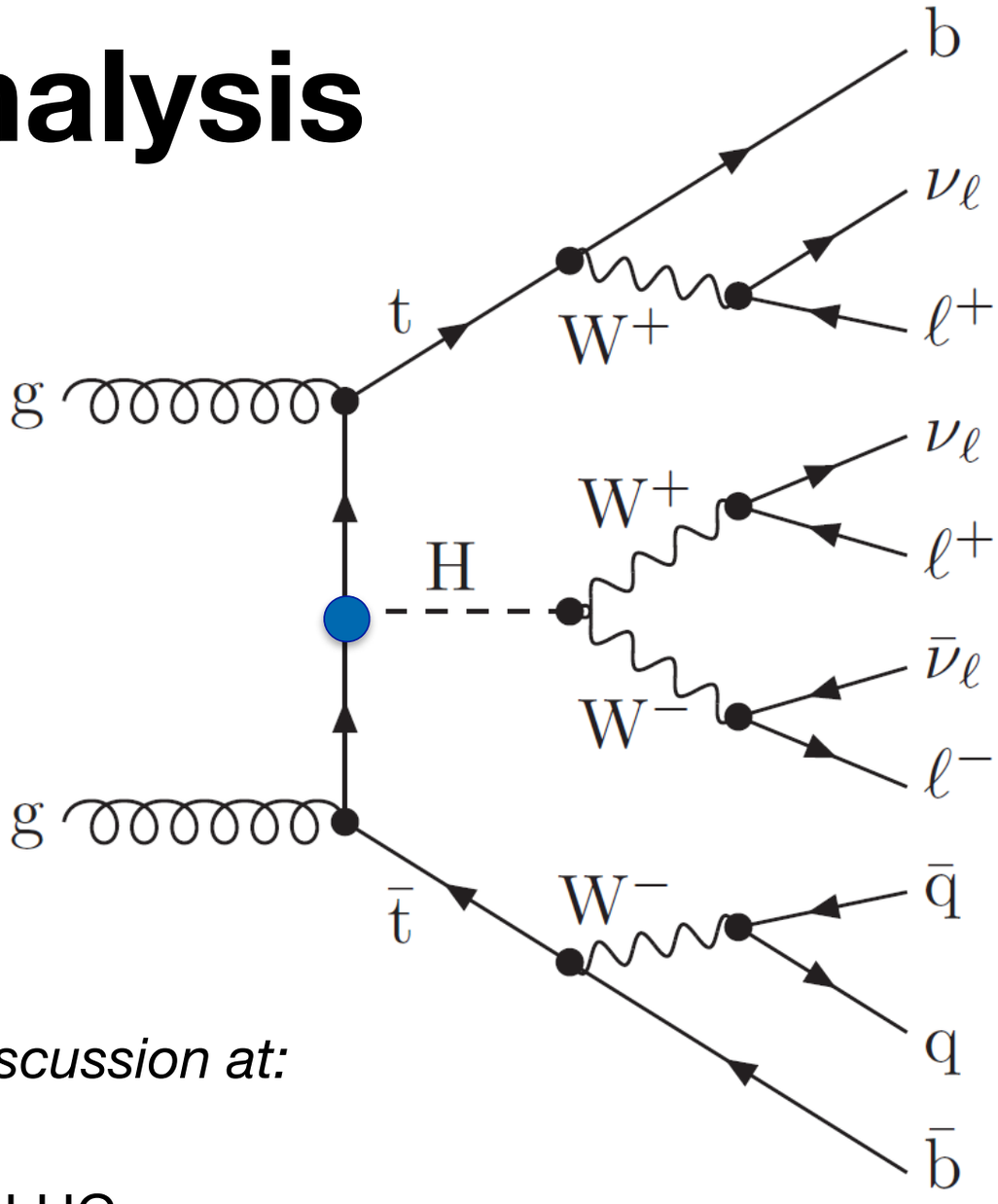
Data-driven background estimate in the CMS ttH multi-lepton analysis



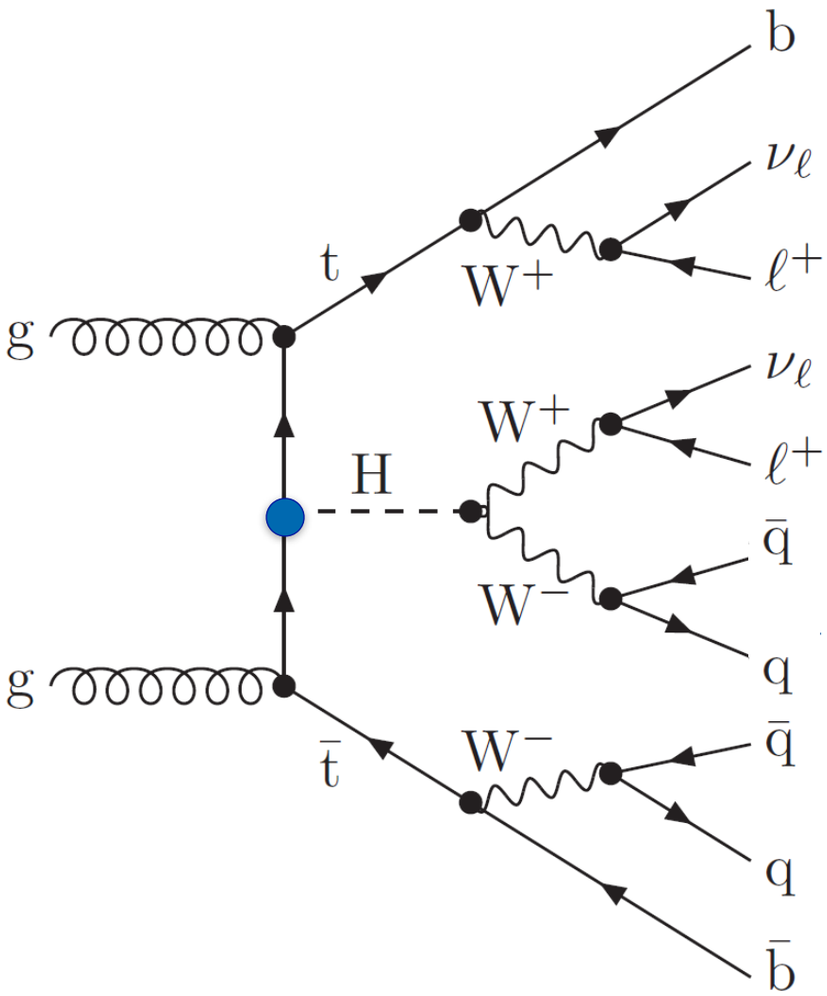
Marco Peruzzi (CERN)

Some material to continue the breakout session discussion at:

Higgs Toppings Workshop
Probing Top-Higgs Interactions at the LHC
Benasque, May 27th - June 2nd, 2018

Final states

CMS HIG-17-018
arXiv:1803.05485



- Targeting WW*, ZZ*, $\tau\tau$ Higgs decays
- Analysis categories:
  - **1 lepton + 2 τ_h**
  - **2 same-sign leptons + 0,1 τ_h**
  - **3 leptons + 0,1 τ_h**
  - **4 leptons**
- Jet multiplicity and b-tagging requirements

- **Irreducible backgrounds** (mainly ttV and di-boson), estimated from simulation
- **Non-prompt lepton** contribution from tt events, predicted from data

Data-driven background estimate in the CMS ttH multi-lepton analysis

Event categories

CMS HIG-17-018
arXiv:1803.05485

Selection	2ℓss	2ℓss $+ 1\tau_h$
Targetted $t\bar{t}H$ decay	$t \to b\ell\nu$, $t \to bqq$, $H \to WW \to \ell\nu qq$	$t \to b\ell\nu$, $t \to bqq$, $H \to \tau\tau \to \ell\tau_h + \nu$'s
Trigger	Single- and double-lepton triggers	
Lepton p_T	$p_T > 25$ / 15 GeV	$p_T > 25$ / 15 (e) or 10 GeV (μ)
τ_h p_T	—	$p_T > 20$ GeV
Charge requirements	2 same-sign leptons and charge quality requirements	2 same-sign leptons and charge quality requirements $\sum_{\ell,\tau_h} q = \pm 1$
Jet multiplicity	$\geq$4 jets	$\geq$3 jets
b tagging requirements	$\geq$1 tight b-tagged jet or $\geq$2 loose b-tagged jets	
Missing transverse momentum	$L_D > 30$ GeV	$L_D > 30$ GeV *
Dilepton mass	$m_{\ell\ell} > 12$ GeV and $\lvert m_{ee} - m_Z \rvert > 10$ GeV *	

Selection	3ℓ	$3\ell + 1\tau_h$
Targetted $t\bar{t}H$ decays	$t \to b\ell\nu$, $t \to b\ell\nu$, $H \to WW \to \ell\nu qq$ $t \to b\ell\nu$, $t \to bqq$, $H \to WW \to \ell\nu\ell\nu$ $t \to b\ell\nu$, $t \to bqq$, $H \to ZZ \to \ell\ell qq$ or $\ell\ell\nu\nu$	$t \to b\ell\nu$, $t \to b\ell\nu$, $H \to \tau\tau \to \ell\tau_h + \nu$'s
Trigger	Single-, double- and triple-lepton triggers	
Lepton p_T	$p_T > 25$ / 15 / 15 GeV	$p_T > 20$ / 10 / 10 GeV
τ_h p_T	—	$p_T > 20$ GeV
Charge requirements	$\sum_{\ell} q = \pm 1$	$\sum_{\ell,\tau_h} q = 0$
Jet multiplicity	$\geq$2 jets	
b tagging requirements	$\geq$1 tight b-tagged jet or $\geq$2 loose b-tagged jets	
Missing transverse momentum	No requirement if $N_j \geq 4$ $L_D > 45$ GeV † $L_D > 30$ GeV otherwise	
Dilepton mass	$m_{\ell\ell} > 12$ GeV and $\lvert m_{\ell\ell} - m_Z \rvert > 10$ GeV ‡	

* Applied only if both leptons are electrons.
† If the event contains a SFOS lepton pair and $N_j \leq 3$.
‡ Applied to all SFOS lepton pairs.

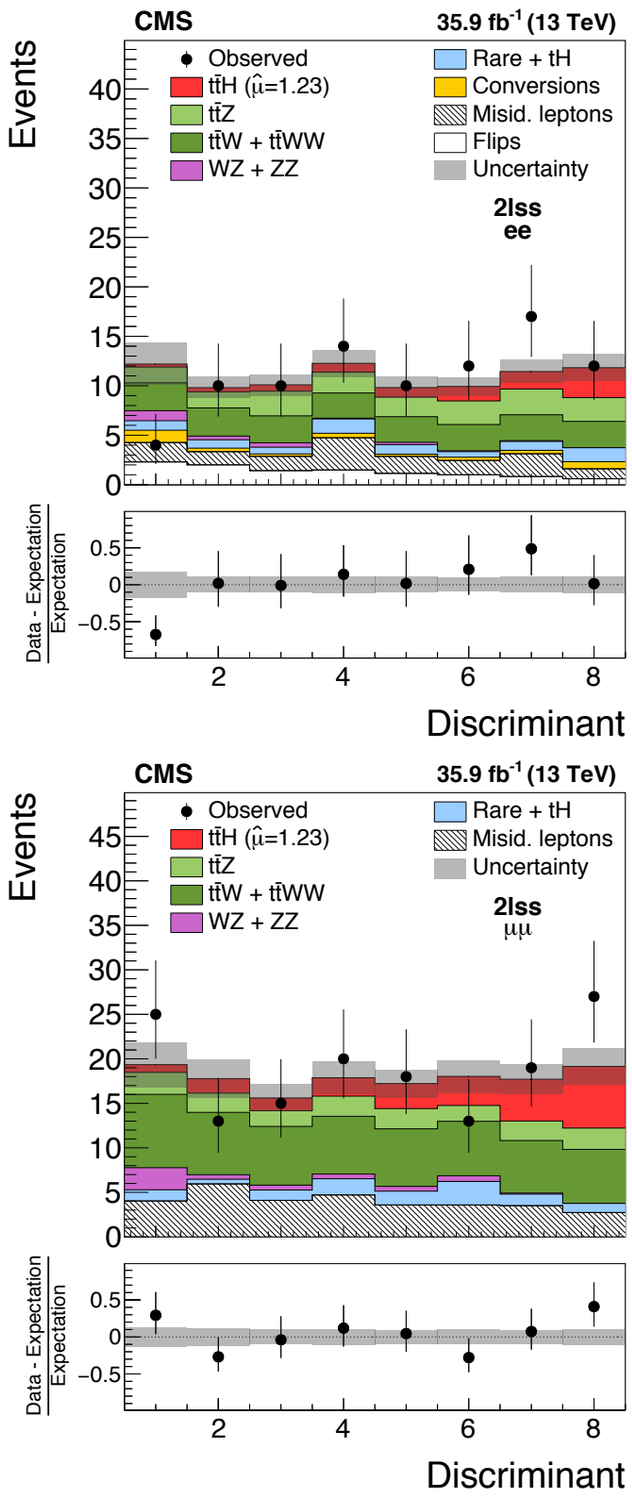
Selection	$1\ell + 2\tau_h$	4ℓ
Targetted $t\bar{t}H$ decays	$t \to b\ell\nu$, $t \to bqq$, $H \to \tau\tau \to \tau_h\tau_h + \nu$'s	$t \to b\ell\nu$, $t \to b\ell\nu$, $H \to WW \to \ell\nu\ell\nu$ $t \to b\ell\nu$, $t \to b\ell\nu$, $H \to ZZ \to \ell\ell qq$ or $\ell\ell\nu\nu$
Trigger	Single=lepton and lepton+τ_h triggers	Single-, double- and triple-lepton triggers
Lepton p_T	$p_T > 25$ (e) or 20 GeV (μ)	$p_T > 25$ / 15 / 15 / 10 GeV
τ_h p_T	$p_T > 30$ / 20 GeV	—
Charge requirements	$\sum_{\tau_h} q = 0$ and $\sum_{\ell,\tau_h} q = \pm 1$	$\sum_{\ell} q = 0$
Jet multiplicity	$\geq$3 jets	$\geq$2 jets
b tagging requirements	$\geq$1 tight b-tagged jet or $\geq$2 loose b-tagged jets	
Missing transverse momentum	—	No requirement if $N_j \geq 4$ $L_D > 45$ GeV † $L_D > 30$ GeV otherwise
Dilepton mass	$m_{\ell\ell} > 12$ GeV	$m_{\ell\ell} > 12$ GeV and $\lvert m_{\ell\ell} - m_Z \rvert > 10$ GeV ‡
Four-lepton mass	—	$m_{4\ell} > 140$ GeV §

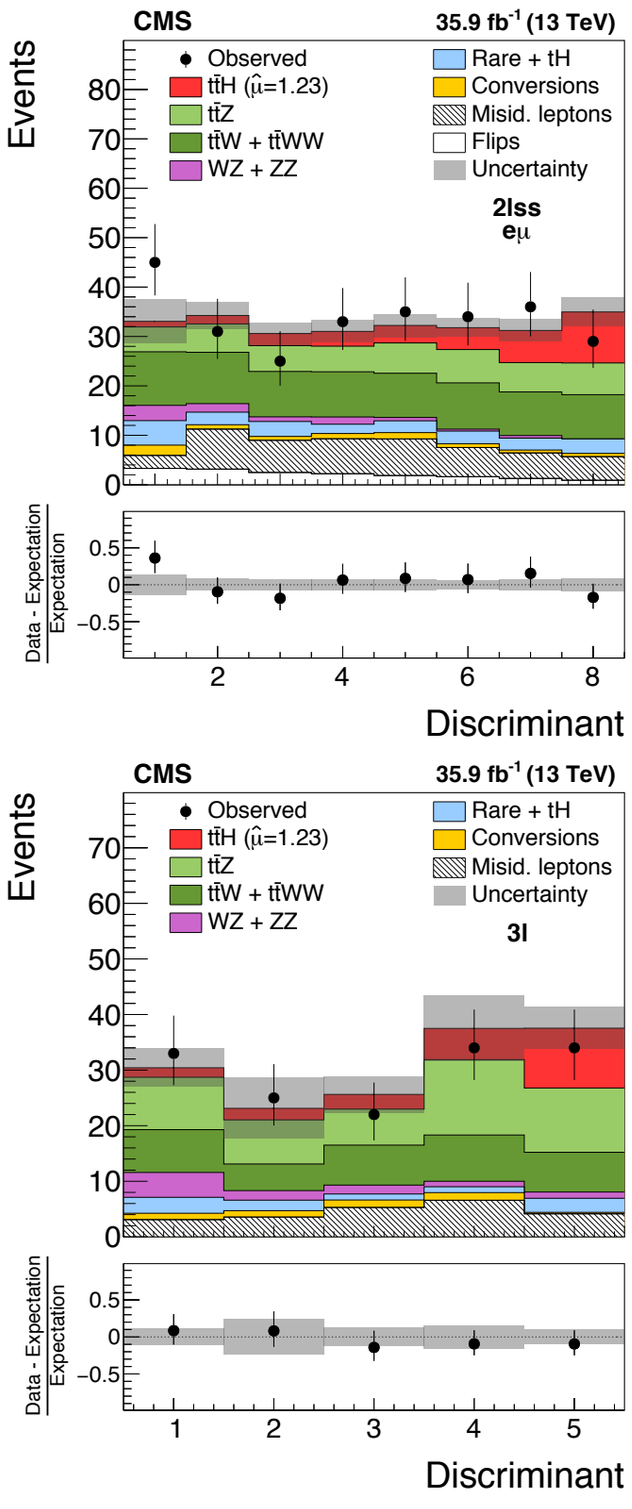
† If the event contains a SFOS lepton pair and $N_j \leq 3$.
‡ Applied to all SFOS lepton pairs.
§ Applied only if the event contains 2 SFOS lepton pairs.

- Large set of kinematic requirements define the categories used in the fit
- Estimation of non-prompt lepton background performed with the same strategy everywhere

Event yields

CMS HIG-17-018
arXiv:1803.05485

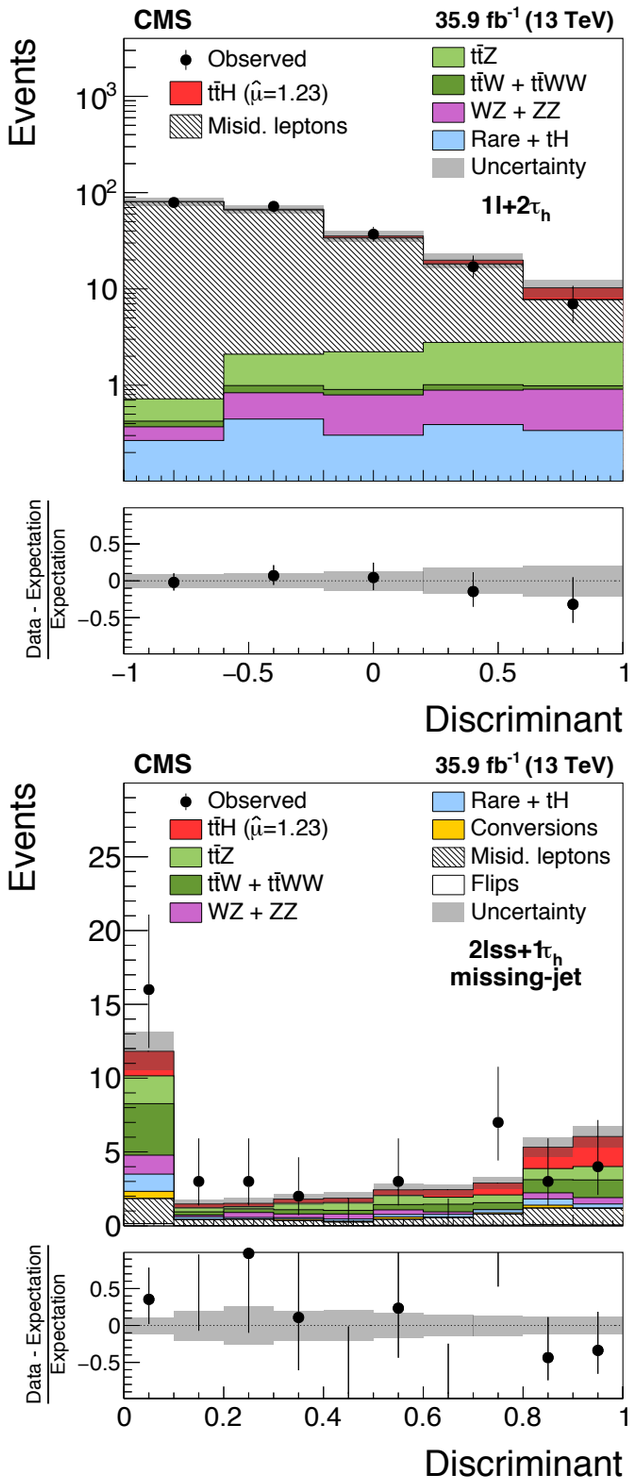


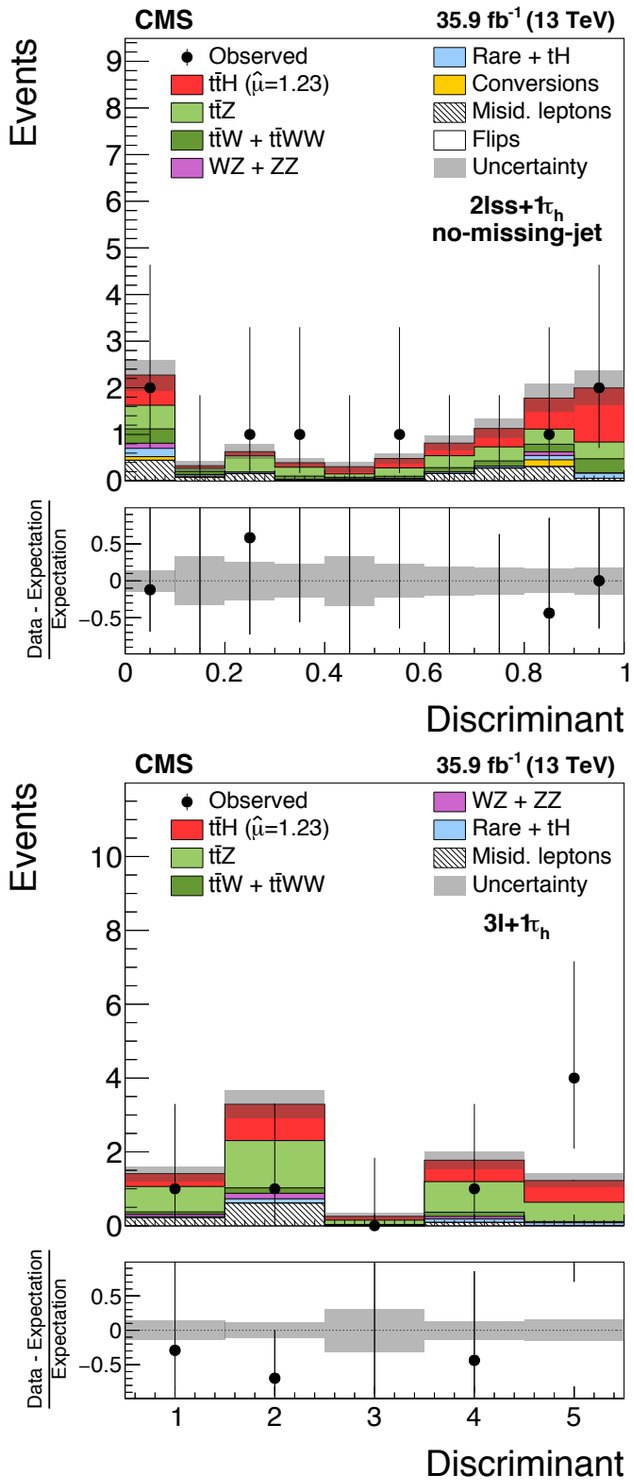


- Yields in:
  - **2 same-sign leptons + 0 τ_h**
  - **3 leptons + 0 τ_h**
- Relative contribution varies as a function of b-tag requirements and event kinematics exploited by multivariate discriminants
- Small contribution from electrons from photon conversions
  - estimated from simulation

Event yields

CMS HIG-17-018
arXiv:1803.05485





- Yields in:
  - **1 lepton + 2 τ_h**
  - **2/3 leptons + 1 τ_h**
- Large yield of non-prompt τ_h from hadronic jets in 1 lepton category
- Categories with more leptons are very pure: non-prompt background is sub-dominant there

Background prediction

- Based on a **“fake factor” method**:
  - **relax** the lepton selection to a looser definition
    which does not include the requirement on the “lepton MVA”
  - **weight** events that have at least one lepton failing the tight selection
    to obtain the data-driven background prediction in the signal region
  - input needed: probability for a non-prompt lepton to pass the tight cut

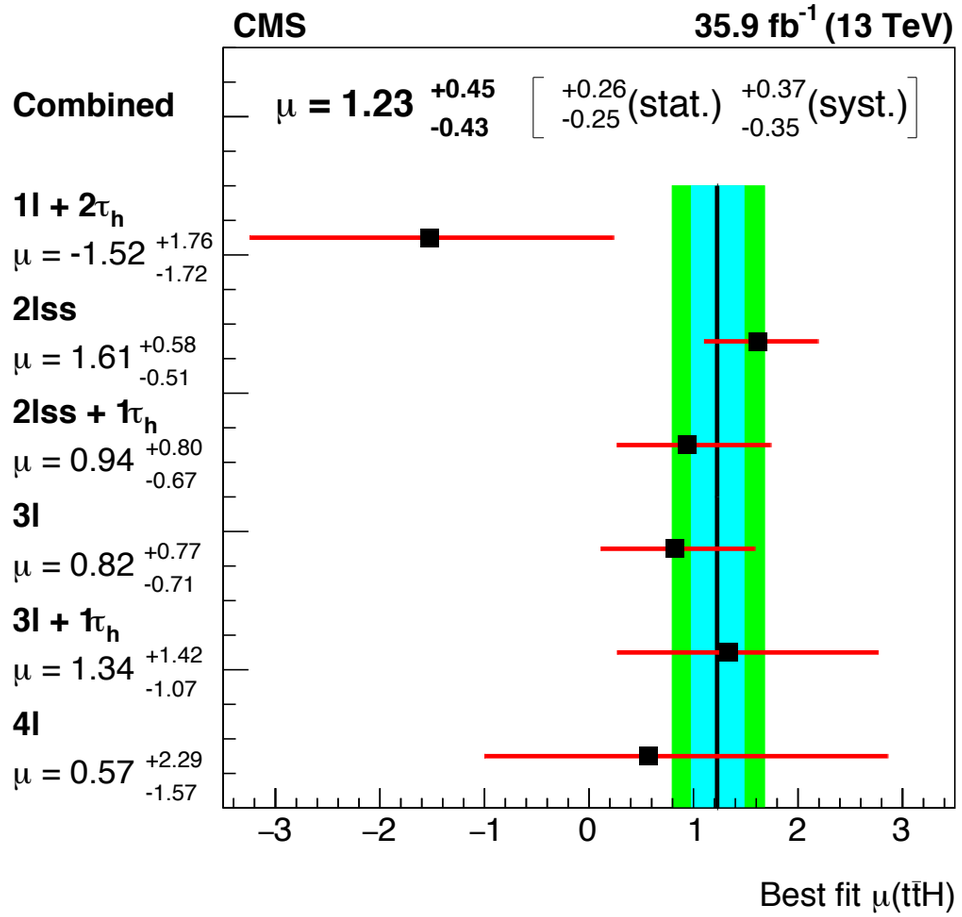
- Light lepton fake rate measured in a **QCD-enriched control region**:
  - as a function of corrected lepton p_T and η
  - use transverse mass to reduce contamination from EWK processes

- **Closure** of the method:
  - how **universal** is the fake rate for different sources of non-prompt leptons?
  - selection tuned to control potential flavour dependence of the fake rate

Systematic uncertainties

CMS HIG-17-018
arXiv:1803.05485

Source	Uncertainty [%]	$\Delta\mu/\mu$ [%]
e, μ selection efficiency	2–4	11
τ_h selection efficiency	5	4.5
b tagging efficiency	2–15 [57]	6
Reducible background estimate	10–40	11
Jet energy calibration	2–15 [65]	5
τ_h energy calibration	3	1
Theoretical sources	≈10	12
Integrated luminosity	2.5	5



- Uncertainties on the fake background have an impact of about 11% on the fitted signal strength:
  - similar impact to efficiency measurements and theoretical sources
- Both uncertainties in the measurement of the fake rate and in the closure of the method in different categories are taken into account