Hannah Jacobi, Arno Straessner, Tom Kreße, Manuel Gutsche, Christian Schmidt
Institut für Kern- und Teilchenphysik, TU Dresden

# Search for a light CP-odd Higgs boson with the ATLAS detector – data-driven modelling of background processes

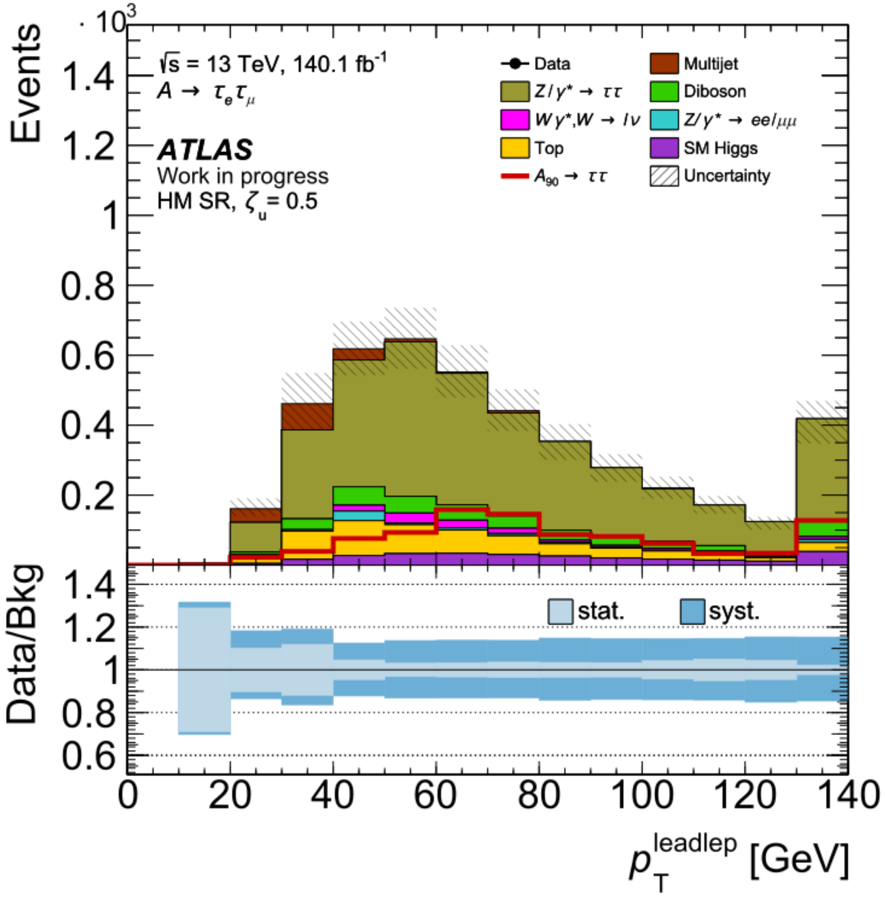
Session T 82: Higgs, Di-Higgs II
DPG Spring Meeting Dresden, 22nd of March, 2023

# Introduction

- **Follow up talk on the analysis that Tom Kreße introduced**
- **Search for light CP-odd Higgs boson A in the decay channel: A→ τ+τ→ e+μ+4ν**

**Main background in the signal region: Z→ ττ**

**Background processes modelled using Monte Carlo simulation, except "Multijet" → data-driven Fake Factor method**



# Fake background estimation

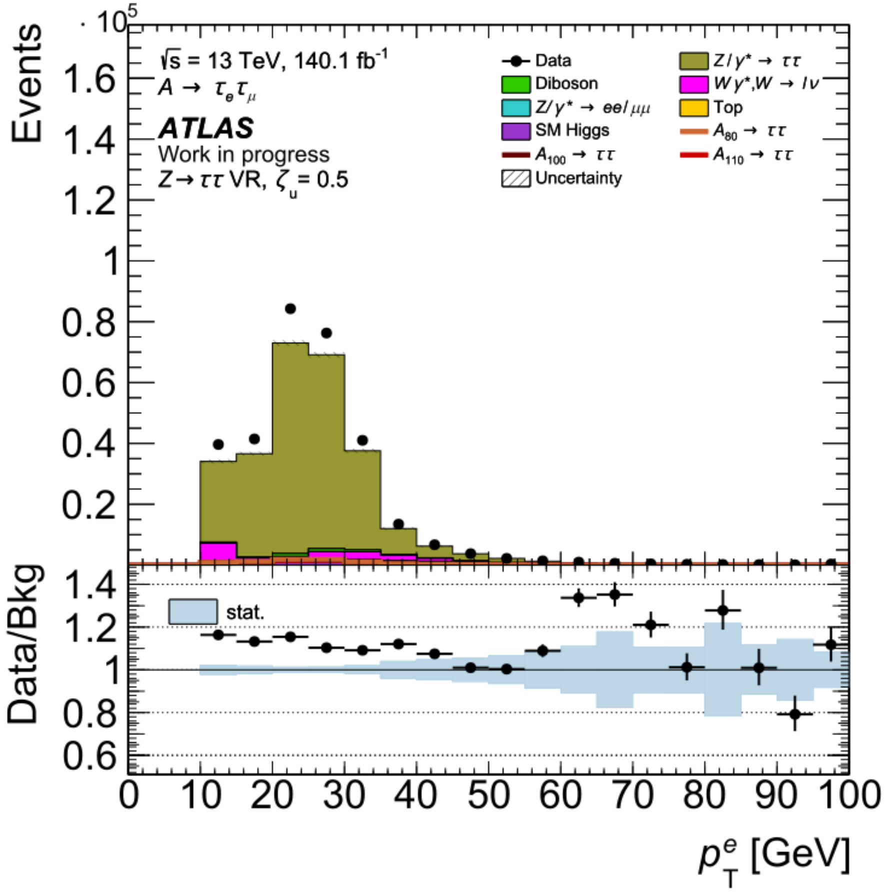
# Fake events

- **Fake events = Events from misidentifying jets as leptons**
  **OR actual leptons from the decay of non-prompt particles**

- **Wide variety in sources with very large cross-sections:**
  **semi-leptonic heavy-flavour decay, photon conversion, charged hadrons, meson decay**

- **Especially prominent in low $p_T$ regions**

- **Hard to simulate in Monte Carlo**
  **→ use data-driven approach to estimate this background**



# Fake Factor Method

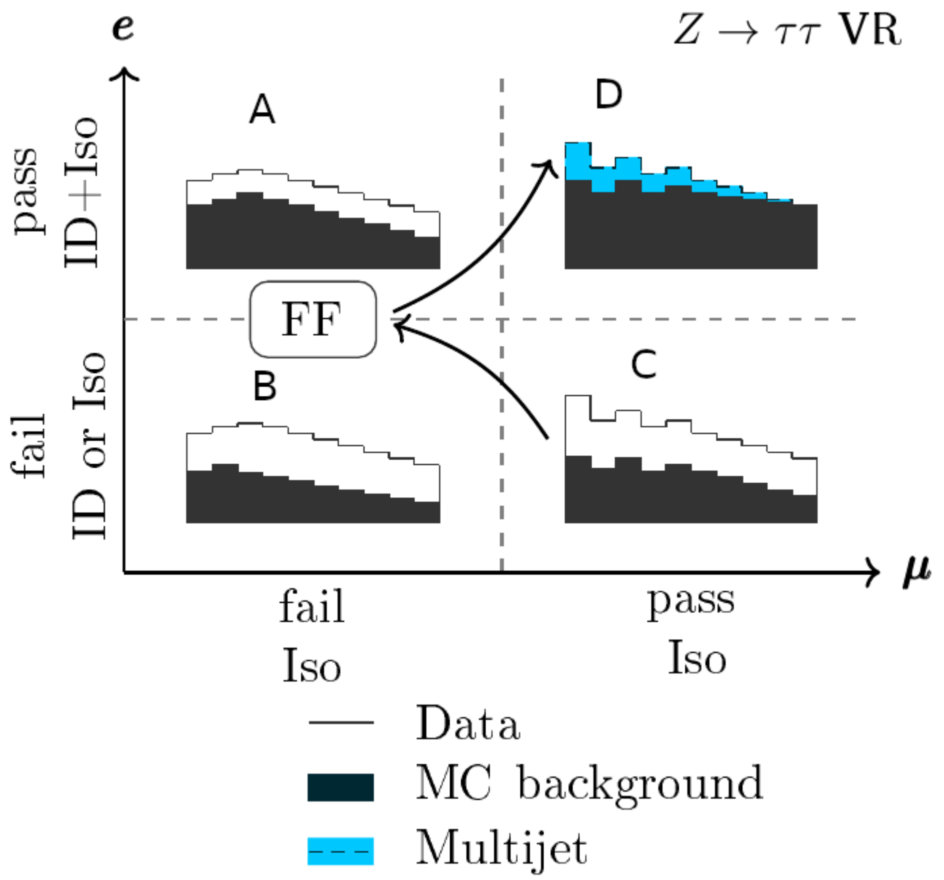
- **Used for estimating double Fake events**
- **Fake Factors "FF":**

$$FF = \frac{N_A^{fake}}{N_B^{fake}} = \frac{N_A^{data} - N_A^{MC}}{N_B^{fake} - N_B^{MC}}$$

- **Number of fake events:**

$$N_D^{fake} = FF \cdot N_C^{fake} = FF \cdot \left( N_C^{data} - N_C^{MC} \right)$$

- **Fake Factors are calculated in Z→ττ validation region**



# Fake Factor Method

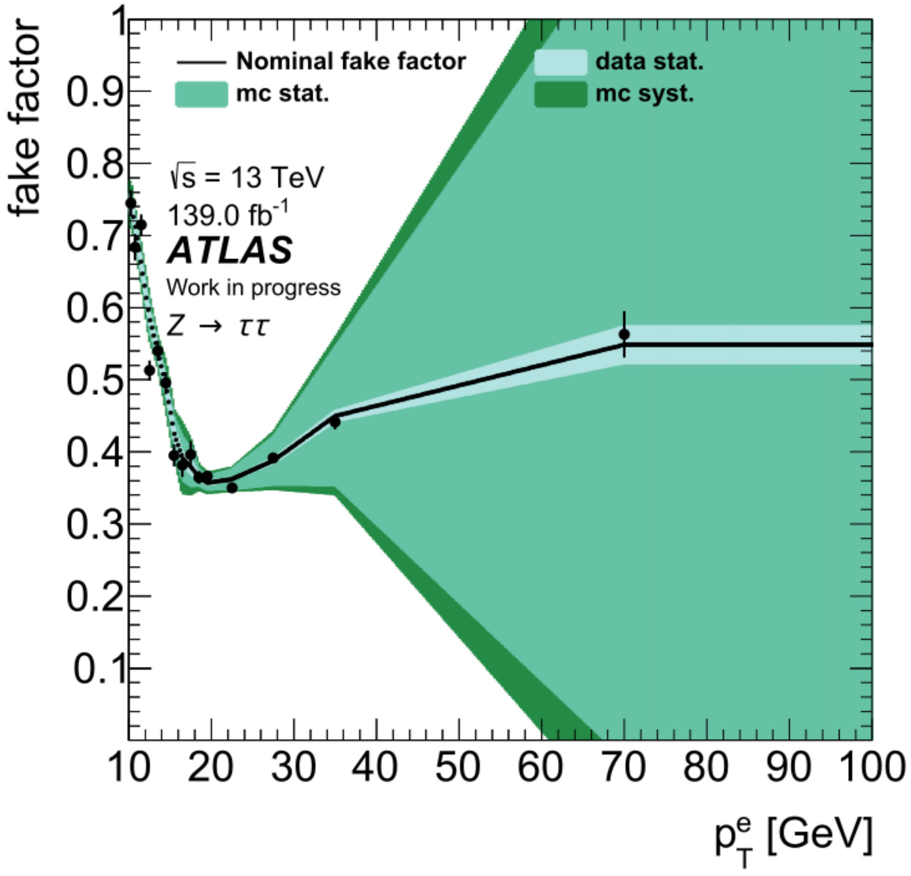
- **Used for estimating double Fake events**
- **Fake Factors "FF":**

$$FF = \frac{N_A^{fake}}{N_B^{fake}} = \frac{N_A^{data} - N_A^{MC}}{N_B^{fake} - N_B^{MC}}$$

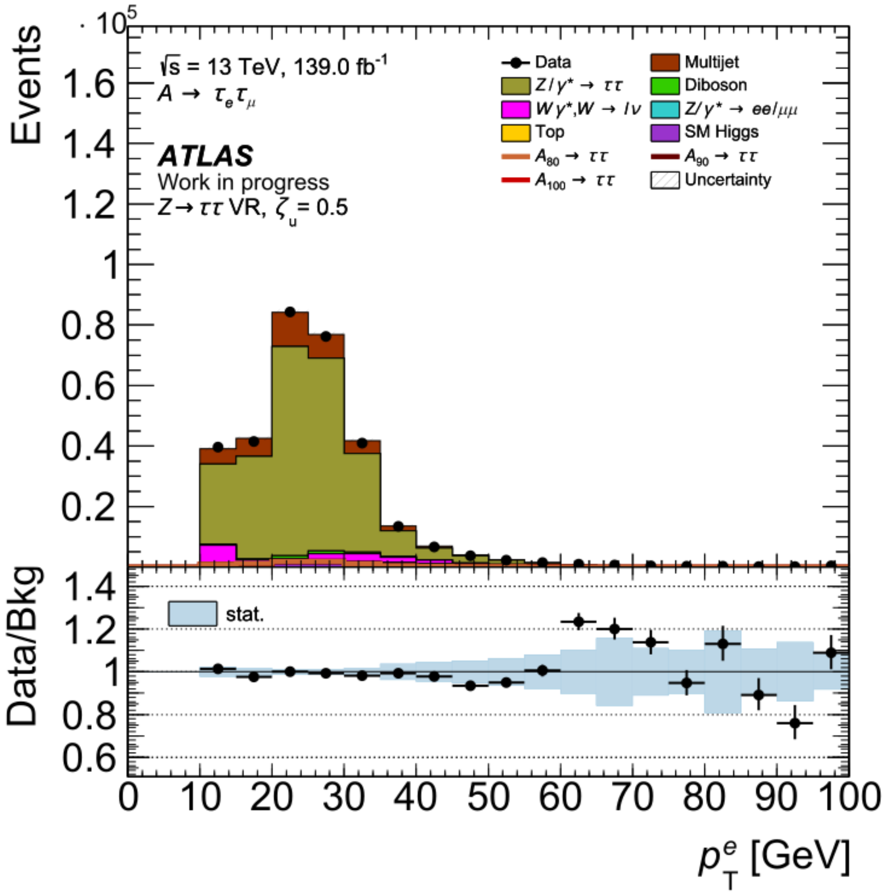
- **Number of fake events:**

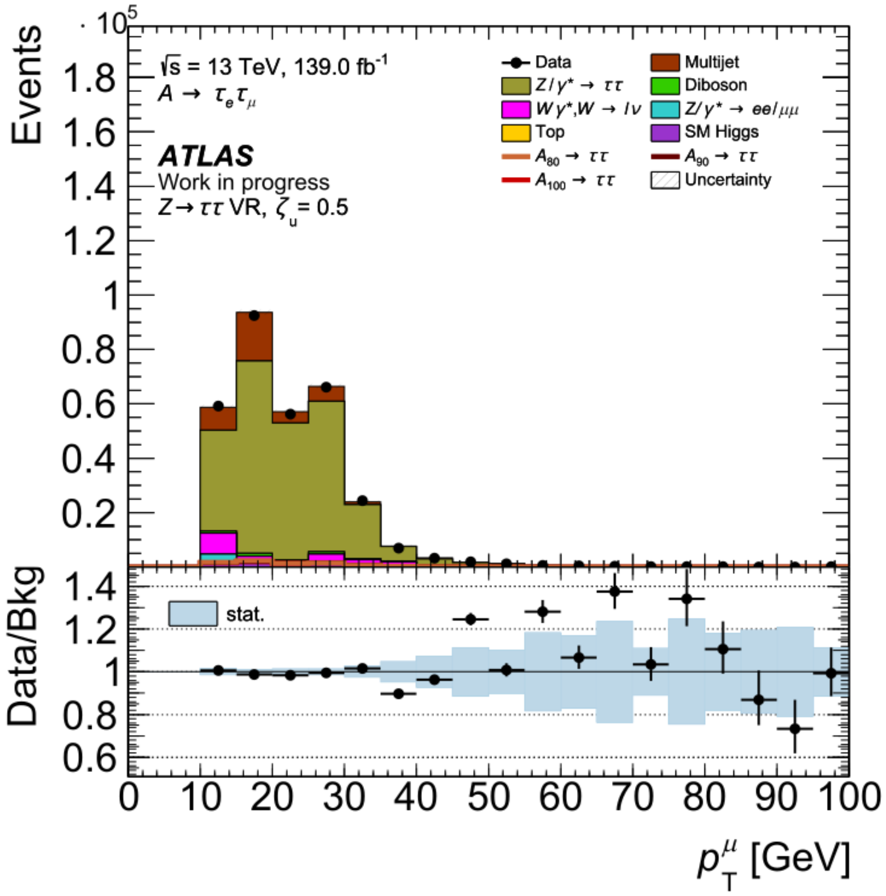
$$N_D^{fake} = FF \cdot N_C^{fake} = FF \cdot \left(N_C^{data} - N_C^{MC}\right)$$

- **Systematics based on: statistical and experimental systematic uncertainties of A, B and C region**



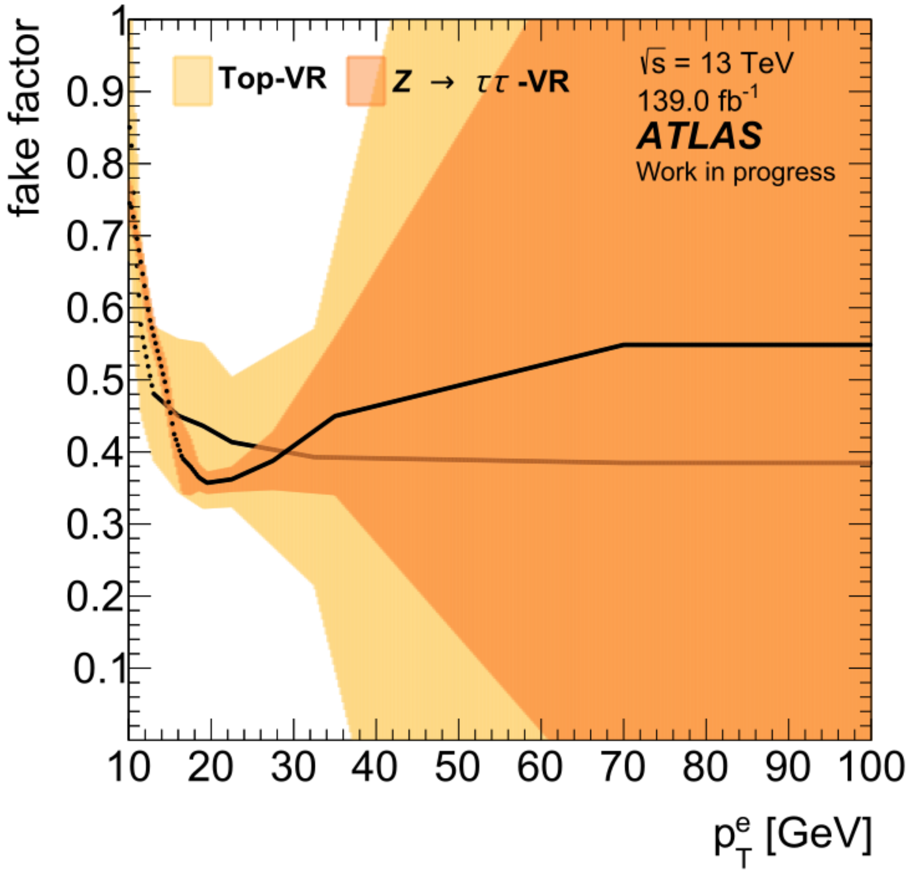
# Z→ττ validation region with multijet background



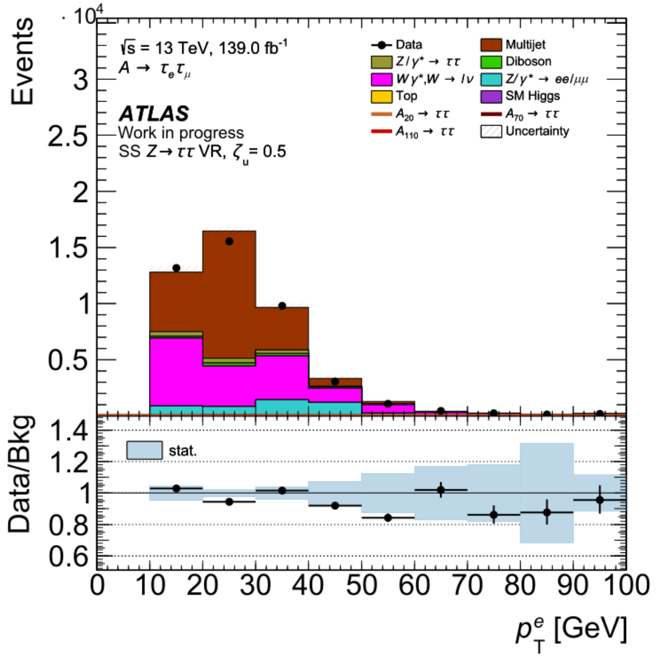



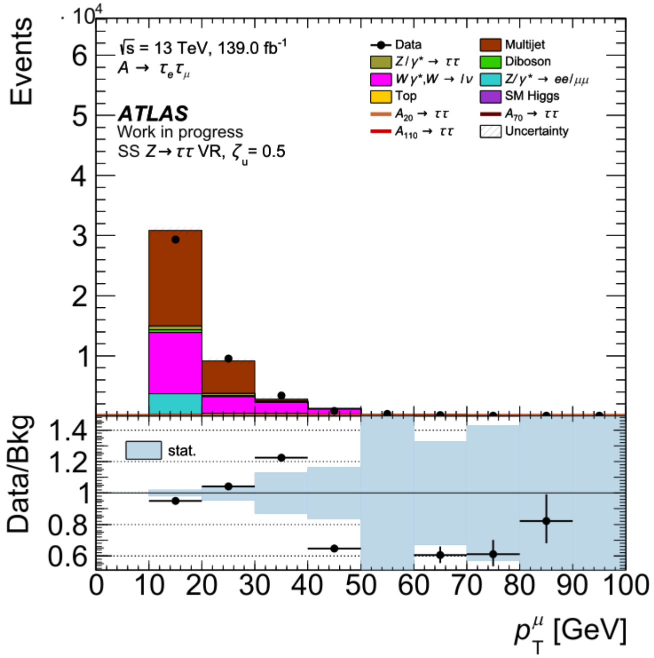
# Validation of Fake Factors

- **FF calculated in Z→ττ and top validation region are statistically compatible**
- **Use same-sign regions for further validation of the multijet background apply FF from Z→ττ region here**



# Same-sign Z→ττ validation region



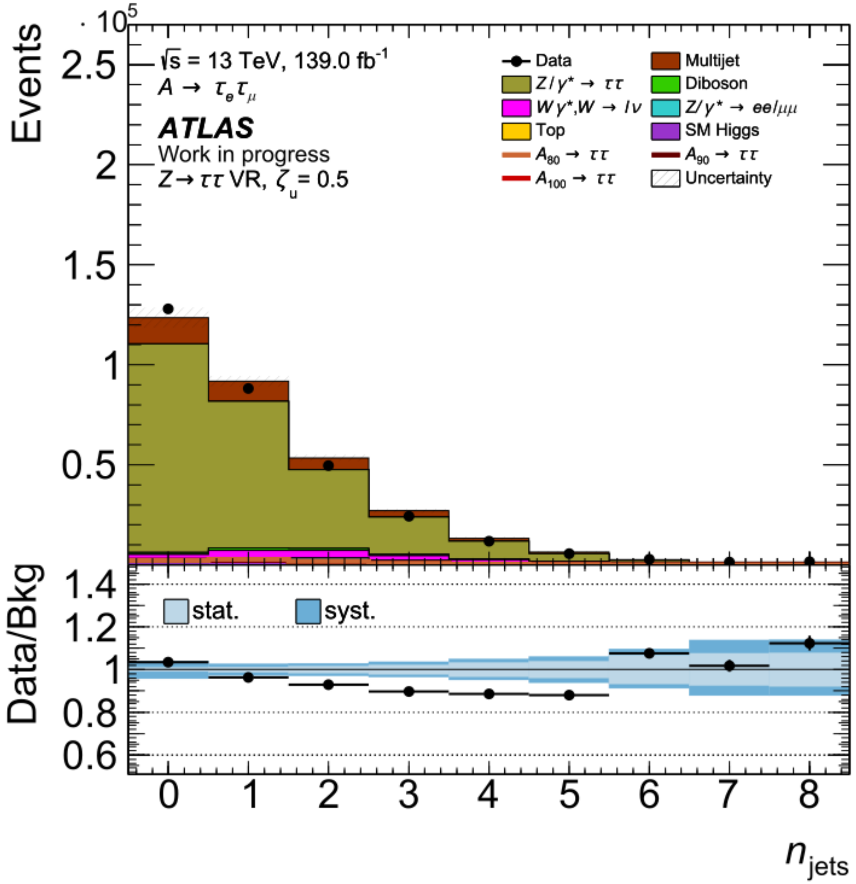




- **Further validation:**
  **Currently implementing alternative Matrix Method**
  **→ estimating single and double fake events using a data-driven method**

# Reweighting of the $Z \rightarrow \tau\tau$ background

# Mismodelling in the Z→ττ validation region

- **Observed: mismodelling in Z→ττ validation region**
  **especially in $n_{jets}$ distribution,**
  **also in missing $E_T$**
- **Important for signal region, as background here dominated by Z→ττ**
- **Unsuccessfully looked into:**
  - Triggers
  - Data-taking years
  - New Sherpa 2.2.11 samples
  - Different lepton IDs

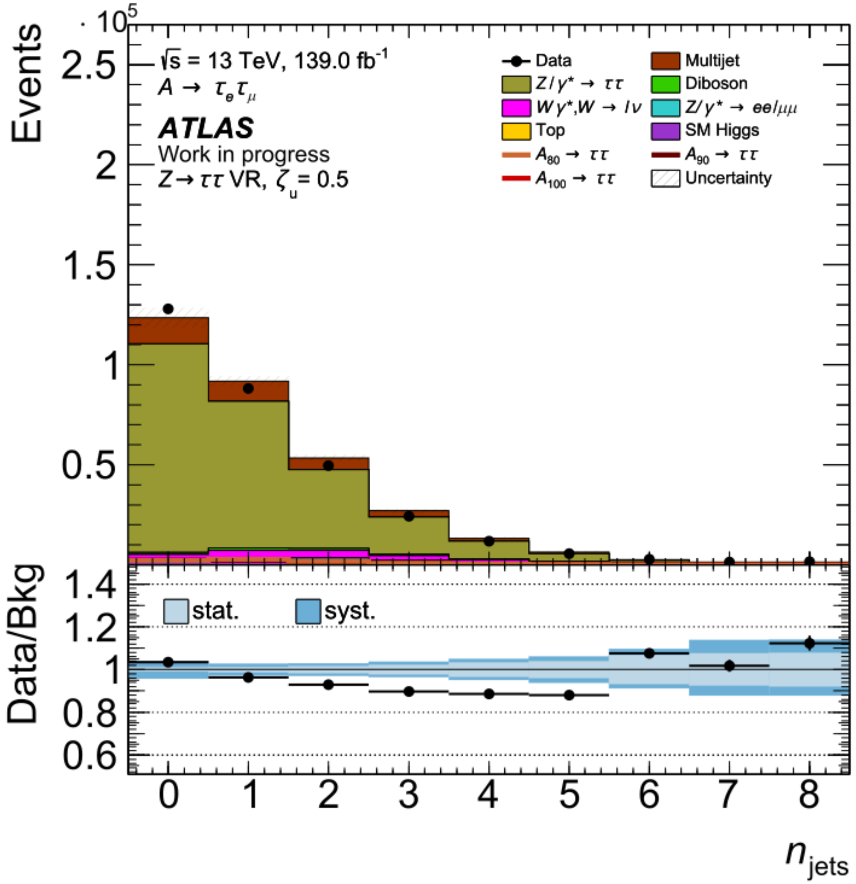


# Data-driven reweighting approach

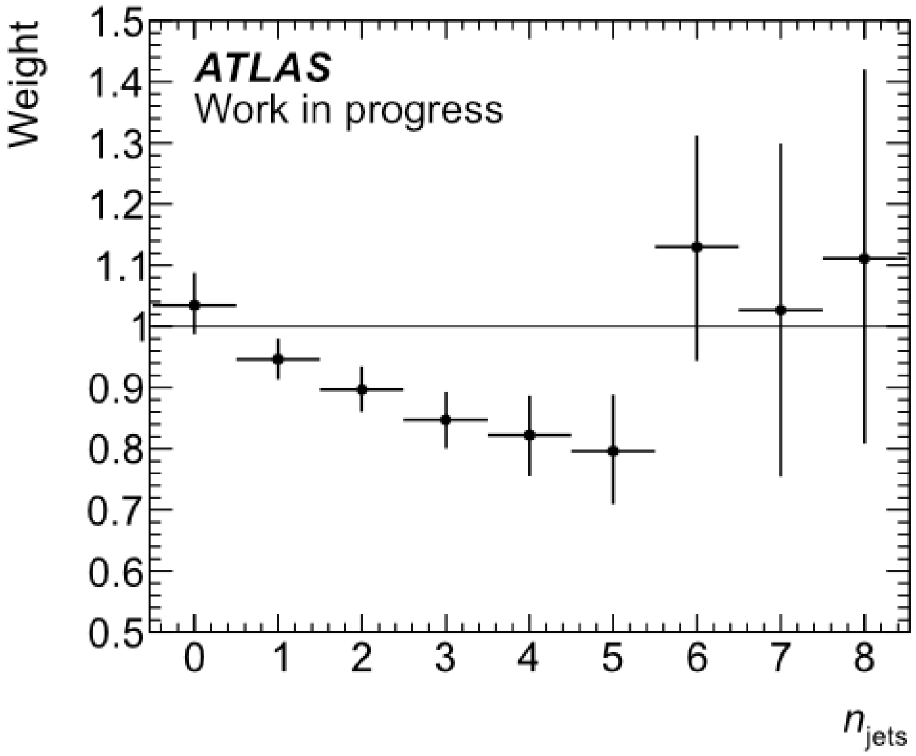
- **Problem pinned down to Z→ττ bkg**
  **→ correct Z→ττ using the njets distribution**

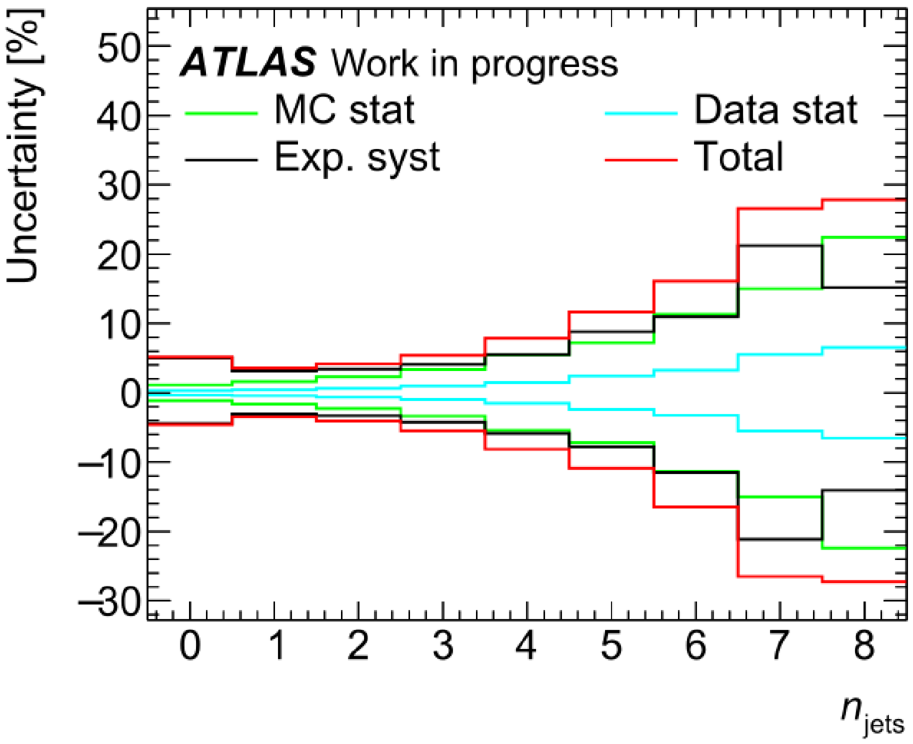
- **Data-driven approach:**
  **event weights for Z→ττ calculated in Z→ττ validation region:**

$$weight = \frac{Data - \sum_{i \neq Z\tau\tau} Bkg_i}{Bkg_{Z\tau\tau}}$$

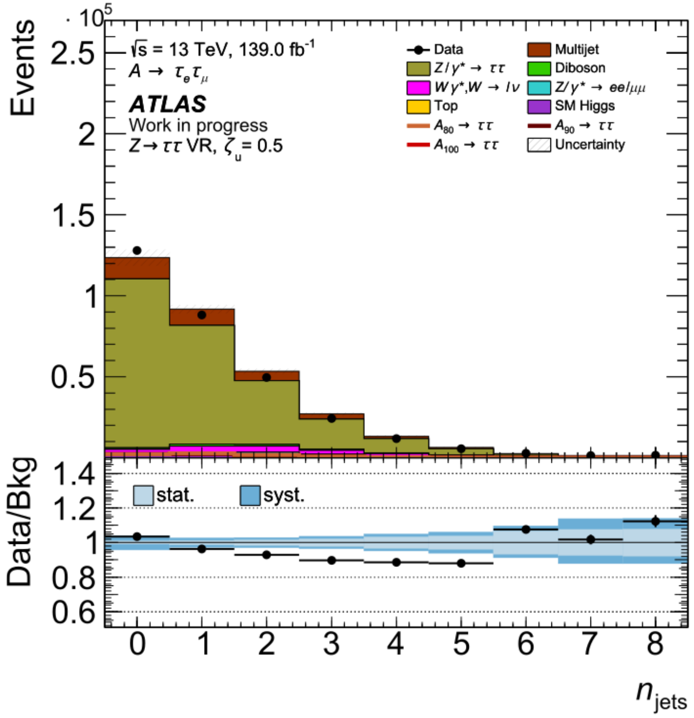


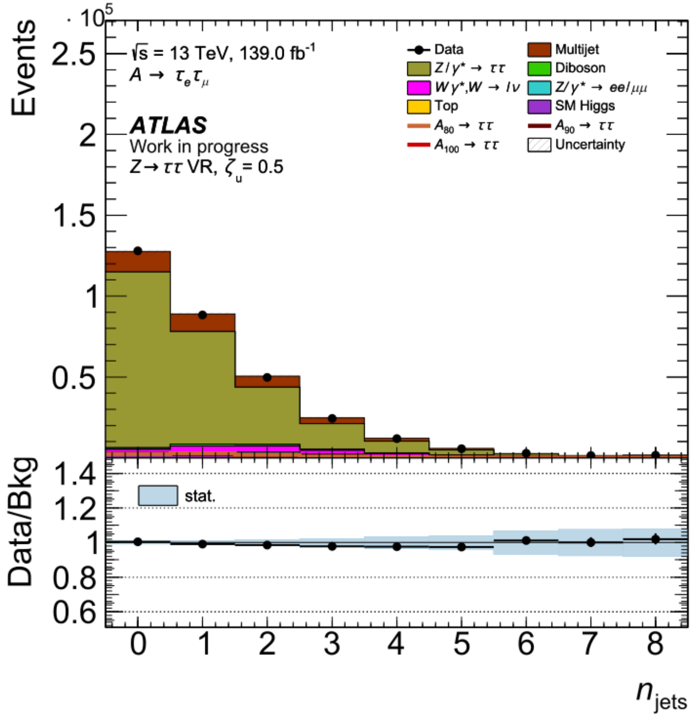
# Data-driven reweighting approach



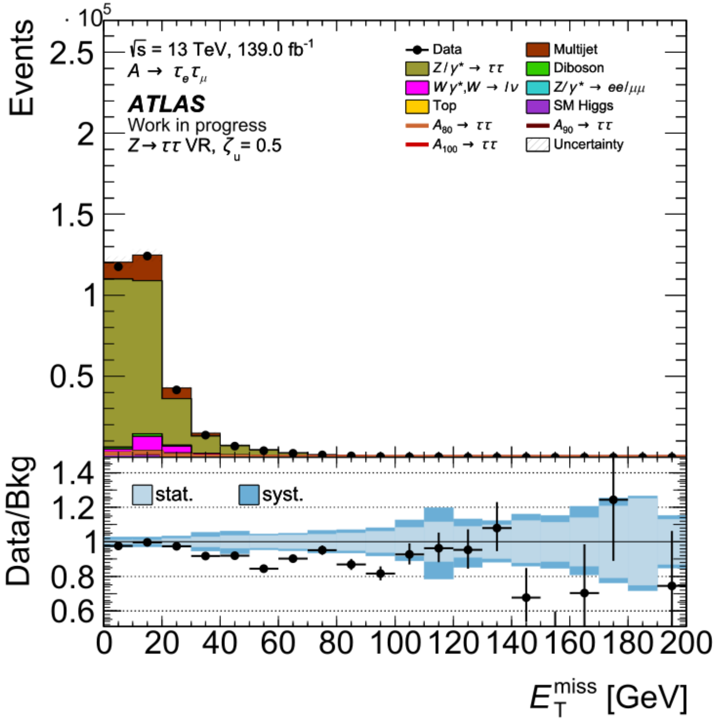
# Results of reweighting in Z→ττ validation region

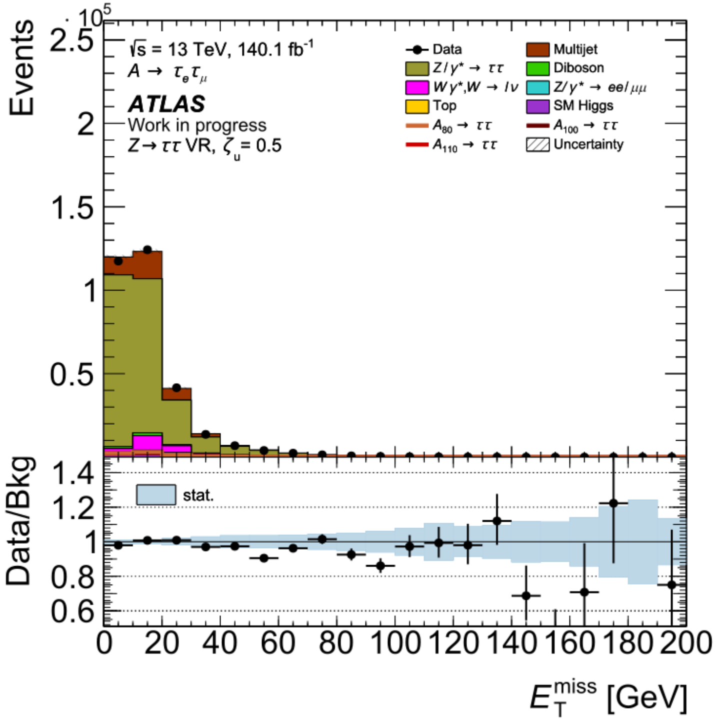



- **Have to reiterate calculation of Fake Factors after reweighting procedure**

# Results of reweighting in Z→ττ validation region




- **$E_{T,miss}$ important selection variable used for estimating mass of A boson**

# Summary and outlook

- **Presented two modelling procedures in the search for a light CP-odd Higgs boson A  in the $A \to \tau\,\tau \to e\,\mu$ channel**

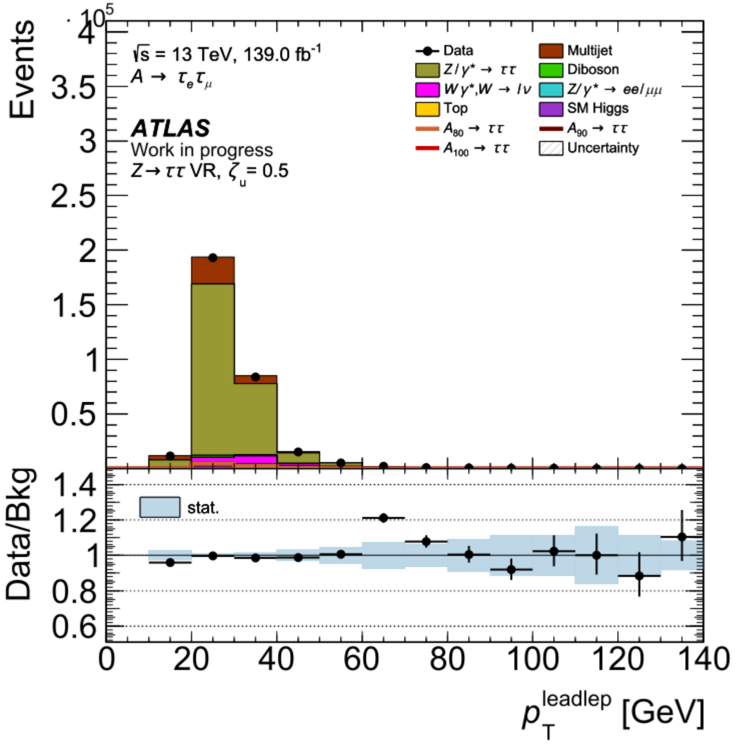
- **Background modelling with Fake Factor Method successful**
- **Reweighting of $n_{jets}$ distribution for $Z \to \tau\tau$ samples necessary**

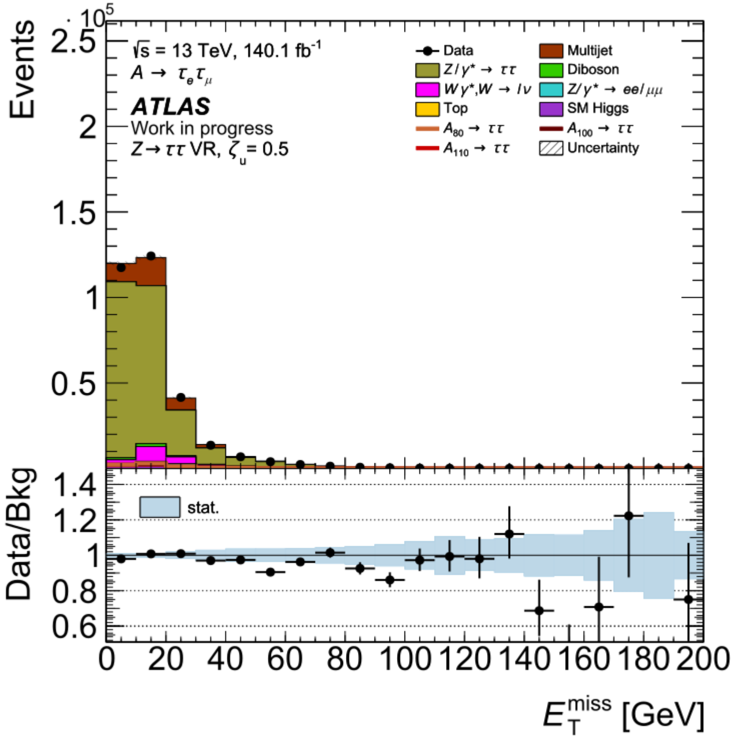
- **Alternative Matrix Method for fake background estimation currently in implementation**

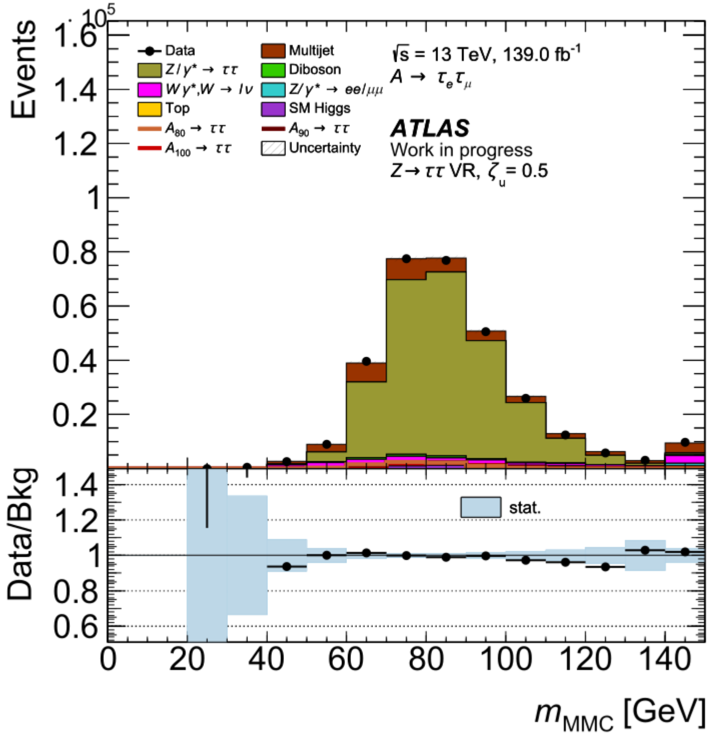
# Thank you for your attention!

# Backup

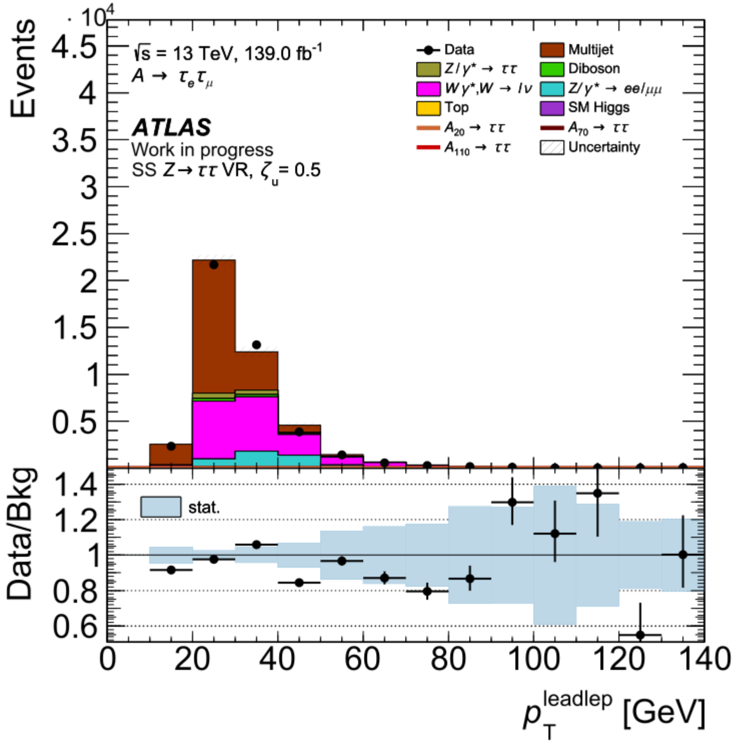
# More distributions in the Z→ττ validation region

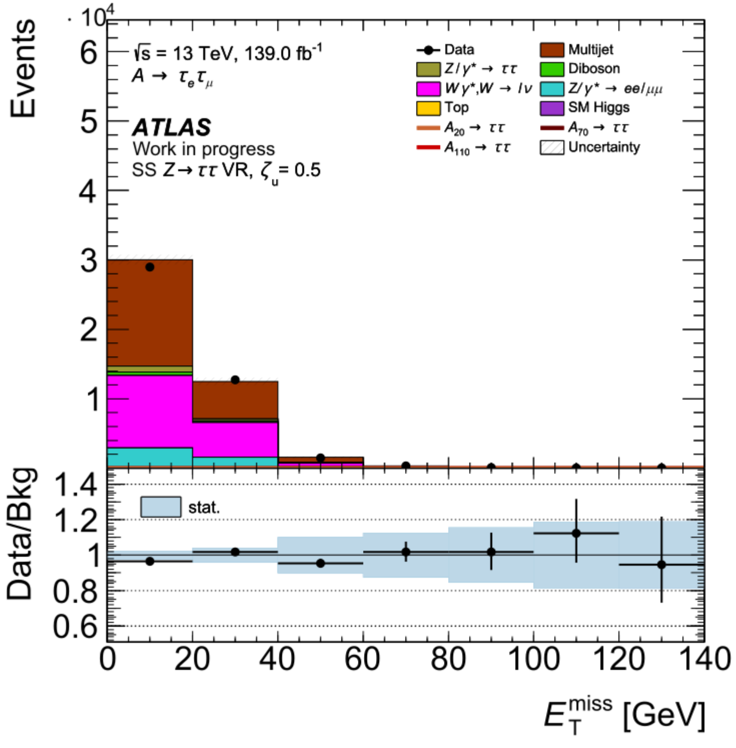


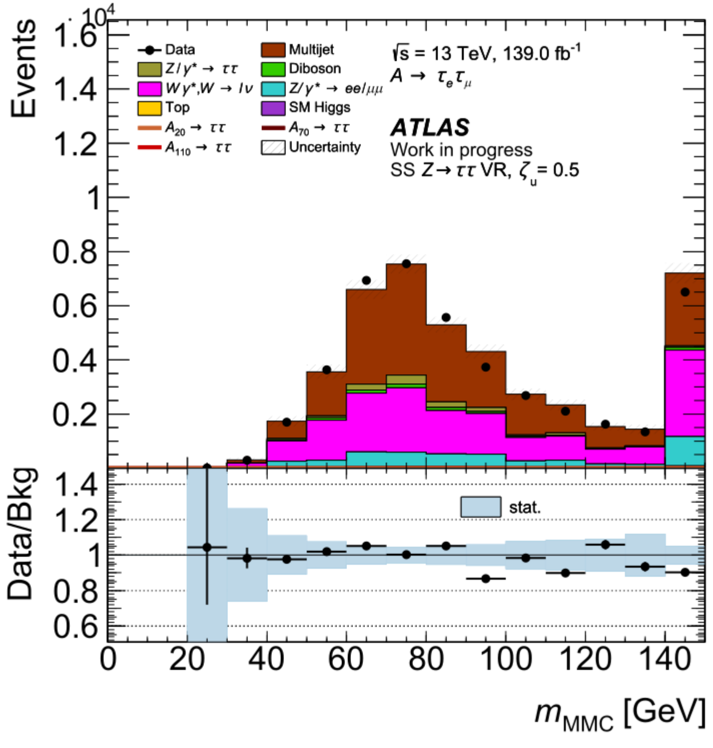



# More distributions in the same-sign Z→ττ validation region



# Alternative: Matrix Method

- estimate single and double fake events
- Find a relation between real/fake leptons reconstructed as tight/loose
- Real efficiency ϵ:
  probability to reconstruct real lepton as tight
- Fake efficiency f:
  probability to reconstruct a fake lepton as tight

- Easy case with one lepton:

$$\begin{pmatrix} T \\ L \end{pmatrix} = \begin{pmatrix} \epsilon & f \\ (1-\epsilon) & (1-f) \end{pmatrix} \cdot \begin{pmatrix} R \\ F \end{pmatrix}$$

- By inverting the matrix: calculate R and F
  → contribution of fakes to tight signal
  (in the two lepton case):

$$TT_{fake} = \epsilon f \cdot R F + f \epsilon \cdot F R + f f \cdot F F$$

# Matrix Method

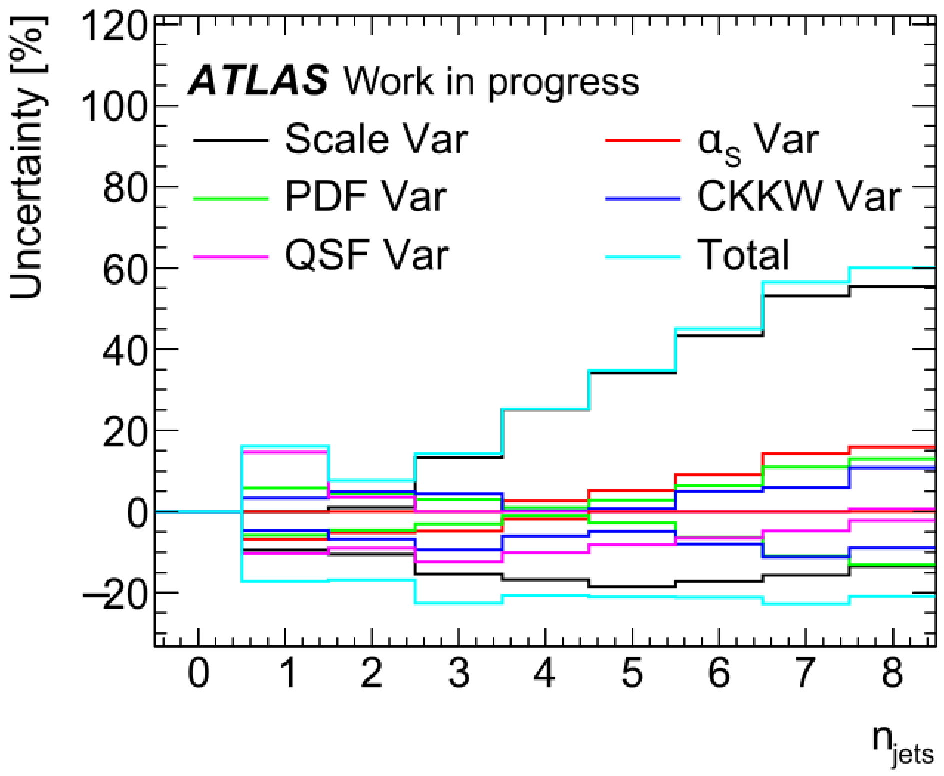
- **Formula for the case of one electron and one muon in the final state:**

$$\begin{pmatrix} T_e T_\mu \\ T_e L_\mu \\ L_e T_\mu \\ L_e L_\mu \end{pmatrix} = \begin{pmatrix} \epsilon_e \cdot \epsilon_\mu & \epsilon_e \cdot f_\mu & f_e \cdot \epsilon_\mu & f_e \cdot f_\mu \\ \epsilon_e \cdot (1-\epsilon_\mu) & \epsilon_e \cdot (1-f_\mu) & f_e \cdot (1-\epsilon_\mu) & f_e \cdot (1-f_\mu) \\ (1-\epsilon_e) \cdot \epsilon_\mu & (1-\epsilon_e) \cdot f_\mu & (1-f_e) \cdot \epsilon_\mu & (1-f_e) \cdot f_\mu \\ (1-\epsilon_e) \cdot (1-\epsilon_\mu) & (1-\epsilon_e) \cdot (1-f_\mu) & (1-f_e) \cdot (1-\epsilon_\mu) & (1-f_e) \cdot (1-f_\mu) \end{pmatrix} \cdot \begin{pmatrix} R_e R_\mu \\ R_e F_\mu \\ F_e R_\mu \\ F_e F_\mu \end{pmatrix}$$

# Reweighting of the $Z \rightarrow \tau\tau$ background

# Mismodelling in the Z→ττ validation region

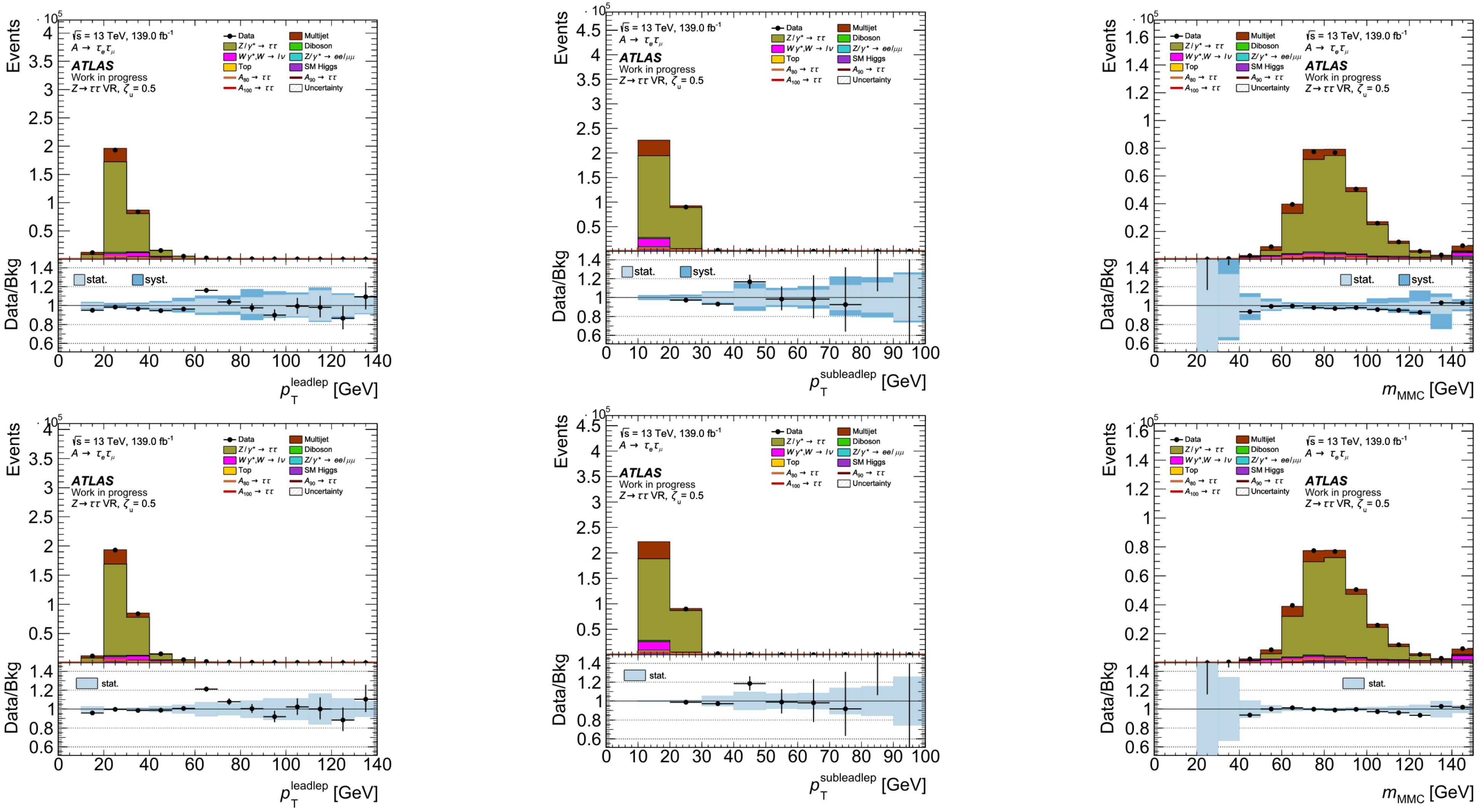
- **Recent development:**
  **similar behaviour in generator uncertainties of renormalization scale of the Z→ττ samples**



# More changes in Z→ττ validation region due to reweighting

**Before reweighting**

**After reweighting**



# Change in limit due to reweighting procedure