Fake estimation in the ttW+jets analysis

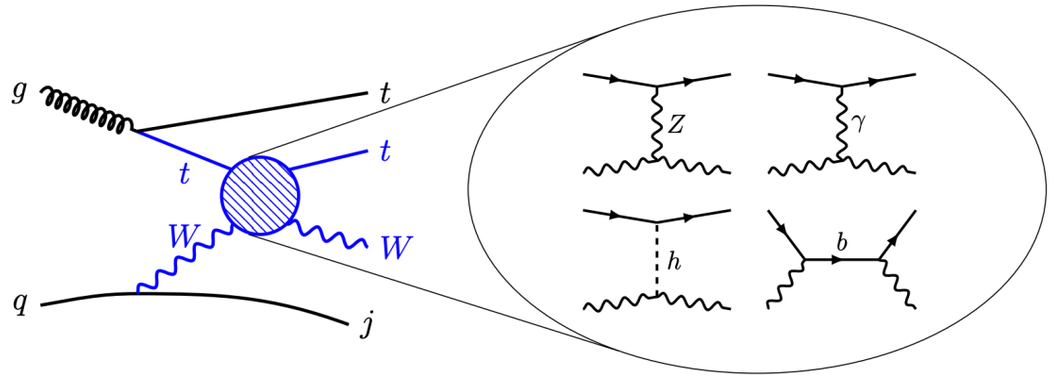
Laura Barranco Navarro, **Jörgen Sjölin**, Nabila Shaikh

06.04.2020

Isolation+Fakes Forum Meeting

ttW+jets measurement

arXiv:1511.03674v2



- *tW->tW* scattering allows to access *ttZ* couplings
  - *ttWj* signature at the LHC
  - Measurement challenging
  - Interesting: strong dependence in cross section in the presence of EFT couplings
  - Effective dim-6 lagrangian affecting *ttZ* and *bbZ* SM couplings

$$\Delta\mathcal{L}_t = \frac{i\bar{c}_L^{(1)}}{v^2} H^\dagger \overleftrightarrow{D_\mu} H \bar{q}_L \gamma^\mu q_L + \frac{i\bar{c}_L^{(3)}}{v^2} H^\dagger \sigma^a \overleftrightarrow{D_\mu} H \bar{q}_L \gamma^\mu \sigma^a q_L$$
$$+ \frac{i\bar{c}_R}{v^2} H^\dagger \overleftrightarrow{D_\mu} H \bar{t}_R \gamma^\mu t_R + \frac{i\bar{c}_R^b}{v^2} H^\dagger \overleftrightarrow{D_\mu} H \bar{b}_R \gamma^\mu b_R + \left(\frac{i\bar{c}_R^{tb}}{v^2} \tilde{H}^\dagger \overleftrightarrow{D_\mu} H \bar{t}_R \gamma^\mu b_R + \text{h.c.} \right)$$
$$+ \frac{\bar{c}_u y_t}{v^2} H^\dagger H \bar{q}_L \tilde{H} t_R + \text{h.c.} \ ,$$

$$\bar{c}_R = \frac{v^2}{\Lambda^2} c_{\phi t} = \frac{v^2}{\Lambda^2} c_{\phi u}^{(33)}$$

- Related to Zt_Rt_R SM
- Weakly constrained by standard measurements
- tW scattering very sensitive to it if there is new physics

Object reconstruction and event preselection

- **Electrons offline**:
  - IDTightLH + ECIDS > -0.33
  - Isolation: FCTight
- **Muons. offline**:
  - Medium
  - Isolation: FCTight
- **Event preselection**:
  - 2 same-sign leptons: p_T (sub)leading lepton > (20)27 GeV
  - Veto Z window in *ee* channel
  - m_{ll} > 30 GeV
  - Electrons abs(η)<2
  - At least 3 jets
  - At least 1 *b*-jet
  - Leading jet p_T > 60 GeV (not included for real and fake rate measurement)

Loose and Tight lepton definitions

- Different identification and isolation requirements are used in the selection of tight and loose leptons

	Loose	Tight
Electrons	LooseAndBLayerLH No isolation ECIDS	TightLH FCTight ECIDS
Muons	Medium No isolation	Medium FCTight

Control regions for fake estimations

- Control regions enriched in real and fake leptons to estimate the rates

Real rate CR	Fake rate CR
2 opposite sign ee >=3j >=1*b*-jet m_{ll}>80 GeV	2 same sign ee,μμ or eμ ==3j ==1*b*-jet ee Z window veto m_{ll}<160 GeV (sub)leading lepton pT >(20)28 GeV
2 opposite sign μμ >=3j >=1*b*-jet m_{ll}>80 GeV	2 same sign ee,μμ or eμ ==3j >=2*b*-jet ee Z window veto m_{ll}<160 GeV (sub)leading lepton pT >(20)28 GeV

SR preselection definition (not optimized)

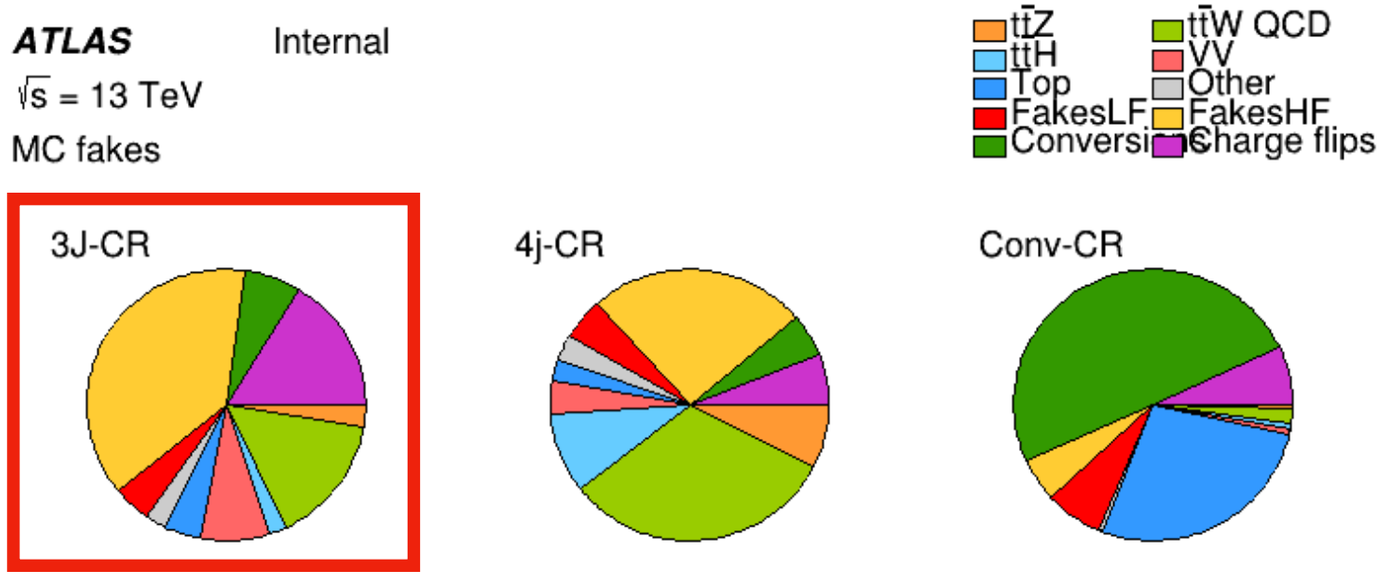
Signal region

Preselection
>=4j
radius_conv< 50 mm
$|\eta_{jj}| > 2$
$\sum j_{pT} > 250\text{GeV}$
$m_{ll} > 125\text{GeV}$

$|\eta_{jj}|$ is the absolute value of the difference in eta between the most forward non b-tagged jet and the jet that has the largest invariant mass (with the most forward jet).

Fakes composition in CRs and SRs

- Fake leptons origin:
  - Heavy-flavor jets (in yellow)
  - Light-flavor jets (in red)
- Conversions and charge-flips are not included in the fakes definition



Matrix Method: dilepton case

- MM relates the number of real loose and the number of fake loose leptons to the observed number of tight and loose leptons

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

- T: reconstructed Tight lepton
- L: reconstructed Loose lepton that fails the tight requirements
- l : reconstructed loose lepton
- R: real lepton
- F: fake lepton
- r: probability for a loose real lepton to be reconstructed as a tight lepton r<1 (real rate)
- f: probability for a loose fake lepton to be reconstructed as a tight lepton f<r (fake rate)
- 1 and 2 refer to lepton ordered by p_T

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

$$\begin{pmatrix} N^{ll}_{RR} \\ N^{ll}_{RF} \\ N^{ll}_{FR} \\ N^{ll}_{FF} \end{pmatrix} = \mathbf{M^{-1}} \begin{pmatrix} N_{TT} \\ N_{TL} \\ N_{LL} \\ N_{LL} \end{pmatrix}$$

$$N^{TT}_{fakes} = (N^{ll}_{RF} + N^{ll}_{FR} + N^{ll}_{FF}) f$$

Matrix Method code framework

- Local ROOT scripts used for both determining rates and applying them
- The script have been used in previous *ttV* analyses
- MM rates starts from preselected events with njets==3 and nbjets>=1.
- Charge-flip rate use OS and SS events with njets<=1 and bjets==0.

Procedure

- Definition of tight and loose leptons.
- Definition of control regions to estimate real and fake rates.
- Estimation of real and fake rates using the matrix equation.
- Application of rates to data using inverted matrix event by event.

Assumptions

- During the rate estimation assume that there are no events in which both leptons are fakes
  - From the matrix equation, this implies that the fourth inverted equation is zero

First term N_{RR} not included

$$N_{fakes}^{TT} = N_{RF}^{TT} + N_{FR}^{TT} + \overset{0}{N_{FF}^{TT}} = r_1 f_2 N_{RF}^{ll} + f_1 r_2 N_{FR}^{ll} + \overset{0}{f_1 f_2 N_{FF}^{ll}}$$

$$= \frac{r_1 f_2}{(r_1 - f_1)(r_2 - f_2)}[(f_1 - 1)(1 - r_2)N_{TT} + (1 - f_1)r_2 N_{TL} + f_1(1 - r_2)N_{LT} - f_1 r_2 N_{LL}]$$

$$+ \frac{f_1 r_2}{(r_1 - f_1)(r_2 - f_2)}[(r_1 - 1)(1 - f_2)N_{TT} + (1 - r_1)f_2 N_{TL} + r_1(1 - f_2)N_{LT} - r_1 f_2 N_{LL}]$$

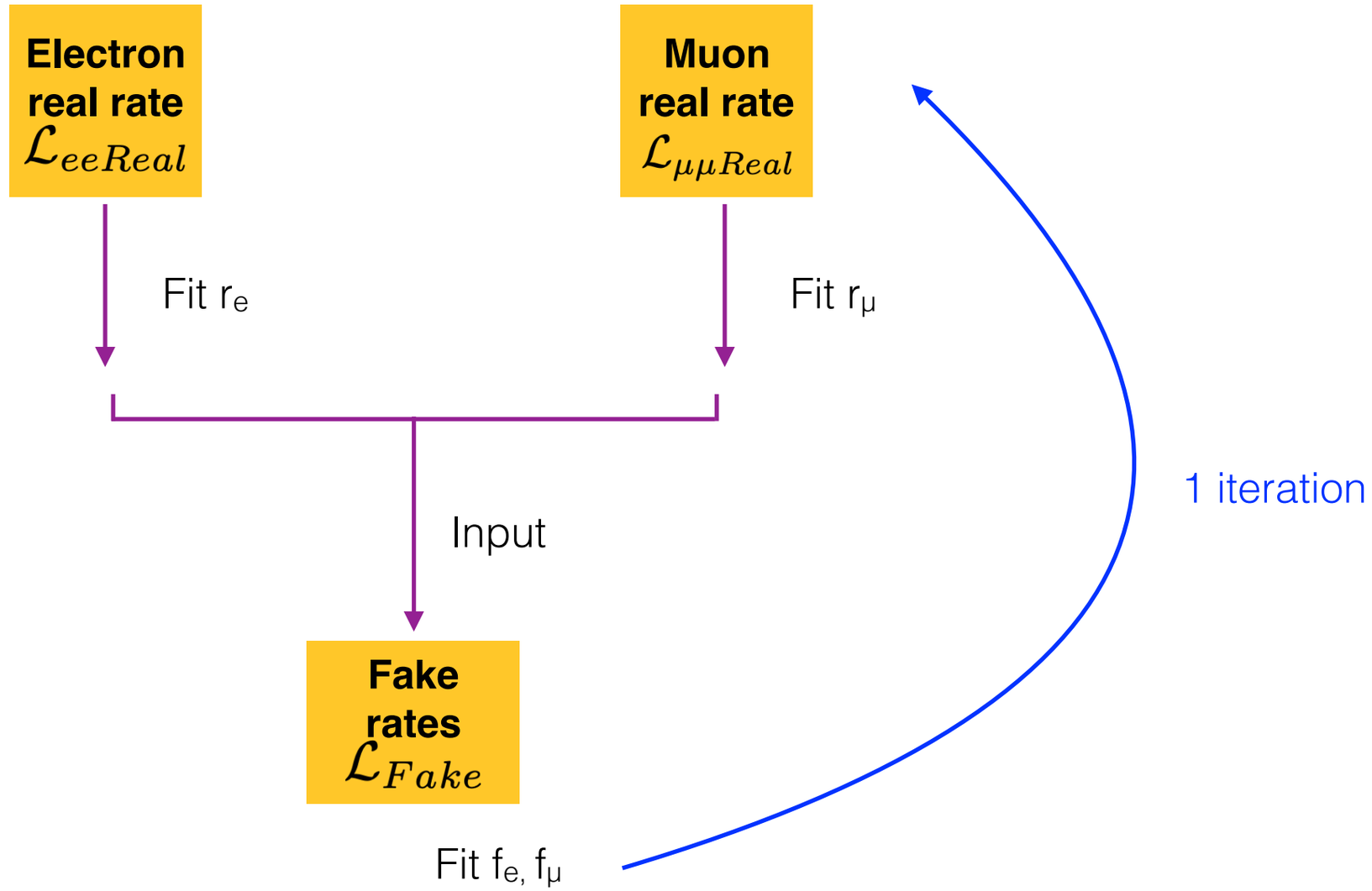
$$+ \frac{f_1 f_2}{(r_1 - f_1)(r_2 - f_2)}\underbrace{[(1 - r_1)(1 - r_2)N_{TT} + (r_1 - 1)r_2 N_{TL} + r_1(r_2 - 1)N_{LT} + r_1 r_2 N_{LL}]}_{0}$$

No explicit dependence on fake rates in the fourth equation with this assumption.

In the limit when r is large we get the T&P equations (T&P is an approximation of the full MM).

Rates estimates

- First iteration is done with a reasonable initial value for fake rates



- Fake rates are estimated separately for 1b and 2b
- The procedure is done twice, independently

Rates estimates: likelihood fits

$$-\ln \mathcal{L}_{eeReal} = -\sum_{xy}\sum_{ij} \ln \mathrm{Po}(N_{xy,ij}|(\mathbf{M}n)_{xy,ij}) \qquad \textbf{OS(ee)}$$

$$-\ln \mathcal{L}_{\mu\mu Real} = -\sum_{xy}\sum_{ij} \ln \mathrm{Po}(N_{xy,ij}|(\mathbf{M}n)_{xy,ij}) \qquad \textbf{OS}(\mu\mu)$$

$$-\ln \mathcal{L}_{Fakes} = -\sum_{\alpha}\sum_{xy}\sum_{ij} \ln \mathrm{Po}(N_{xy,ij,\alpha}|(\mathbf{M}n)_{xy,ij,\alpha} + N^{CF}_{yx,ij,\alpha} + N^{Conv}_{yx,ij,\alpha} + N^{Real}_{yx,ij,\alpha}) \qquad \textbf{SS(ll)}$$

- Po: Poisson term

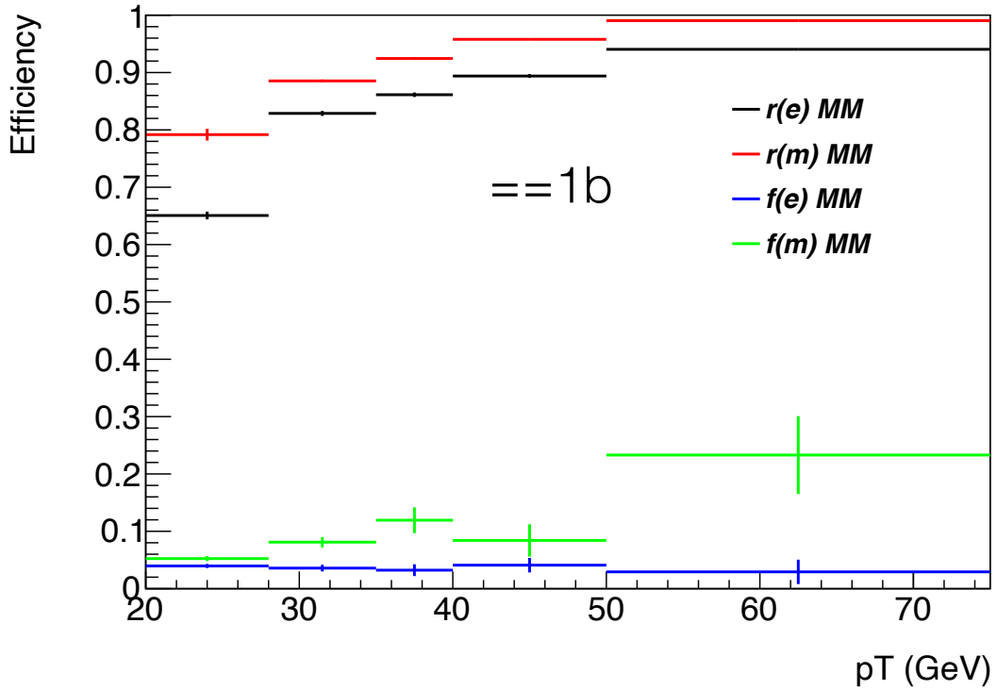
Applied in the corresponding CR:

- N: data events
- **M**: dilepton matrix
- CF: charge flips
- Conv: photon conversions
- Real: real background events
- n: $\begin{pmatrix} N^{ll}_{RR} \\ N^{ll}_{RF} \\ N^{ll}_{FR} \\ N^{ll}_{FF} \end{pmatrix}$

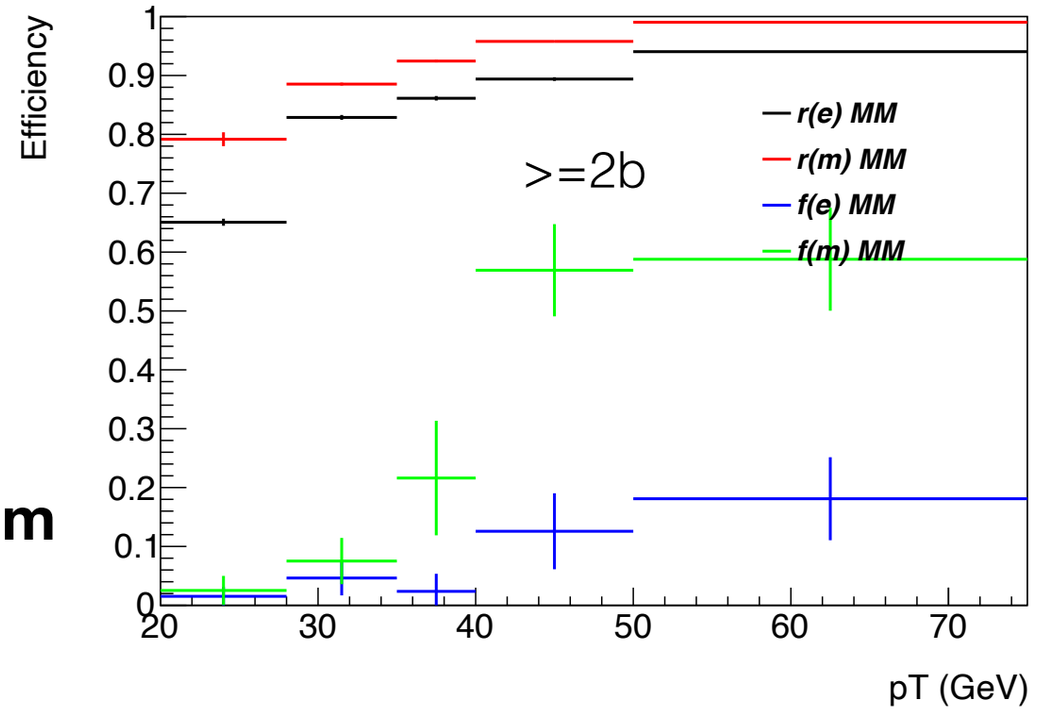
- xy: TT,TL,LT,LL
- ij: bins in leptons p_T ordered in p_T [20,28,35,40,50,inf)
- α: ee, eμ and μμ

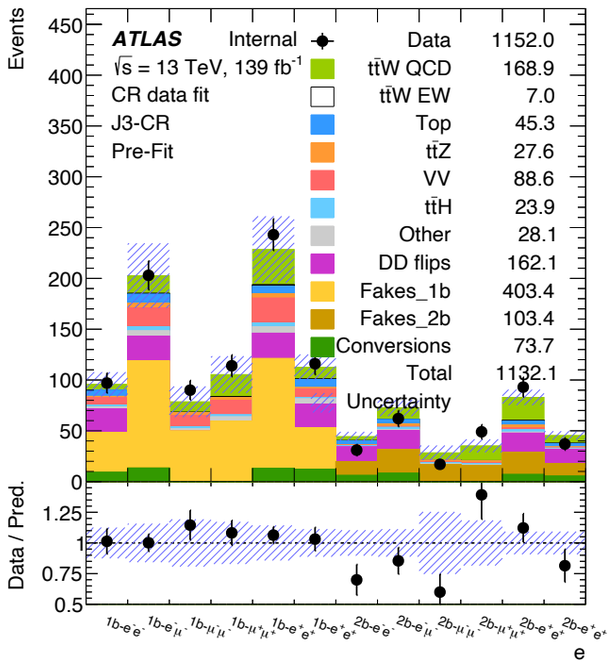
Fit results of rate estimation

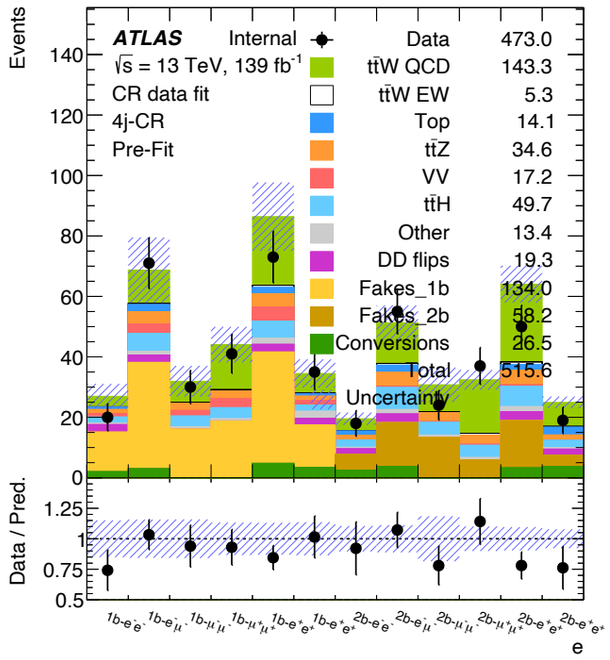
- Real rates fit stable
- Fake rates increase at high pT in the 2b region

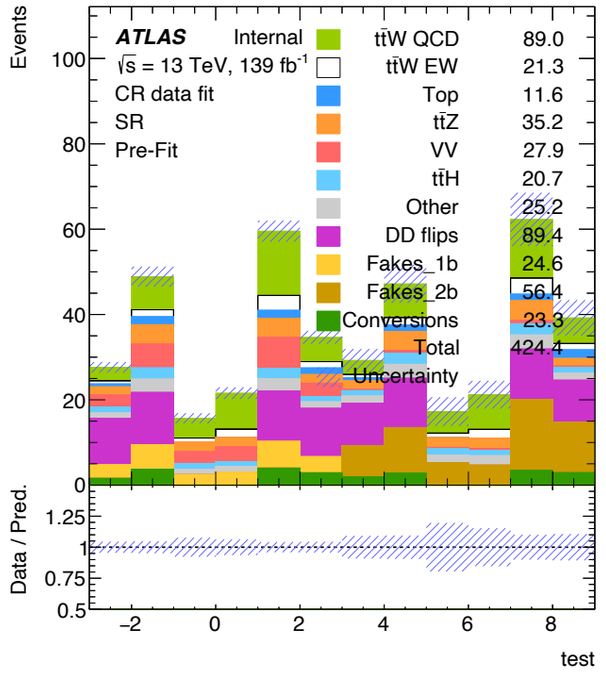


m





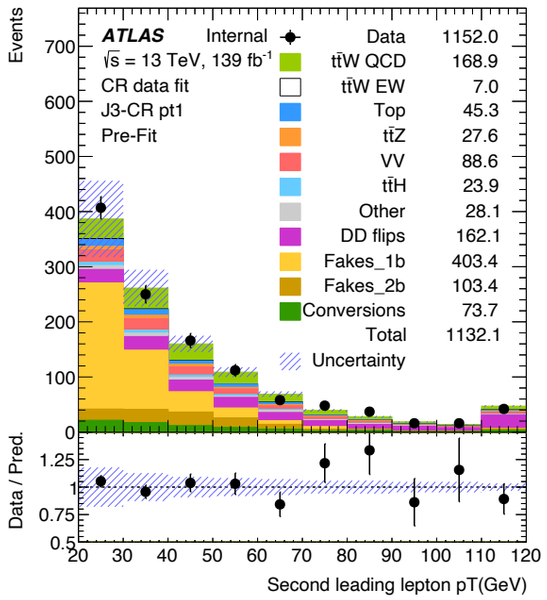


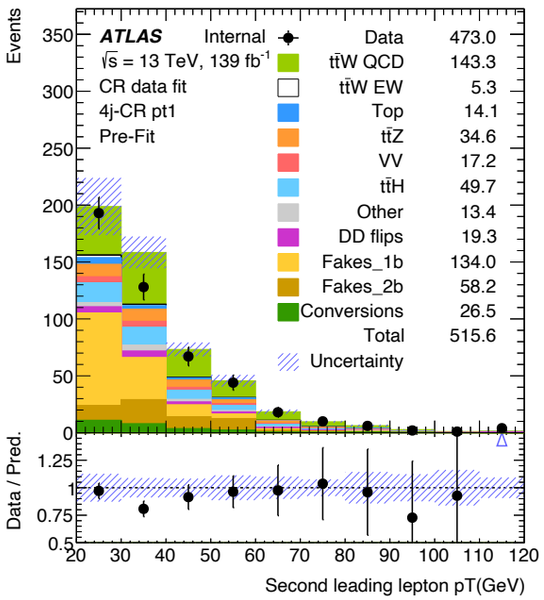


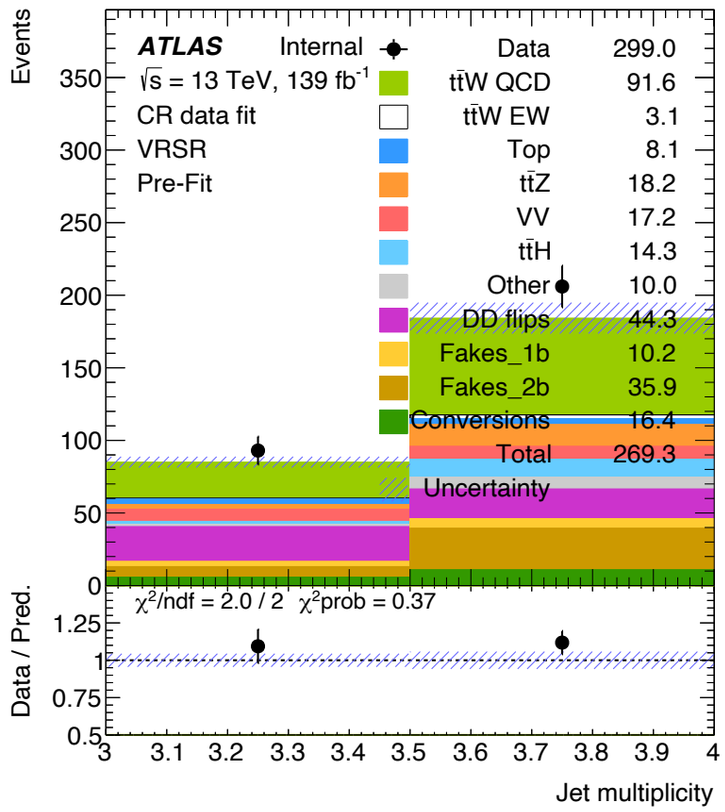
The uncertainty bands include statistics uncertainties and an overall normalisation uncertainty for the main backgrounds:

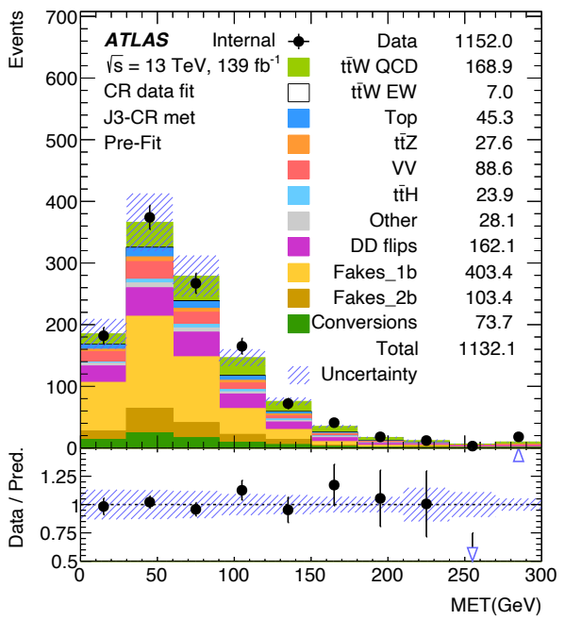
- ttZ (10%)
- ttH (20%)
- Fakes 1b (30%)
- Fakes 2b (30%)
- Charge Flips (20%)

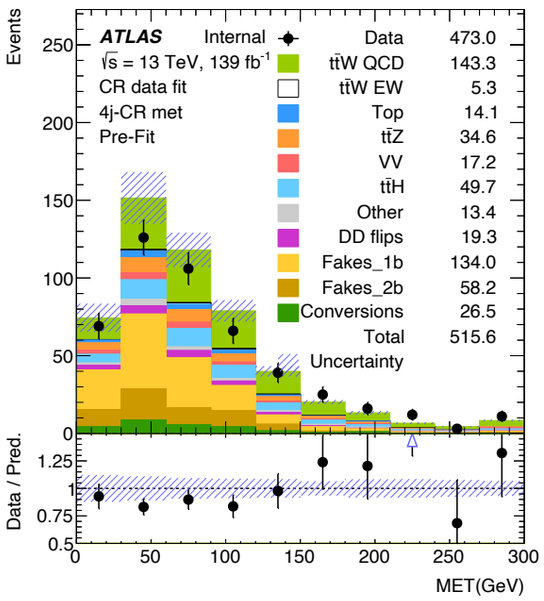
Example distributions after the rate fit





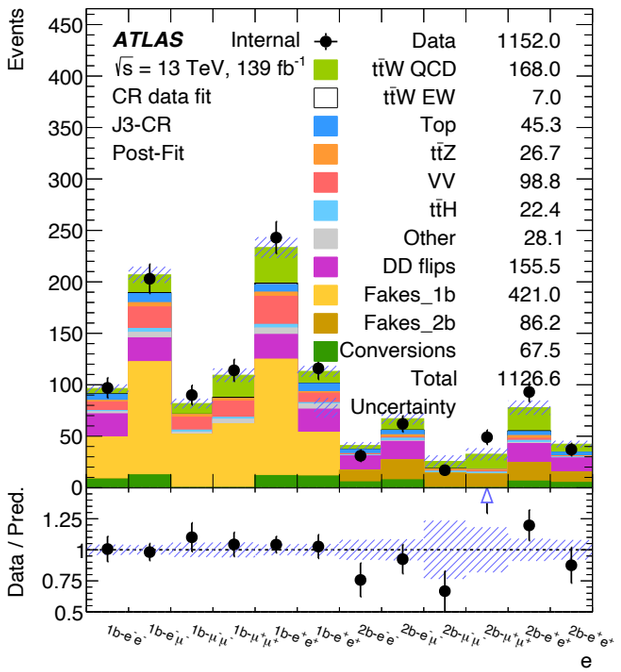


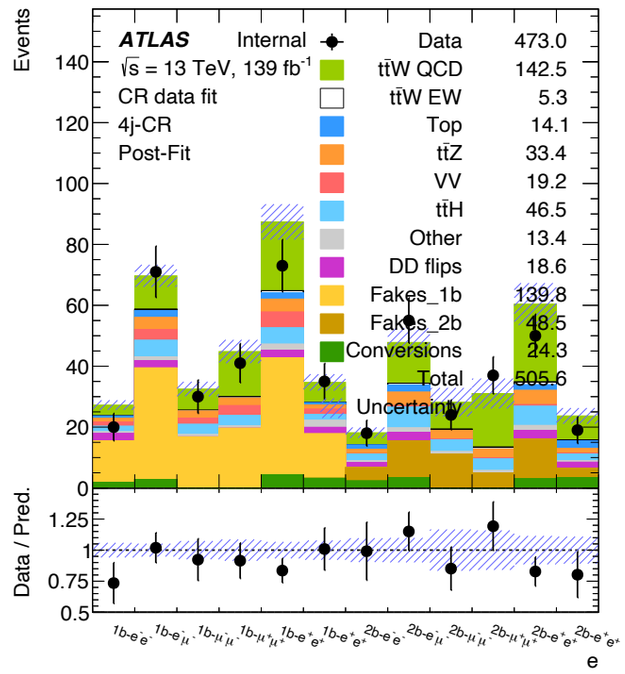




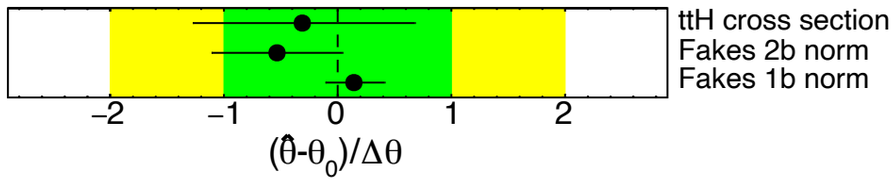
SR like selection in CR:
- mll>125 GeV
- Lepton pT > 40 GeV
- HT>450 GeV

Fit results of all CR only regions

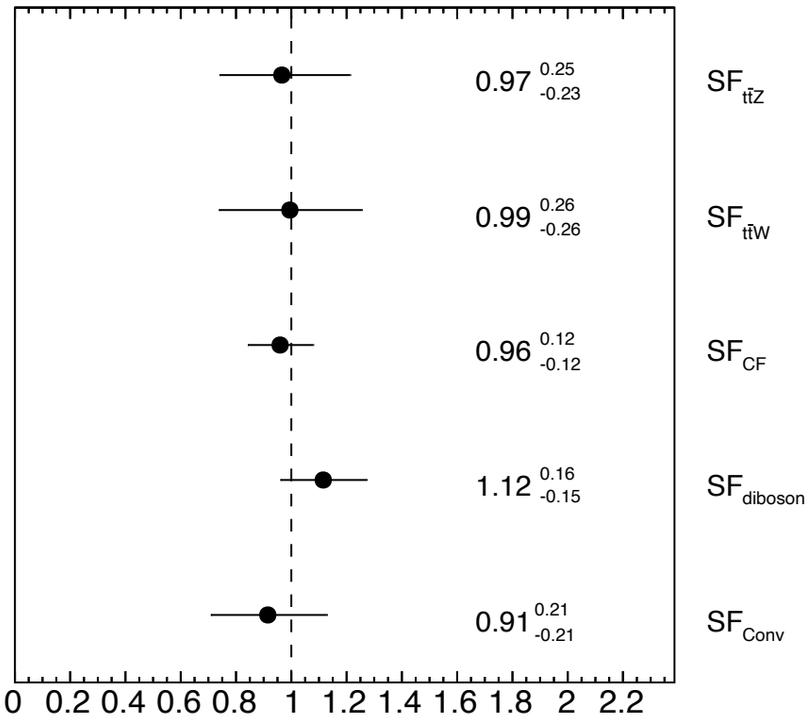




Nuisance parameters



Free floating parameters

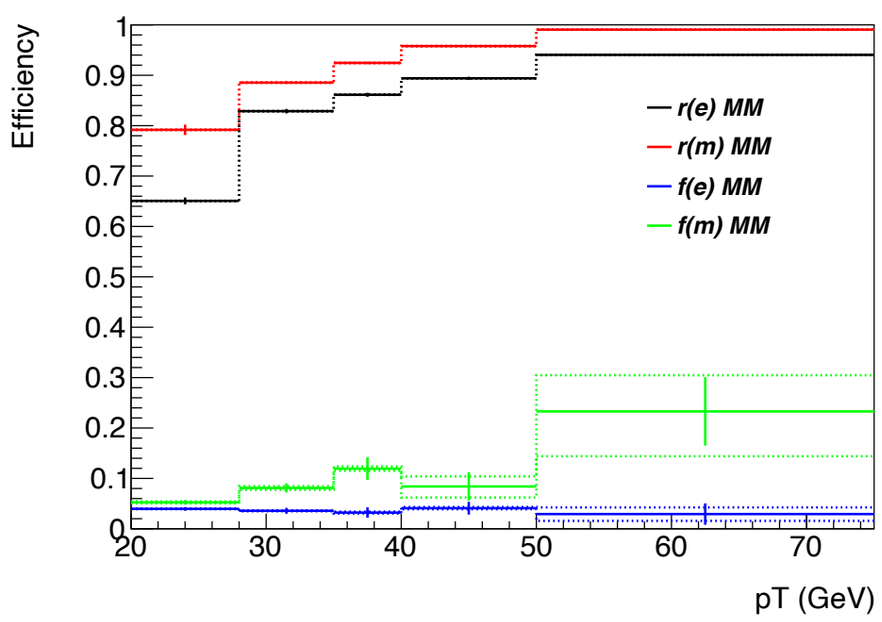


Estimation of uncertainties

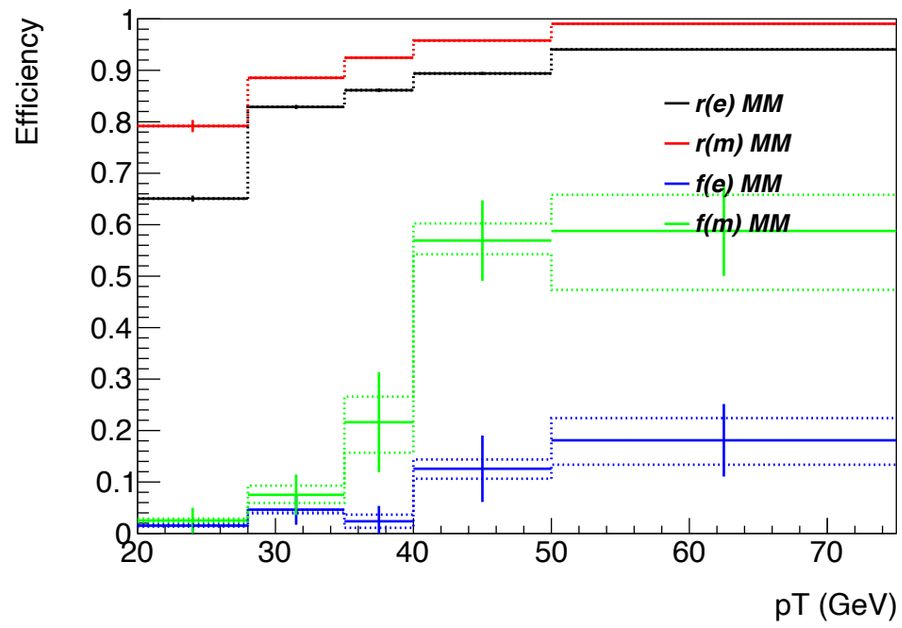
- The contribution of the different backgrounds (CF, Conv, ttW, other real) are shifted independently and rates are refitted, bin-by-bin shifts (shape) are derived using the bin stat. unc. from Minos.

Example: rate shifts due to 50% shift in ttW QCD, this is used in the final fit
to derive #Fakes(norm_ttW_QCD) while norm_ttW_QCD floating is in the fit.
50% is arbitrary (next slide) and only used for the numerical derivative in the Taylor expansion.

==1b



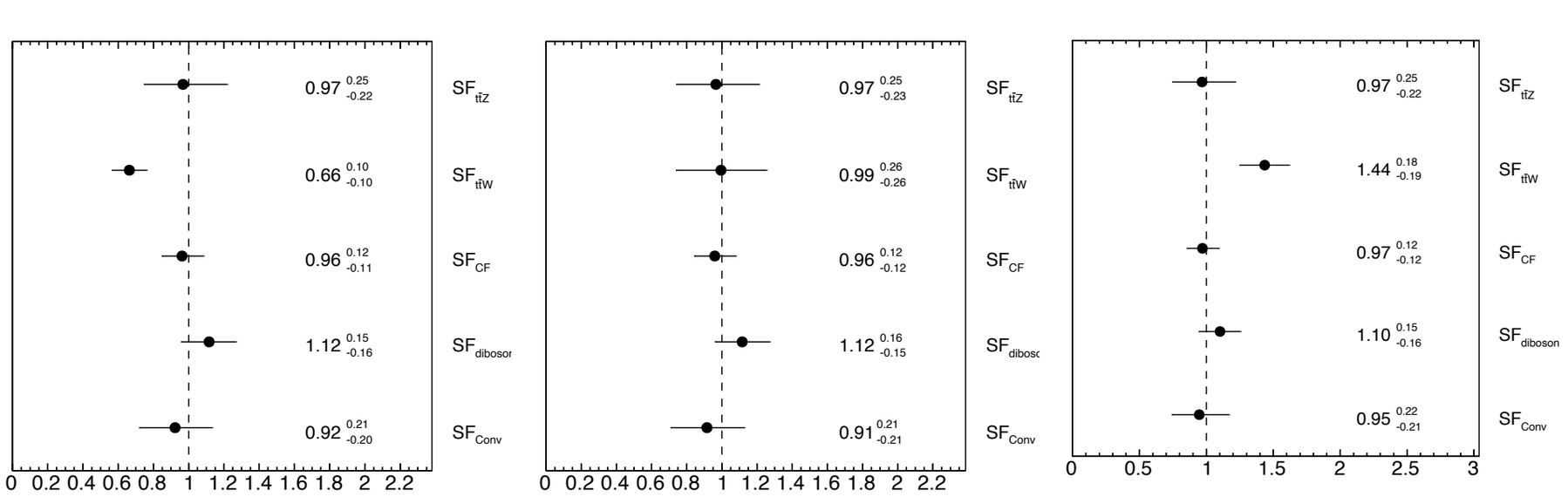
>=2b



Check fake rate independence of ttW

- Large fraction of ttW in the fake rate CR.
- Remove fake rate dependence on ttW by constructing #fake(norm_ttW) using Taylor expansion: f(nom+h) = f(nom) + h * f'(x), f'(x)~(f_up-f_nom)/h_shift. Then use that in the fit.
- Redo full analysis including rate fit for assumed σ(ttW)=+50, 0, -25%.

+50% **0%** **-25%**

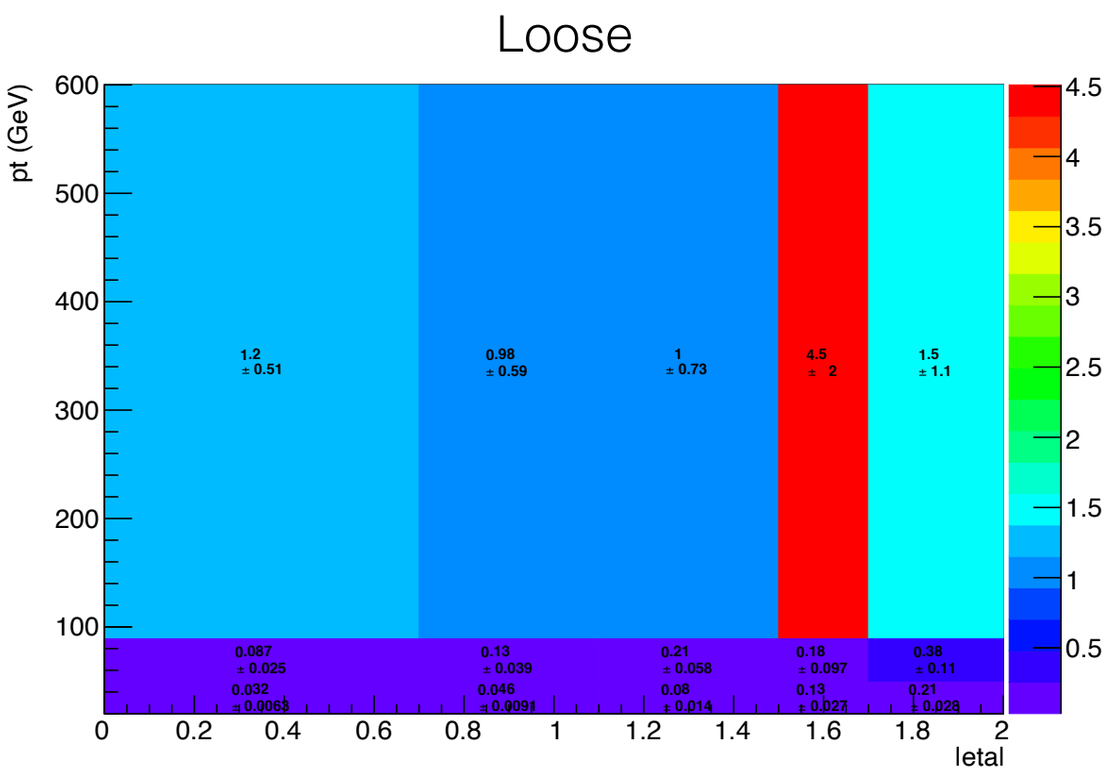


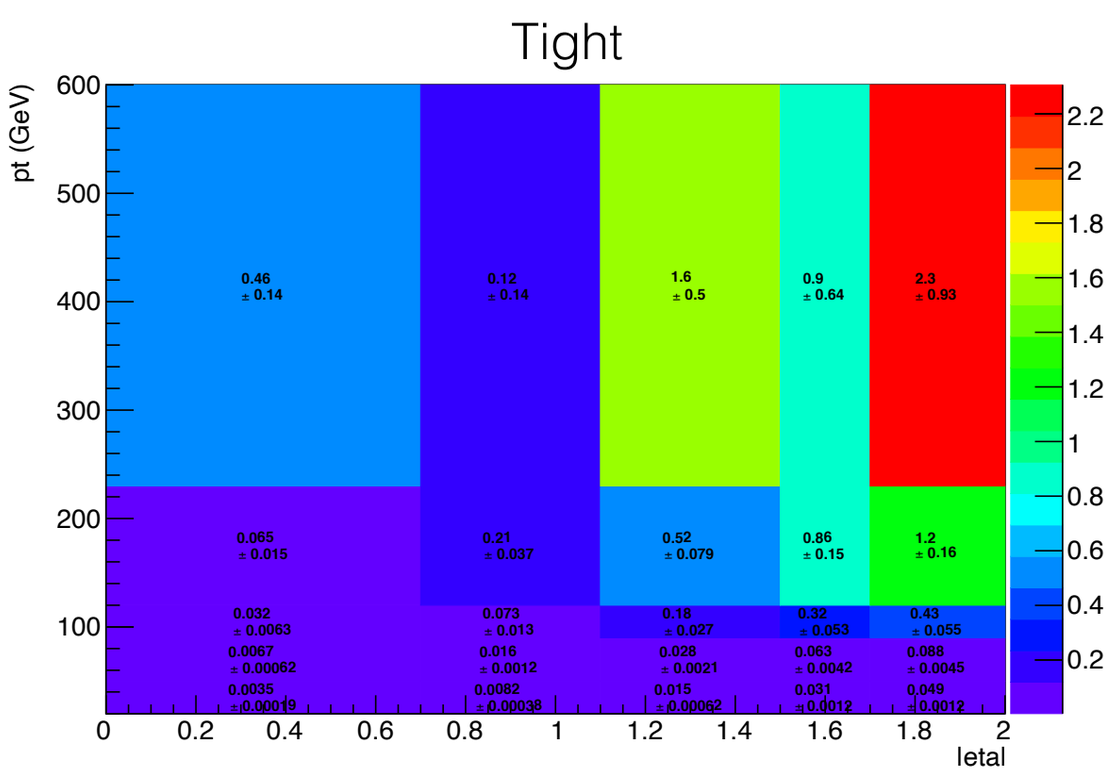
Assumed ttW	$\sigma(ttW)^{*}_{fit}$ / $\sigma(ttW)^{SM}$
50 %	0.99
0 %	0.99
–25 %	1.08

Data driven Charge Flips estimation

- Z peak same-sign ee events + side-bands
  - Side-bands used to interpolate the background contribution
  - Subtract interpolated background in Z window ->Gives charge flips
- This observed number is then used in a likelihood to estimate the rate of charge flips

$$-\ln \mathcal{L}_{CF} = -\sum \ln \mathrm{Po}(N_{ss}|(\epsilon_1+\epsilon_2)N) \qquad N = \frac{N_{OS}}{1-\epsilon_1-\epsilon_2}$$





Conclusions

- Constraining the QCD production of ttW is challenging due to the relevant fake lepton CR contains significant fraction of ttW.
- A critical aspect of the analysis is to assert independence of the assumed ttW normalization.
- Very recently managed to convince ourself that the fake lepton estimation using the matrix method can be made independent of the ttW(QCD) normalization by a Taylor expansion approximation of #fakes(ttW), validated for different assumptions.
- This has been the main challenge for quite some time, now need to include the remaining fake lepton uncertainties in the fit.
- Optimization of the signal region not final yet.
- Preliminary checks in the CR with selections that have signal like source compositions look promising.