as prompt light lepton or from genuine tau leptons.
  - Events with a jet faking a light lepton and the other object is from genuine tau leptons in the  $e\tau_h$ ,  $\mu\tau_h$  or  $\tau_h\tau_h$  channels.
  - Events with one jet faking a light lepton and the other object from a prompt light lepton in the  $e\mu$  channel.

Backgrounds from (i) consists of largely  $Z/\gamma^* \rightarrow \tau\tau$  events but there are also smaller contributions from other processes. This background is modelled by a data-simulation hybrid method called the embedding method and this is described in detail in Section 1.6. Group (ii) is dominated by QCD,  $W + \text{jets}$  and  $t\bar{t}$  events with a  $\text{jet} \rightarrow \tau_h$  misidentification. This is modelled from data by the fake factor method ( $F_F$ ) and is explained in Section 1.7. Group (iii) is modelled from data to describe QCD multijet contribution to the background in the  $e\mu$  channel. The method to obtain this background is described in Section 1.5. The data driven background estimations for (i), (ii) and (iii) contribute  $>98\%$  of all expected background events in the  $\tau_h\tau_h$  channel,  $>90\%$  in  $e\tau_h$  and  $\mu\tau_h$  channels and  $>50\%$  in the  $e\mu$  channel. The final groups, (iv) and (v), are modelled with MC. The  $t\bar{t}$  process is separated due to its large contribution to the phase space where a b jet is required.

The  $W + \text{jets}$  and  $Z \rightarrow ll$  processes are simulated at leading order (LO) using the MADGRAPH5\_aMC@NLO 2.2.2 (2.4.2) event generator [?, ?] for the simulation of the data taken in 2016 (2017–2018). To increase the number of simulated events in regions of high signal purity, supplementary samples are generated with up to four outgoing partons in the hard interaction. For diboson production, MADGRAPH5\_aMC@NLO is used at next-to-LO (NLO) precision. In each case,

the FxFx [?] (MLM [?]) prescription is used to match the NLO (LO) matrix element calculation with the parton shower model. For  $t\bar{t}$  [?] and (t-channel) single top quark production [?], samples are generated at NLO precision using POWHEG 2.0 [?, ?, ?, ?]. The POWHEG version 1.0 at NLO precision is used for single top quark production in association with a W boson (tw channel) [?].

When compared with data,  $W + \text{jets}$ ,  $Z \rightarrow l\bar{l}$ ,  $t\bar{t}$ , and single top quark events in the tW channel are normalised to their cross sections at next-to-NLO (NNLO) precision [?, ?, ?]. Single top quark (t-channel) and diboson events are normalized to their cross sections at NLO precision or higher [?, ?, ?].

## 1.5 QCD Estimation in the $e\mu$ Channel

## 1.6 Embedding Method

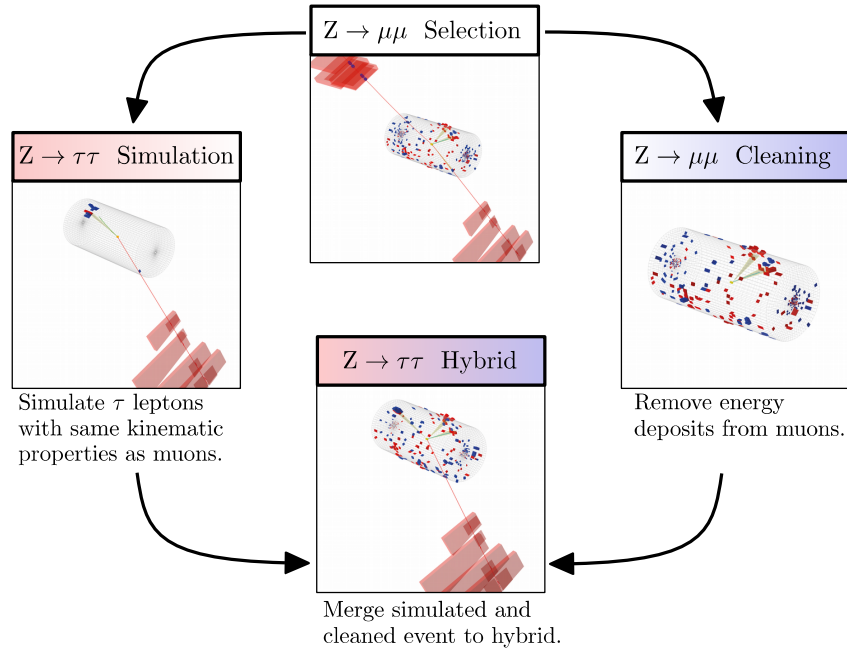


Figure 1.8: Schematic of the embedding method to model genuine di-tau backgrounds from di-muon events in data.

Validation plots of it working

## 1.7 Fake Factor Method

Backgrounds in which a jet fakes a  $\tau_h$  can be difficult to model using MC due to the poor description of the  $\text{jet} \rightarrow \tau_h$  fake rate in simulation. In addition, the small probability of a jet being misidentified as a  $\tau_h$  necessitates the production of high statistics MC samples at a significant computational expense. These shortcomings motivate the use of data-driven estimates for these processes. One such procedure is the fake factor ( $F_F$ ) method.

The  $F_F$  method utilises regions in the data to model the  $\text{jet} \rightarrow \tau_h$  background. Firstly, the determination regions (DR), which are  $\text{jet} \rightarrow \tau_h$  enriched control regions orthogonal to the signal region (SR). It is used to calculate  $F_F$  by taking the ratio of number of jet fake events that pass the nominal hadronic tau ID requirement ( $N(\text{Nominal})$ ), to the number of jet fake events that fail the nominal hadronic tau ID but pass a looser alternative hadronic tau ID requirement ( $N(\text{Alternative} \ \&\& \ !\text{Nominal})$ ), as shown in Equation 1.4.

$$F_F = \frac{N(\text{Nominal})}{N(\text{Alternative} \ \&\& \ !\text{Nominal})}. \quad (1.4)$$

In the remaining text this numerator and denominator are referred to as the pass and fail regions. The derivation of this ratio is done differentially with respect to key parameters that differ in the two regions. Once  $F_F$  have been derived it is common to calculate corrections in other sideband regions (a region orthogonal to the signal region) and combine  $F_F$  measured from different processes. Finally, the  $F_F$  are applied to the application region (AR). This is defined as the SR but with the criteria that the jet fakes fail the nominal hadronic tau ID but pass the looser alternative tau ID requirement. This now models the background from  $\text{jet} \rightarrow \tau_h$  events in SR.

The following Sections 1.7.1–1.7.5 detail the complexities of how this method is applied to this analysis. For these searches the nominal hadronic tau ID used is the Medium  $D_{\text{jet}^{\text{WP}}}$  and the alternative hadronic tau ID used is the VVLoose  $D_{\text{jet}^{\text{WP}}}$ .

### 1.7.1 Determination Regions

The fake factors are measured separately in each year of data taking period (2016, 2017, 2018), in each channel containing hadronic taus ( $e\tau_h$ ,  $\mu\tau_h$ ,  $\tau_h\tau_h$ ) and in enriched regions of dominant processes that contribute  $\text{jet} \rightarrow \tau_h$  events. In the  $e\tau_h$  and  $\mu\tau_h$  channels  $F_F$  are measured for three processes: QCD, W + Jets and  $t\bar{t}$ . In the  $\tau_h\tau_h$  channel  $F_F$  are measured only for the dominant QCD process. The QCD



process is assumed to produce two jet fakes and so the fake factors is chosen to be calculated from leading  $p_T$  hadronic tau candidate only. Section 1.7.5 discusses how single jet fake events in the  $\tau_h\tau_h$  channel are modelled.

Each separate measurement region is split into three sideband regions based off two cuts that surround the signal region. These regions are named the **Determination Region (C)**, **Alternative Determination Region (D)** and **Correction Region (B)** and are schematically shown in Figure 1.9.

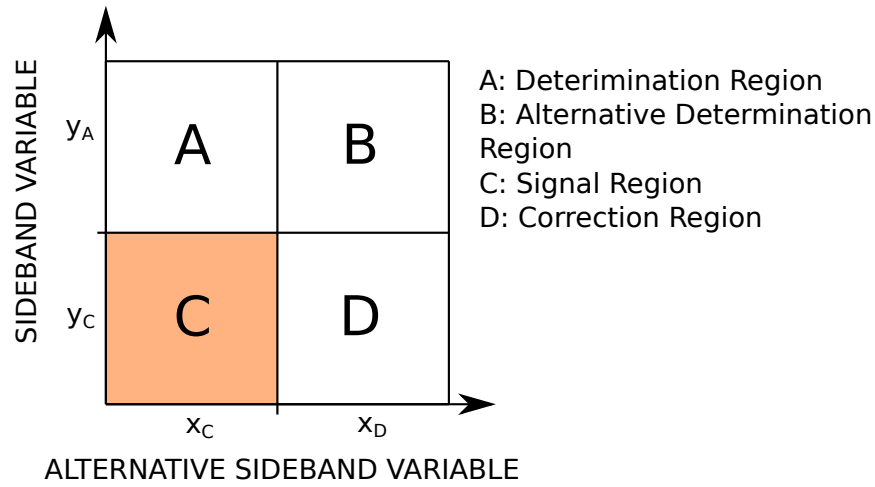


Figure 1.9: Schematic of the regions used for fake factor derivation.

Region A is used to measure and fit fake factors. Region B is an alternative region used to measure and fit fake factor to account for the difference in fake factors between A and C. These alternative fake factors are applied to the fail region in D and corrections are calculated comparing it to the pass region in D. The total fake factor per measurement region is calculated as the fake factors derived in region A multiplied by the correction calculated from region B to D.

The selection for  $x_C$ ,  $x_D$ ,  $y_C$  and  $y_A$ , as defined in Figure 1.9, in each separate measurement region are shown below. These are chosen to balance the number of events and the purity of each background in the region.

i)  $\tau_h\tau_h$  QCD

$y_C$ : The  $\tau_h$  candidates are required to have the opposite sign.

$y_A$ : The  $\tau_h$  candidates are required to have the same sign.

$x_C$ : The subleading tau passes the **Medium**  $D_{\text{jet}}^{\text{WP}}$ .

$x_D$ : The subleading tau fails the VVLoose  $D_{\text{jet}^{\text{WP}}}$  but passes the VVVLoose  $D_{\text{jet}^{\text{WP}}}$ .

ii)  $e\tau_h$  and  $\mu\tau_h$  QCD

$y_C$ : The  $e/\mu$  and  $\tau_h$  candidates are required to have the opposite sign.

$y_A$ : The  $e/\mu$  and  $\tau_h$  candidates are required to have the same sign and the  $e/\mu$  to have  $I_{\text{rel}} > 0.05$ .

$x_C$ : The  $e/\mu$  candidate is required to have  $I_{\text{rel}} < 0.15$ .

$x_D$ : The  $e/\mu$  candidate is required to have  $0.25 < I_{\text{rel}} < 0.5$ .

iii)  $e\tau_h$  and  $\mu\tau_h$  W + Jets

$y_C$ : The  $m_T$  between the  $e/\mu$  and the MET  $< 70$  GeV.

$y_A$ : The  $m_T$  between the  $e/\mu$  and the MET  $> 70$  GeV and no b jets in the event.

$x_C$ : Data.

$x_D$ : W + Jets MC.

iv)  $e\tau_h$  and  $\mu\tau_h$   $t\bar{t}$

$y_C$ : Data.

$y_A$ : MC ( $t\bar{t}$  in B and W + Jets D).

$x_C$ :  $m_T < 70$  GeV.

$x_D$ :  $m_T > 70$  GeV and no b jets.

In the  $\mu\tau_h$  and  $e\tau_h$  channels QCD and W + Jets jet fake events are in general the most significant and contribute with approximately equal weights.  $t\bar{t}$  inclusively is small but becomes more significant when searching for events with a b jet. The additional  $I_{\text{rel}} > 0.05$  requirement in these channels for QCD is to reduce processes producing genuine leptons and the  $N_{\text{b-jets}} = 0$  requirement for W + Jets is to reduce  $t\bar{t}$  contamination. It is not possible to define a DR that is sufficiently pure in  $t\bar{t}$  events to make a reasonable measurement of  $F_F^{t\bar{t}}$  from data. Therefore  $F_F^{t\bar{t}}$  are derived from MC. A comparison of the  $F_F^{W+\text{jets}}$  measured in data and MC shows only  $\sim 10$ – $20\%$  differences in the fake rates in data and MC. This observation coupled with the fact that the  $t\bar{t}$  contribution is small compared to the other processes means that any bias introduced by using  $F_F^{t\bar{t}}$  measured in MC is small compared to the uncertainties on the fake factors, discussed in Section ??.

### 1.7.2 Parametrisation

The  $F_F^i$  take into account dependencies on  $N_{\text{jets}}$  via the analysis tailed variable  $N_{\text{pre b-jets}}$ , the  $p_T$  of the  $\tau_h$  candidate ( $p_T^{\tau_h}$ ) and the  $p_T$  of the jet matched in  $\Delta R$  to the  $\tau_h$  ( $p_T^{\text{jet}}$ ).  $N_{\text{pre b-jets}}$  is defined to map the dependence of  $F_F^i$  on  $N_{\text{jets}}$  and describe the categorising variable  $N_{\text{b-jets}}$  well. It is the number of jets in the event with  $|\eta| < 2.4$  and  $p_T > 20$ . These are the same  $\eta$  and  $p_T$  thresholds required for a b-jet. They are also calculated separately for the three channels where a jet can fake a hadronic tau,  $e\tau_h$ ,  $\mu\tau_h$  and  $\tau_h\tau_h$ . The  $F_F^i$  are then measured separately for the dominant jet  $\rightarrow \tau_h$  processes  $i$ , which for the  $e\tau_h$  and  $\mu\tau_h$  channels includes QCD,  $W$ +jets and  $t\bar{t}r$ , and for the  $\tau_h\tau_h$  channel includes only QCD. The  $F_F^i$  are finally corrected to account for extrapolations from the DR to the SR, and for missing variable dependencies (via non-closure corrections). The  $F_F^i$  measured for the different processes are combined into an overall factor,  $F_F$ , using

$$F_F = \sum_i f_i \cdot F_F^i, \quad (1.5)$$

where the factor  $f_i$  is defined as

$$f_i = \frac{N_{\text{AR}}^i}{\sum_j N_{\text{AR}}^j}, \quad (1.6)$$

which is the fraction of events with a jet  $\rightarrow \tau_h$  originating from process  $i$  over the total number of jet  $\rightarrow \tau_h$  events for all processes in the AR.

### 1.7.3 Corrections

### 1.7.4 Application Region Fractions

### 1.7.5 Applying Fake Factors

using leading tau subtracting off rest w fakes in  $t\bar{t}$

## 1.8 Uncertainty Model

bulleted list of uncertainties used in the analysis.

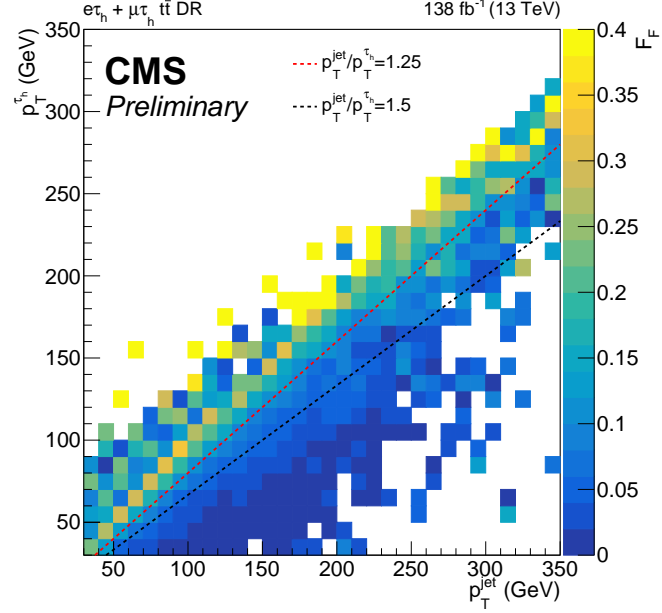


Figure 1.10: A 2D heat map of the fake factors determined from  $t\bar{t}$  MC for the full run-2 dataset in the combined  $e\tau_h$  and  $\mu\tau_h$  channels. This is shown with respect to the hadronic tau  $p_T$  and the  $p_T$  of the jet matched to the hadronic tau. The ratio of jet to hadronic tau  $p_T$  categorisation used is shown split by the dashed lines.

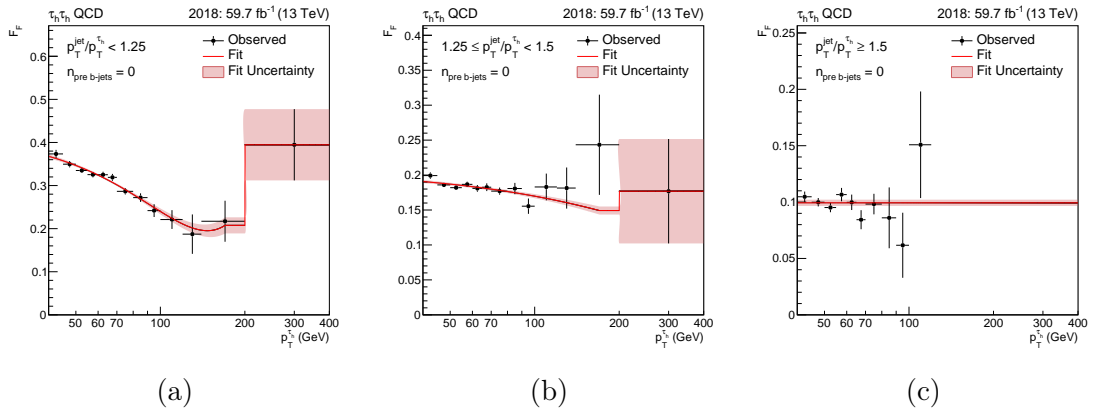


Figure 1.11: Fake factor fits in  $\tau_h\tau_h$  channel for the QCD  $n_{prebjets} = 0$  category with 2018 data. The three jet  $p_T$  to hadronic tau  $p_T$  categories are shown.

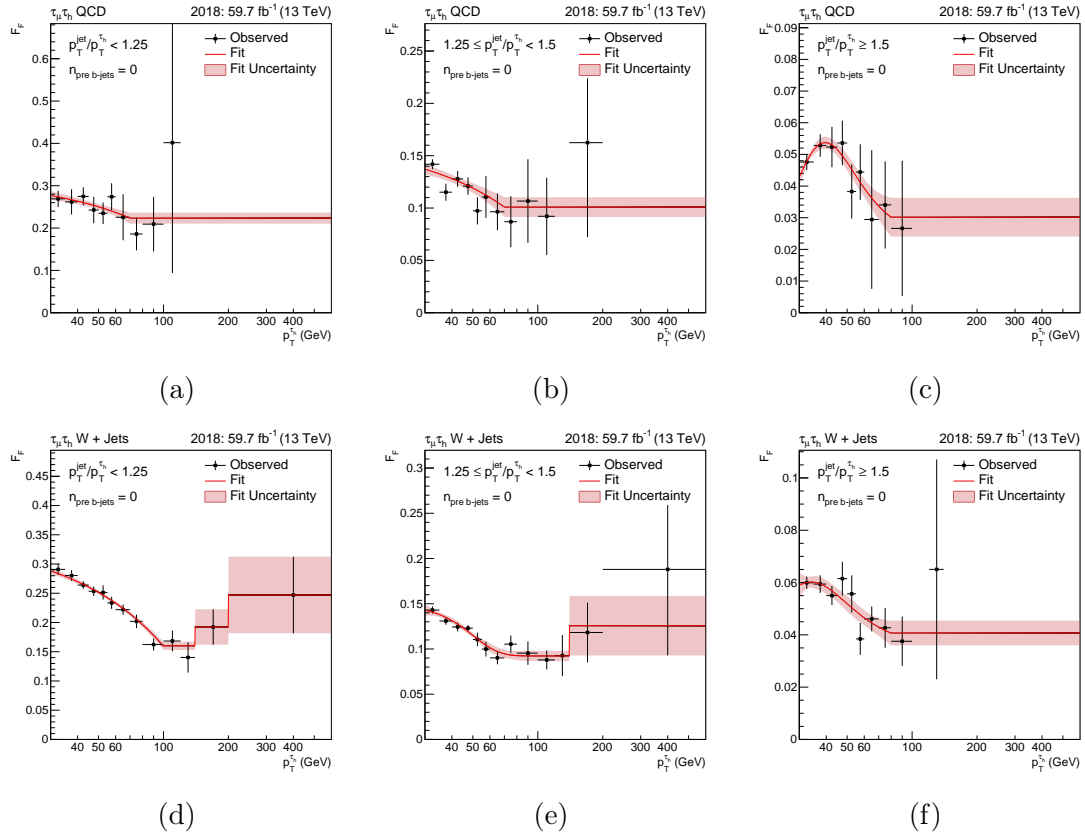


Figure 1.12: Fake factor fits in  $\mu\tau_h$  channel for the QCD and  $W + \text{Jets}$   $n_{prebjets} = 0$  category with 2018 data. The three jet  $p_T$  to hadronic tau  $p_T$  categories are shown for each process.

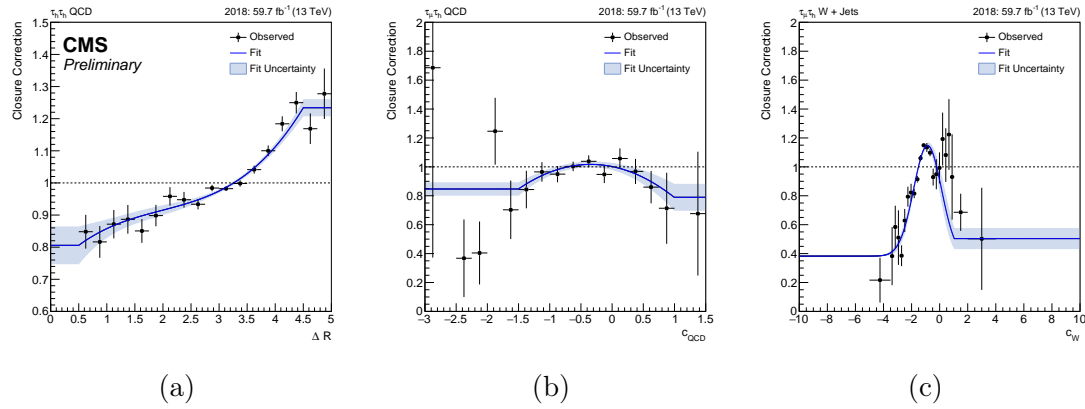


Figure 1.13: DR closures

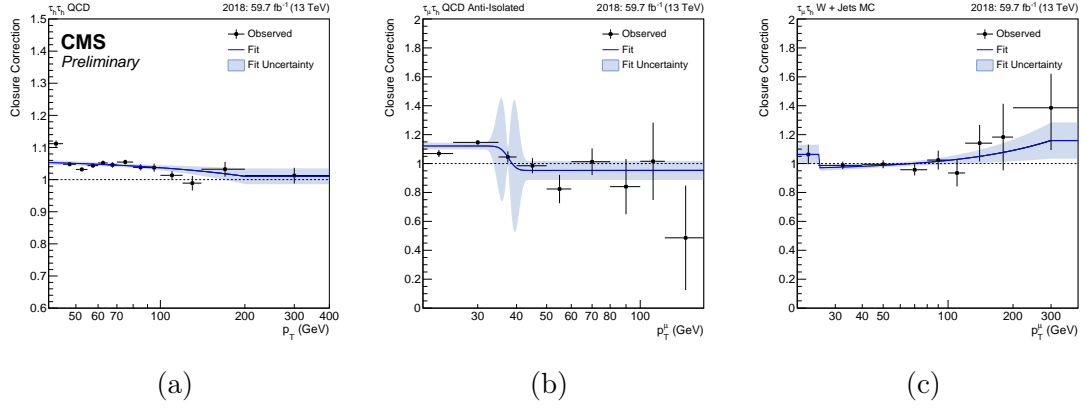


Figure 1.14: DR to AR closures.

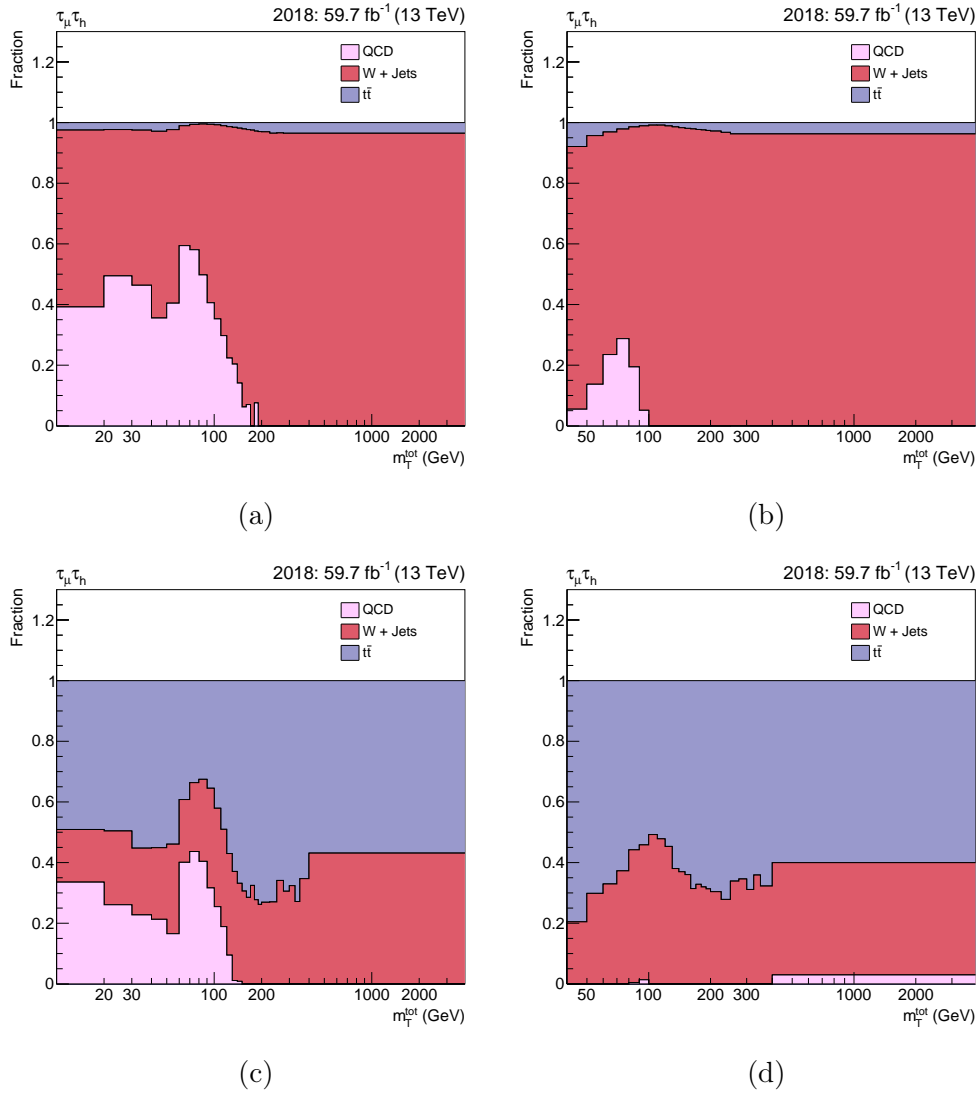


Figure 1.15: Fake factor fractions.

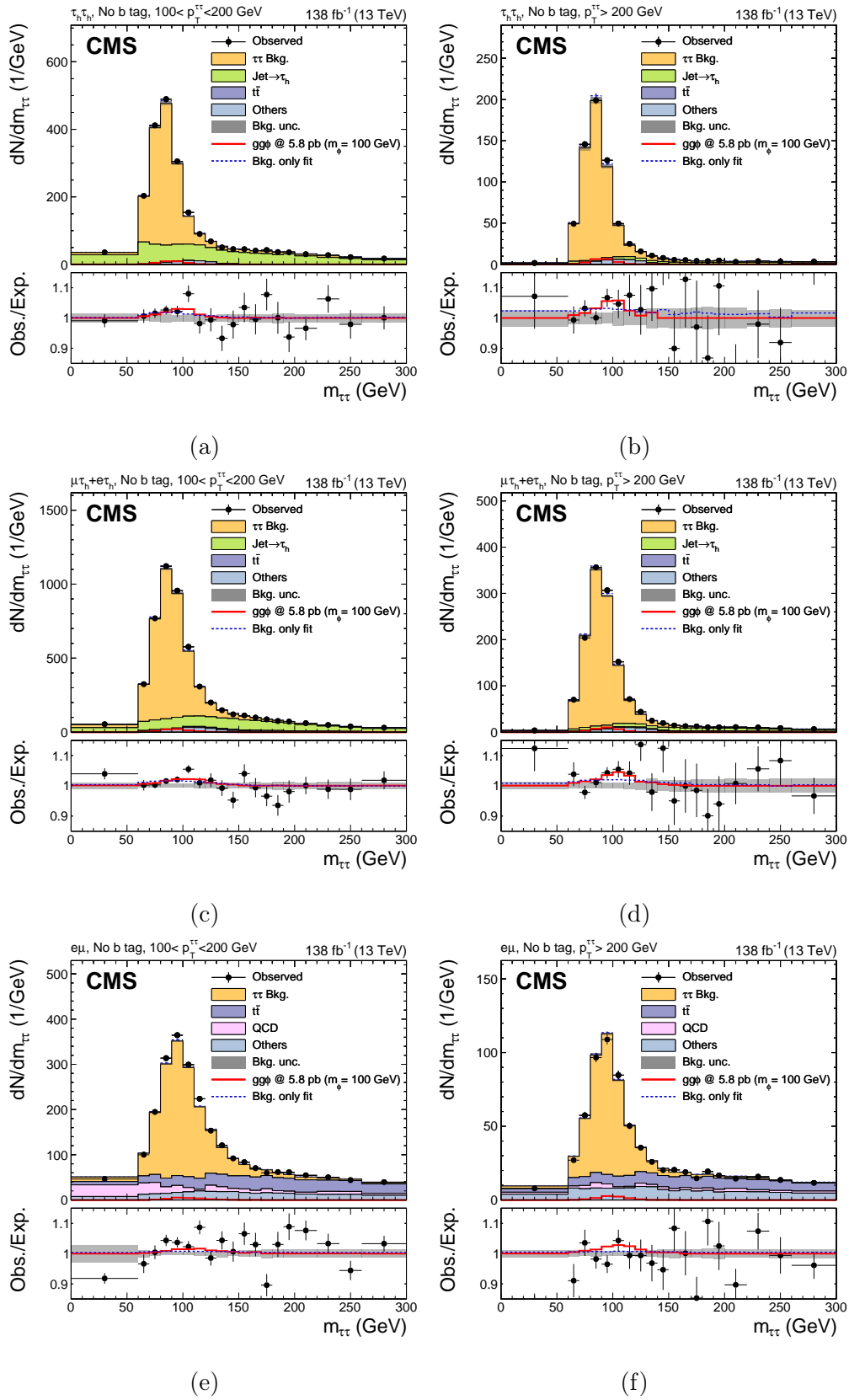


Figure 1.16: Low mass postfit.

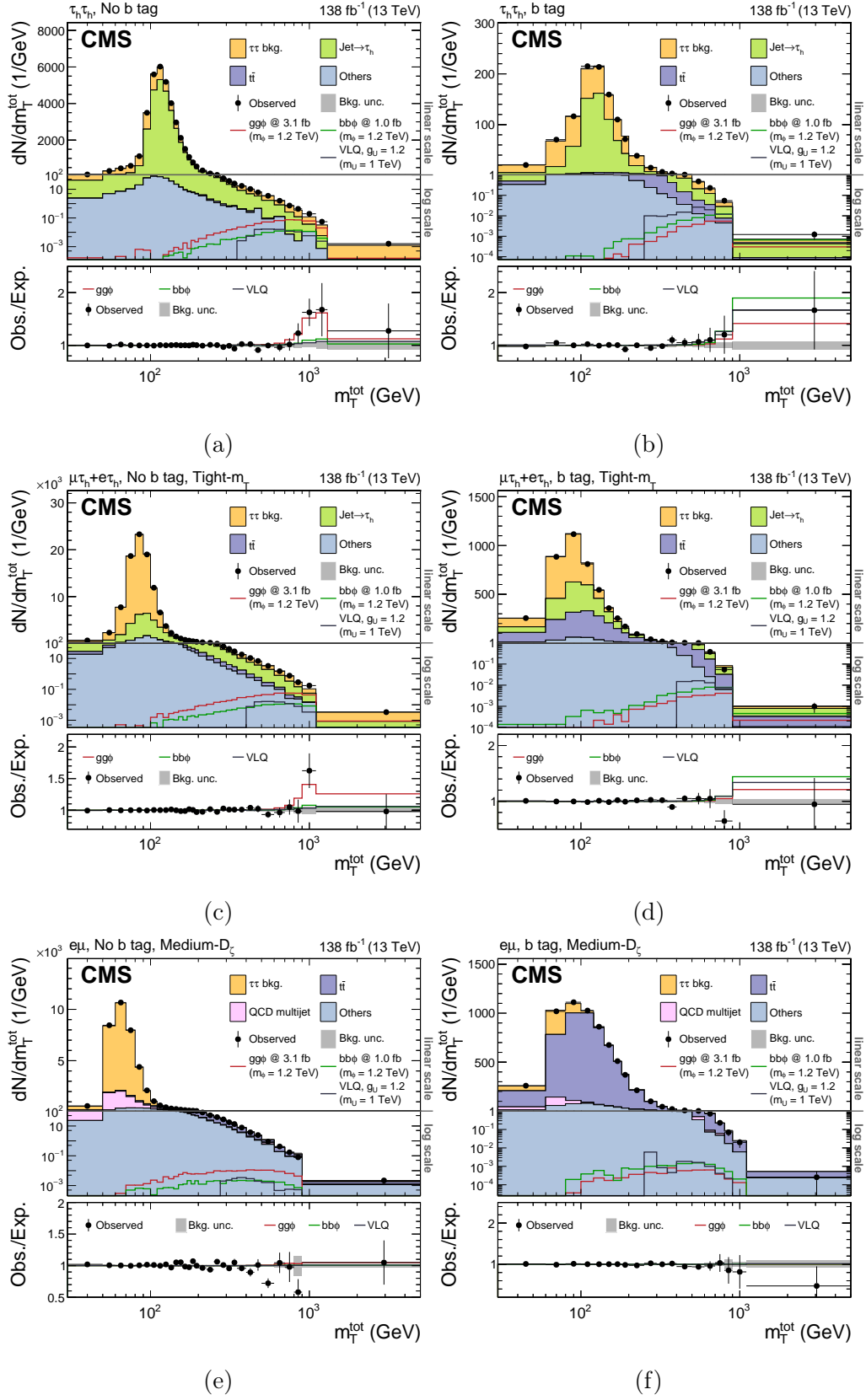


Figure 1.17: High mass postfit.



## 1.9 Postfit Plots

## 1.10 MC Corrections

## 1.11 Model Independent Results

### 1.11.1 Limit Setting

$$q_\mu = -2 \ln \left( \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta}_{\hat{\mu}})} \right), 0 \leq \hat{\mu} \leq \mu, \quad (1.7)$$

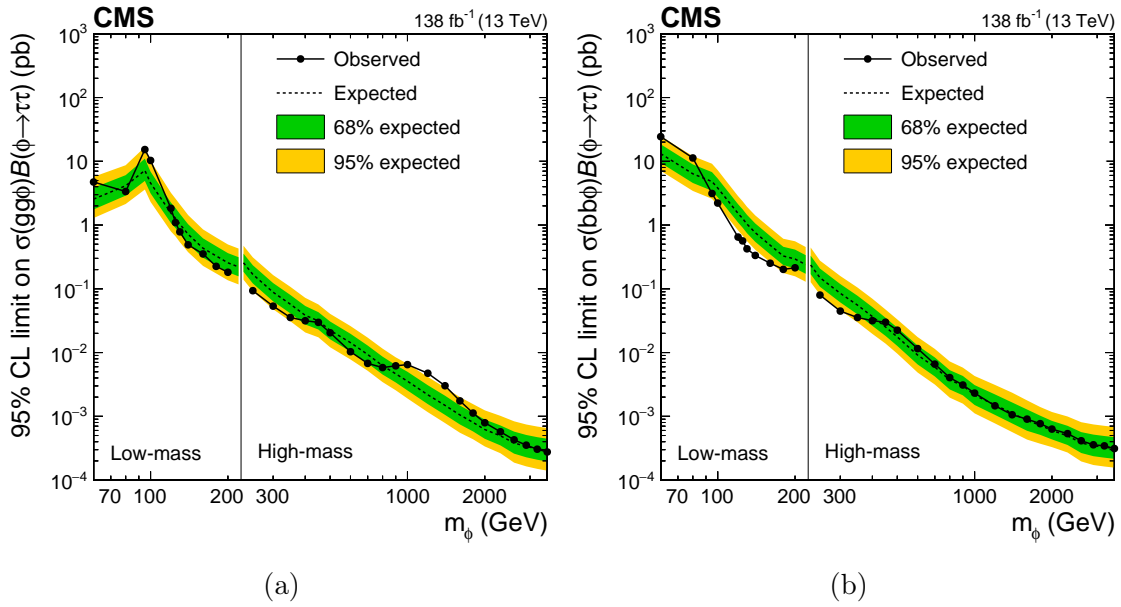


Figure 1.18: Model independent limits.

### 1.11.2 Significance and Compatibility

### 1.11.3 2D Likelihood Scans

## 1.12 Model Dependent Limits

### 1.12.1 Limit Setting

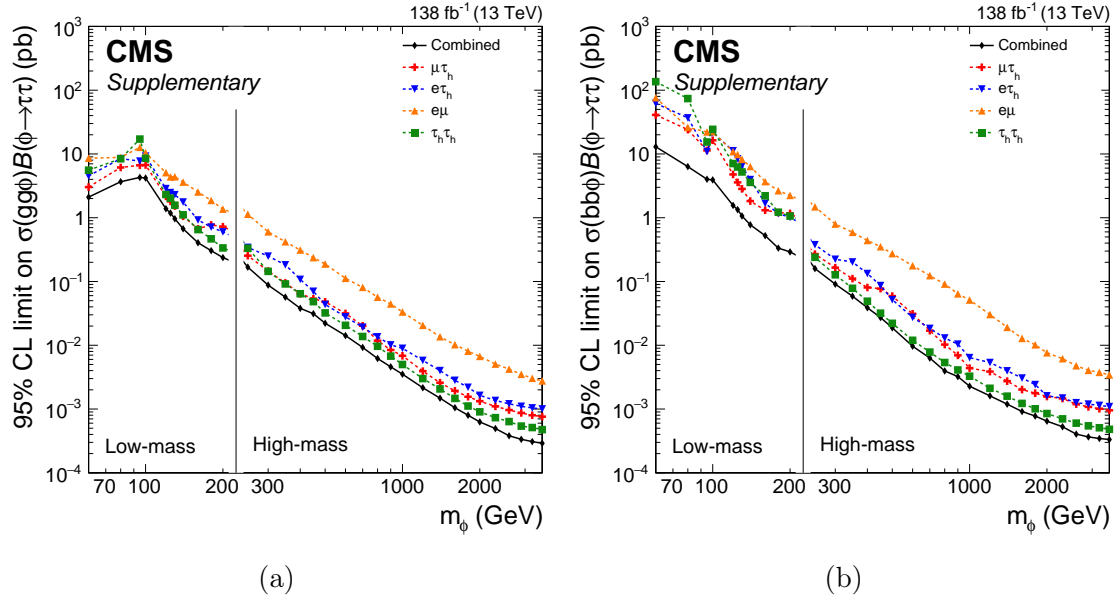


Figure 1.19: Model independent limits.

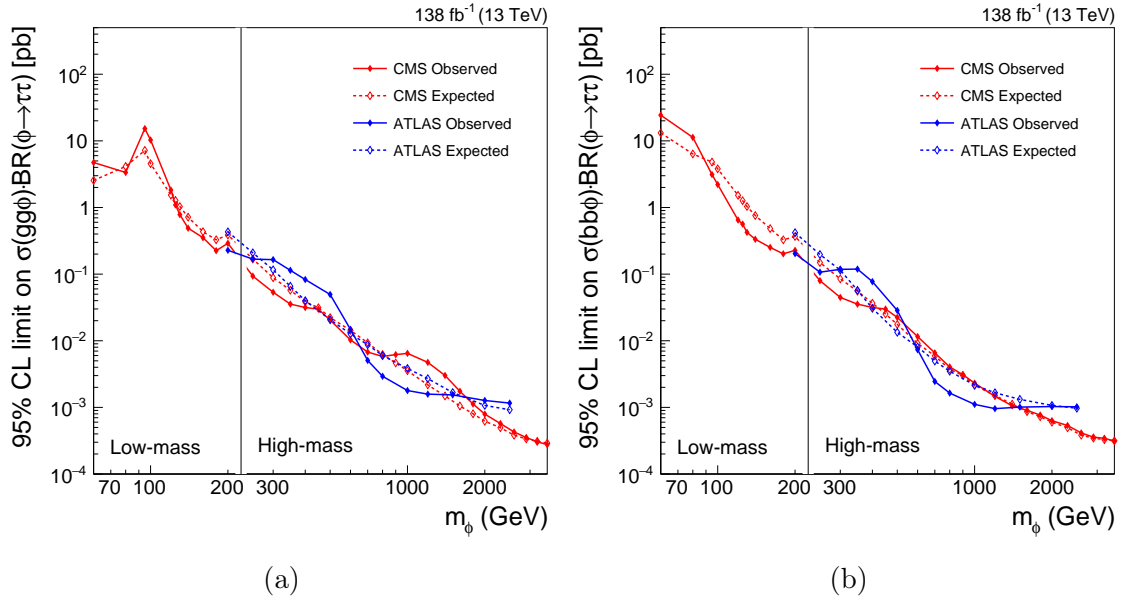


Figure 1.20: Model independent limits.

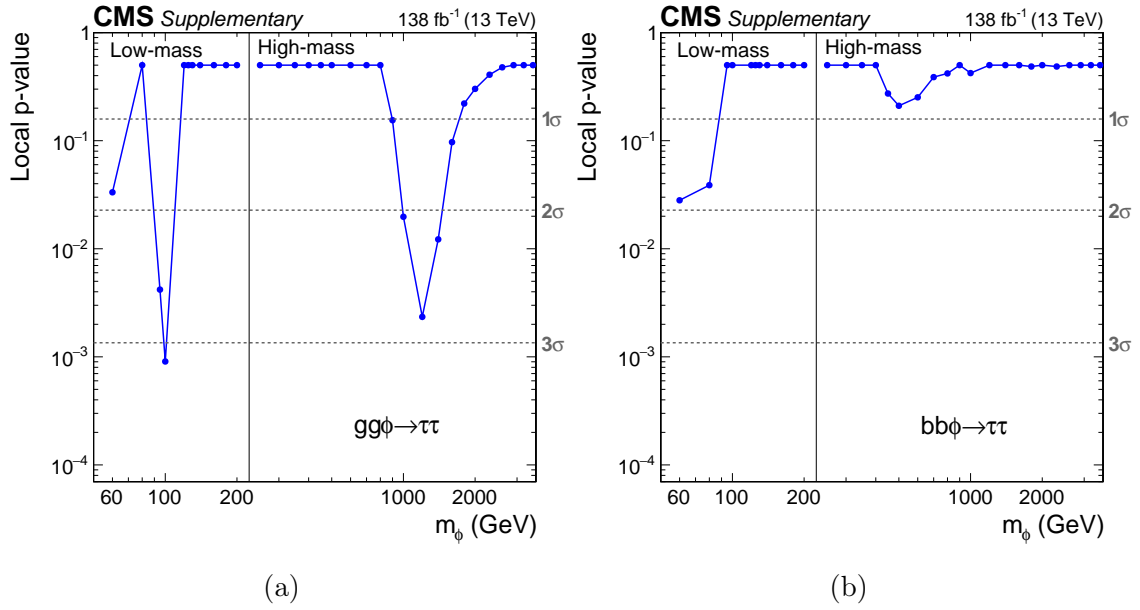


Figure 1.21: Significance.

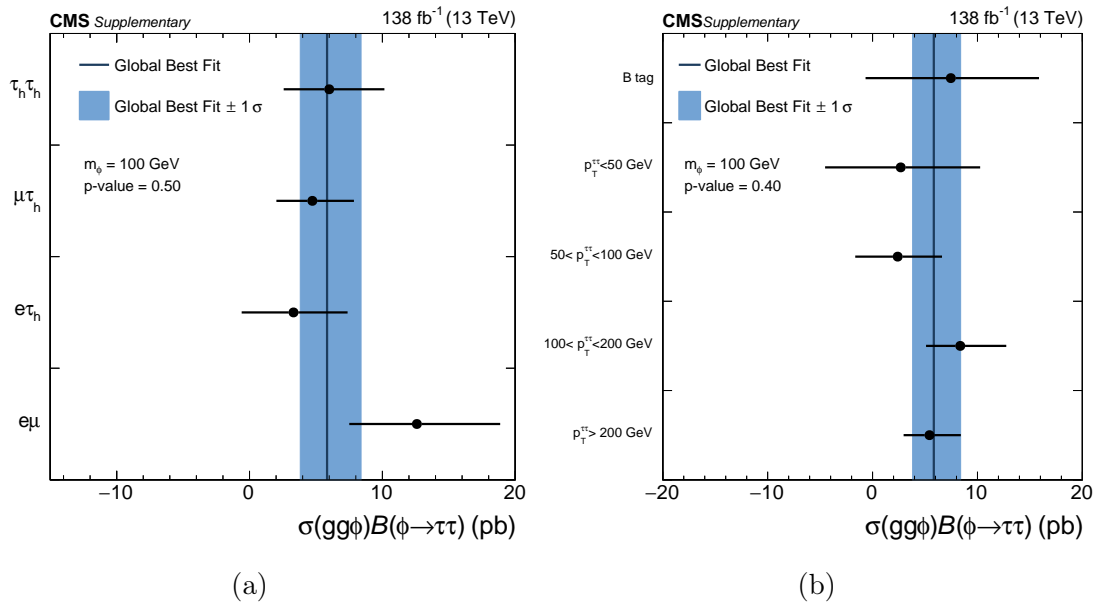


Figure 1.22: Low mass compatibility.

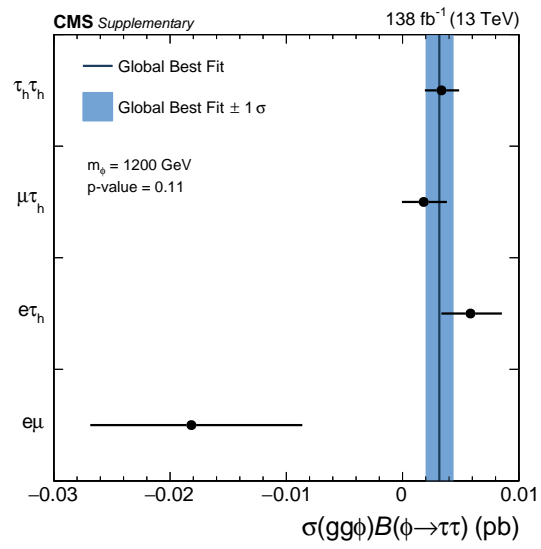


Figure 1.23: High mass compatibility.

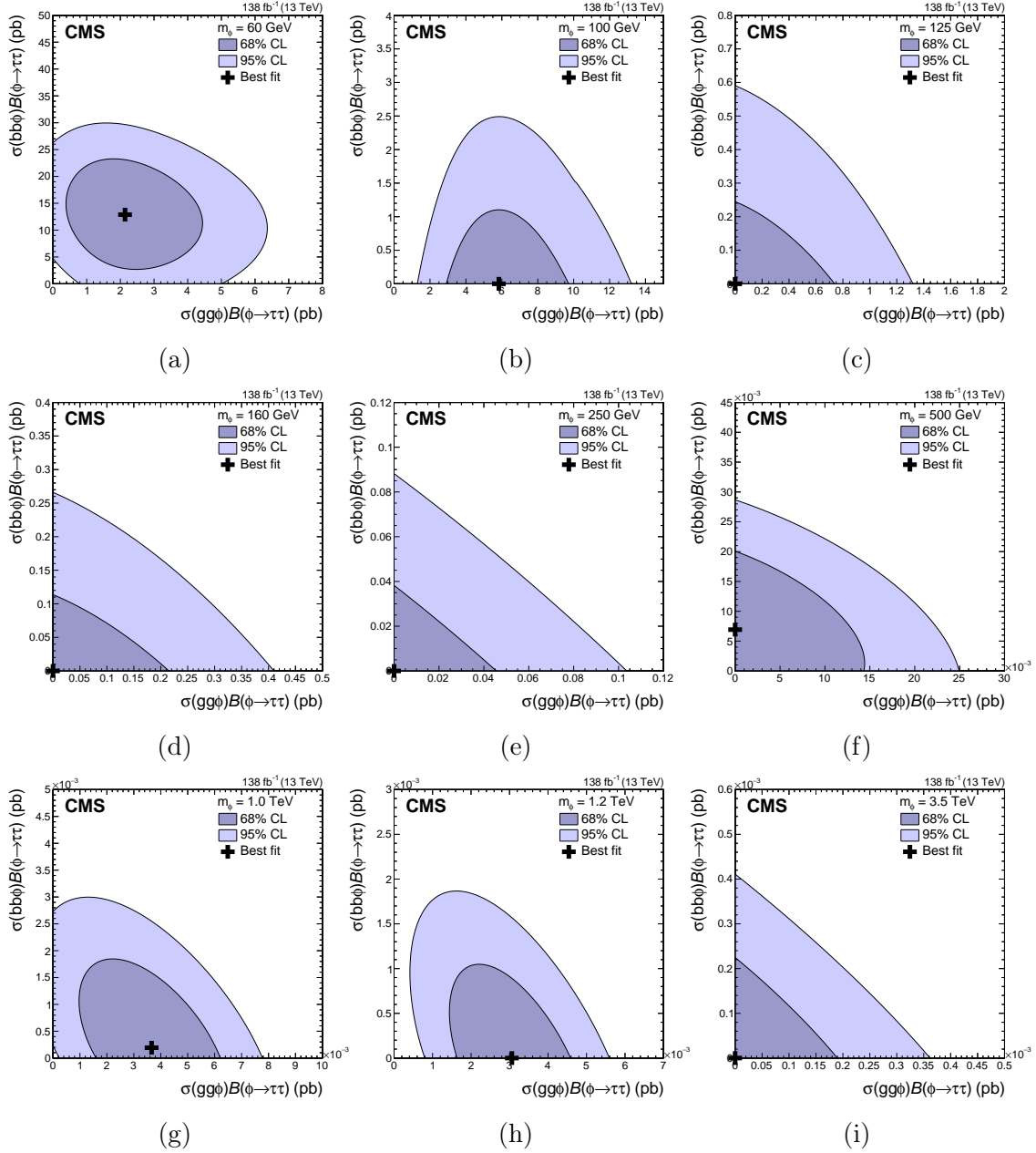


Figure 1.24: 2D Likelihood scans.

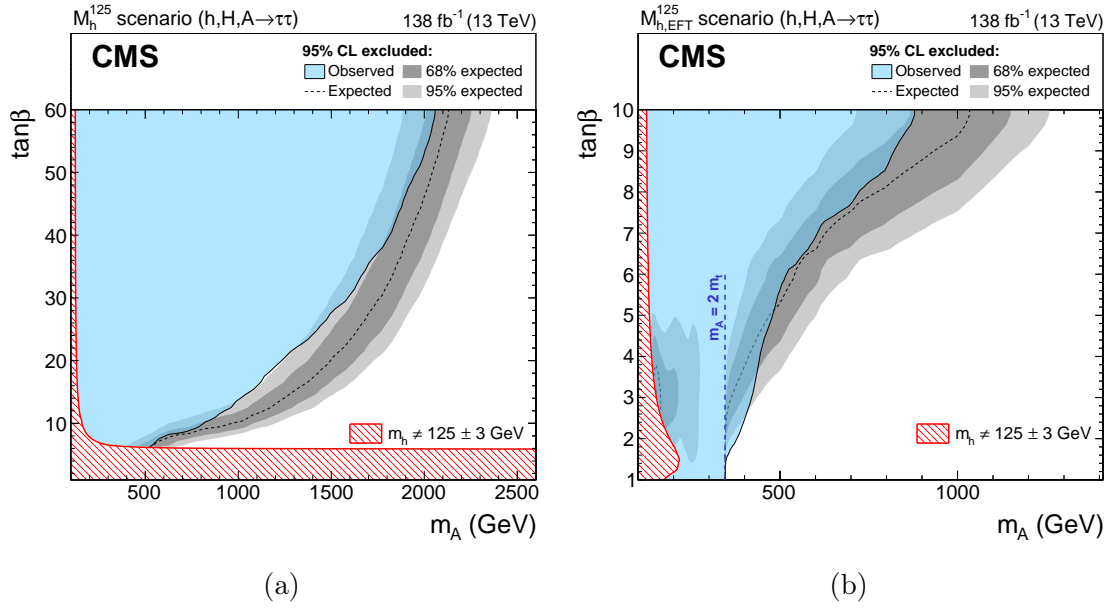


Figure 1.25: MSSM limits.

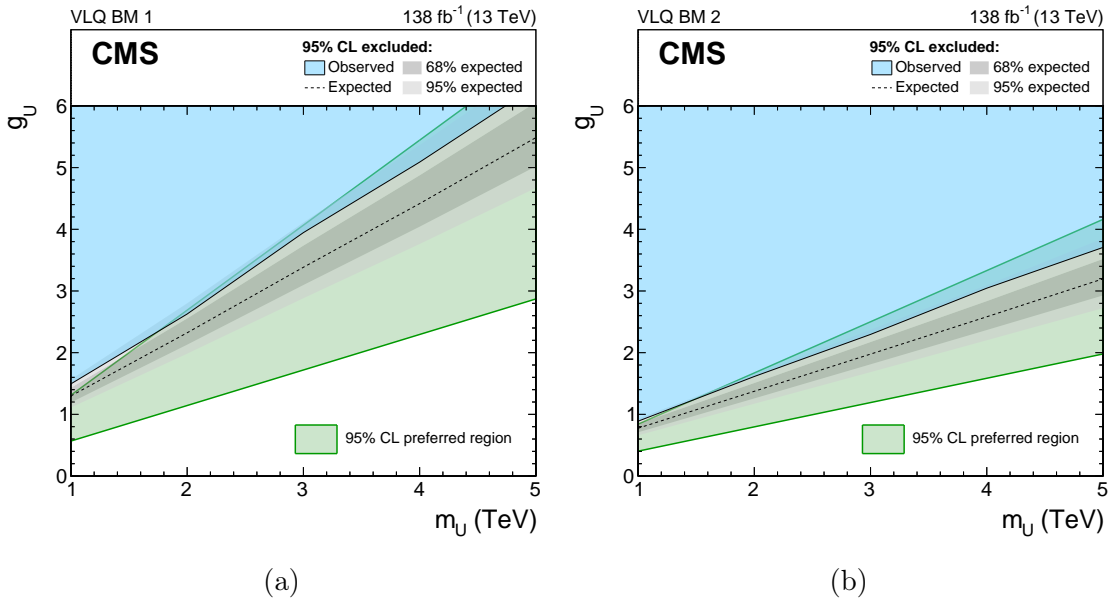


Figure 1.26: VLQ limits.