Estimation des bruits de fond avec fake lepton dans ttH multilepton

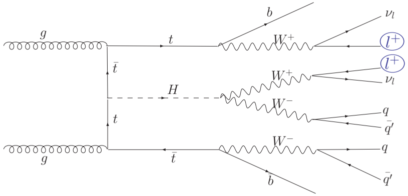
A.Chomont

LPC Clermont-Ferrand
TopLHC France

May 20th 2016

Multileptons signatures

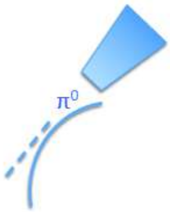
- Decays of Higgs to WW, ZZ and $\tau\tau$ targeted
- Large number of different channels can be considered
  - Light leptons channels: 2lSS, 3l, 4l
  - Light+tau channels: 2τ+1l, 2lSS+τ, 2lOS+τ, (l+τ)SS, 2τ+jets
- Other characteristics of those channels: 1 b-tagged jets, at least 4 jets
  - Main backgrounds: $t\bar{t}V$, events with fakes leptons, with misidentification of lepton charge (2 leptons SS channels)

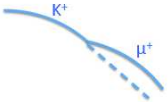


Source	$\Delta\mu$	
$2\ell 0\tau_{\rm had}$ non-prompt muon transfer factor	+0.38	−0.35
$t\bar{t}W$ acceptance	+0.26	−0.21
$t\bar{t}H$ inclusive cross section	+0.28	−0.15
Jet energy scale	+0.24	−0.18
$2\ell 0\tau_{\rm had}$ non-prompt electron transfer factor	+0.26	−0.16
$t\bar{t}H$ acceptance	+0.22	−0.15
$t\bar{t}Z$ inclusive cross section	+0.19	−0.17
$t\bar{t}W$ inclusive cross section	+0.18	−0.15
Muon isolation efficiency	+0.19	−0.14
Luminosity	+0.18	−0.14

Fake leptons

- Leptons fakes are objects reconstructed as prompt leptons, leptons coming from a W boson, a Z boson or a τ (decay results of top or Higgs)
  - Jets
  - Non prompts leptons due to decays of b-hadrons for example $\rightarrow$ majority of the cases (checked at truth level simulation)
  - Trident process with an electron radiating a photon converting to a pair of electrons
- Focus on fake light leptons (regions 2l/3l/4l)



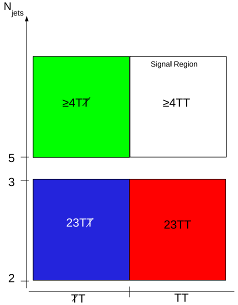


- Process originating fakes for $t\bar{t}H$ multileptons analysis: $t\bar{t}$(by far the main one), single top, dibosons . . .

Methods to extract Fakes yields

- Estimation of this background done using Data-Driven technics
- Three methods to extract the fakes yields in the $t\bar{t}H$ multileptons analysis are used
  - Fake factor method $\rightarrow$ used for Run 1 and Run 2 (Clermont-Fd)
  - Matrix method $\rightarrow$ used for Run 2 (Marseille)
  - MC reweighting using DD factors $\rightarrow$ used for Run 2 (Marseille)

Fake factor method



- Anti-Tight lepton
  - Obtained by inverting some selection on the tight electrons: for example electron not isolated or reversal of p_T cut used for tight muons ...
  - Definition of Anti-Tight lepton will depend on the choice made for signal region object definition

- Fake factor θ is defined as (for electrons):

$$\theta_e = \frac{TT}{T\not{T}}(low\_jets) = \frac{TT(N_{ee}^{data} - N_{ee}^{PromptSS} - N_{ee}^{QMisId})}{T\not{T}(N_{e\not{e}}^{data} - N_{e\not{e}}^{allPrompt})}$$

- PromptSS: $t\bar{t}V$, VV
- QMisId: prompt opposite-sign events with a charge mis-identification (data-driven in TT region)

Fake factor method

- Number of fakes in signal region obtained from θ_e, θ_μ for 2l and 3l channels
  - for $e^\pm e^\pm$ region:

    $N_{ee}(high\_njets) = N_{e\not{e}}(high\_njets) \times \theta_e$
  - for $\mu^\pm \mu^\pm$ region:

    $N_{\mu\mu}(high\_njets) = N_{\mu\not{\mu}}(high\_njets) \times \theta_\mu$
  - for $e^\pm \mu^\pm$ region:

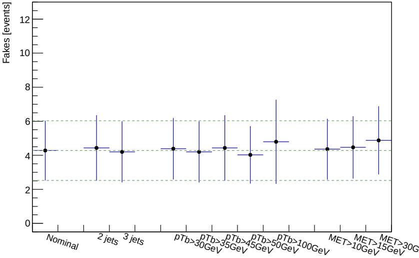
    $N_{e\mu}(high\_njets) = N_{e\not{\mu}}(high\_njets) \times \theta_\mu + N_{\mu\not{e}}(h\_njets) \times \theta_e$
  - for 3l region:

    $N_{estimated\ \ fakes,SR} = N_{ll\not{e}} \times \theta_e + N_{ll\not{\mu}} \times \theta_\mu$

Uncertainties on the fakes yields

- Expected statistical precision on θ and on the data size of $T\not{T}(\geq 5$ jets) region: from 25 to 55%
- Validity of the extrapolation from low jets multiplicity region to high jets multiplicity region
  - Closure test performed on simulated $t\bar{t}$ events
  - Comparison of real ss fakes in signal region to number predicted by $N_{l\not{l}} \times \theta$

- Uncertainty on substracted backgrounds (QMisId, PromptSS): $\sim$ 25%
- Composition of low jets multilplicity region: 7 to 20%
  - Presence of additional non-$t\bar{t}$ fake sources, prompt processes w.r.t signal region $\rightarrow$ bias on the θ estimation
  - Estimated by changing definition of low multiplicity region (Cut on MET, HT ...)



Definition of the Matrix Method (MM)

- Extension of transfer factor method, allowing extraction of shapes by weighting events
  - As for fake factor method, DD method
- Weights applied function of Real lepton efficiency (r) and Fake lepton efficiency (f) to pass the tight requirement, with r$\gg$ f

$$r = \frac{N_R^T}{N_R^l} \quad f = \frac{N_F^T}{N_F^l}$$

- $N_R^T (N_F^T)$ is the number of Real (Fake) leptons passing the tight selection obtained in samples enriched in real leptons
- $N_R^L (N_F^L)$ is the number of Real (Fake) leptons not passing the tight selection obtained in samples enriched in fake leptons
- $N_R^l = N_R^T + N_R^F$
- To estimate r and f, regions enriched in prompt (fake) leptons used

Matrix Method for 2l channel

- Number of events with real and fake leptons to the events related with tight and loose leptons using a 4x4 matrix
- N_{TT}, N_{TL}, N_{LT}, N_{LL} numbers of events with 0,1 (leading or not),2 loose leptons
- N_{ll}^{RR}, N_{ll}^{RF}, N_{ll}^{FR}, N_{ll}^{FF} numbers of events with 2 (signal events),1 or 0 real leptons
- Those numbers are related using:

$$\begin{pmatrix} N_{TT} \\ N_{TL} \\ N_{LT} \\ N_{LL} \end{pmatrix} = \begin{pmatrix} r_1 r_2 & r_1 f_2 & f_1 r_2 & f_1 f_2 \\ r_1(1-r_2) & r_1(1-f_2) & f_1(1-r_2) & f_1(1-f_2) \\ (1-r_1)r_2 & (1-r_1)f_2 & (1-f_1)r_2 & (1-f_1)f_2 \\ (1-r_1)(1-r_2) & (1-r_1)(1-f_2) & (1-f_1)(1-r_2) & (1-f_1)(1-f_2) \end{pmatrix} \begin{pmatrix} N_{ll}^{RR} \\ N_{ll}^{RF} \\ N_{ll}^{FR} \\ N_{ll}^{FF} \end{pmatrix}$$

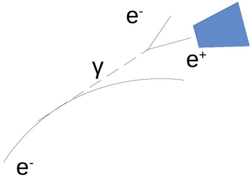
- From inverting the matrix, number of fake in the TT region (signal region) can be obtained
- Fakes in 3l region can be estimated in the same way with a 9x9 matrix

Matrix Method particularities

- Advantages of Matrix Method
  - Extraction of fakes efficiencies in 2D
    - Leptons p_T, η ...
  - Shape for Fakes events in SR to be achieved
- Limitations of MM
  - Suppose that r$\gg$ f $\rightarrow$ can be difficult in case of low statistics but corrected by:
    - Loose leptons to be far looser than tight leptons to cope with statistics
    - Reduce binning in lepton p_T, η $\rightarrow$ now reduce method flexibility
    - Likelihood MM $\rightarrow$ real efficiency and fake rates constraint from fit in fake enriched region

Charge MisId estimation (Clermont)

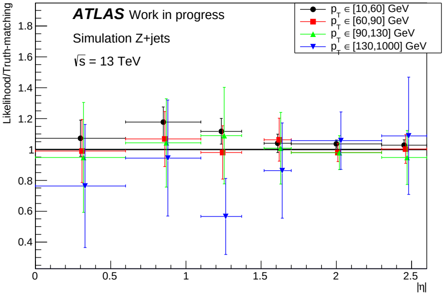
- Mis-identification of the charge of a lepton (QMisId) is an important background originating from two processes
  - High p_T electron with straight track
  - Trident process with an electron radiating a photon converting to a pair of electrons



- Negligible effect on muon

QMisid: method for rates estimation

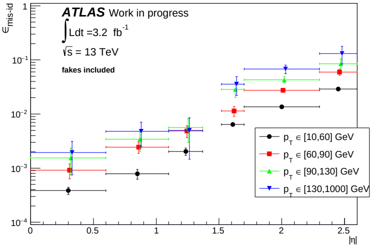
- Rate of QMisid computed from $Z \to e^+e^-$ mass peak region and used to reweight OS data
- ϵ_i rate of charge Misid for a single electron in region i (regions defined in η, p_T,E ...) and we obtain for N_{tot} true opposite-sign events:
$$N_{ss} = N_{tot}[(1-\epsilon_i)\epsilon_j + (1-\epsilon_j)\epsilon_i] \simeq N_{tot}(\epsilon_i + \epsilon_j)$$
- The rates, ϵ_i and ϵ_j, are obtained by likelihood minimization and are highly dependent on the choice of the binning



- Closure test: good agreement between rates from LH method and truth matching

QMisid: Rates estimation for 2015 data

- Rates obtained using Likelihood method from $3.2 fb^{-1}$of data
- Rates for last bin in p_T obtained by extrapolation of rates in the next to last bin in p_T (bin [90,130]GeV)



- Uncertainties included:
  - Statistical uncertainty from the likelihood method (main uncertainty)
  - Statistical uncertainty on the p_T dependent correction factor (last p_T bin, p_T >130GeV)
  - Difference between rates from truth matching and likelihood method on Z samples
  - Stability of rates due to definition of Z-peak region definition

Conclusion

- Considered channels for $t\bar{t}H$ studies relie on leptons presence $\rightarrow$ Misidentification of these leptons is an important uncertainties source
- Processes leading to events with fakes or QMisId leptons events among the main backgrounds in the $t\bar{t}H$ multileptons analysis
- Two methods presented to estimated the presence of the fake leptons background in the signal regions
- Fake Factor method simplification of the matrix method
  - Rates for MM can be estimated vs p_T (or η ...) of the leptons $\rightarrow$ better for analysis based on shapes
  - Thetas factors estimated using Fake Factor method inclusive in p_T
- QMisId events also estimated using a likelihood method