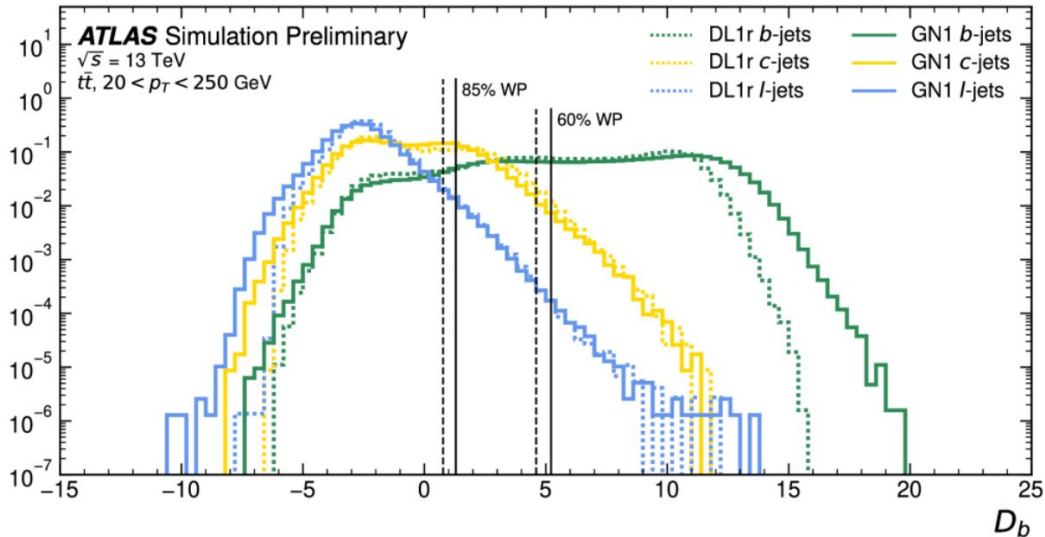# Jet Tagger Calibration

Liam Foster

# Overview

We will go over:

1. A brief reminder about flavor tagging
2. What calibration means
3. What we need to understand to do calibration
4. How we actually do calibration
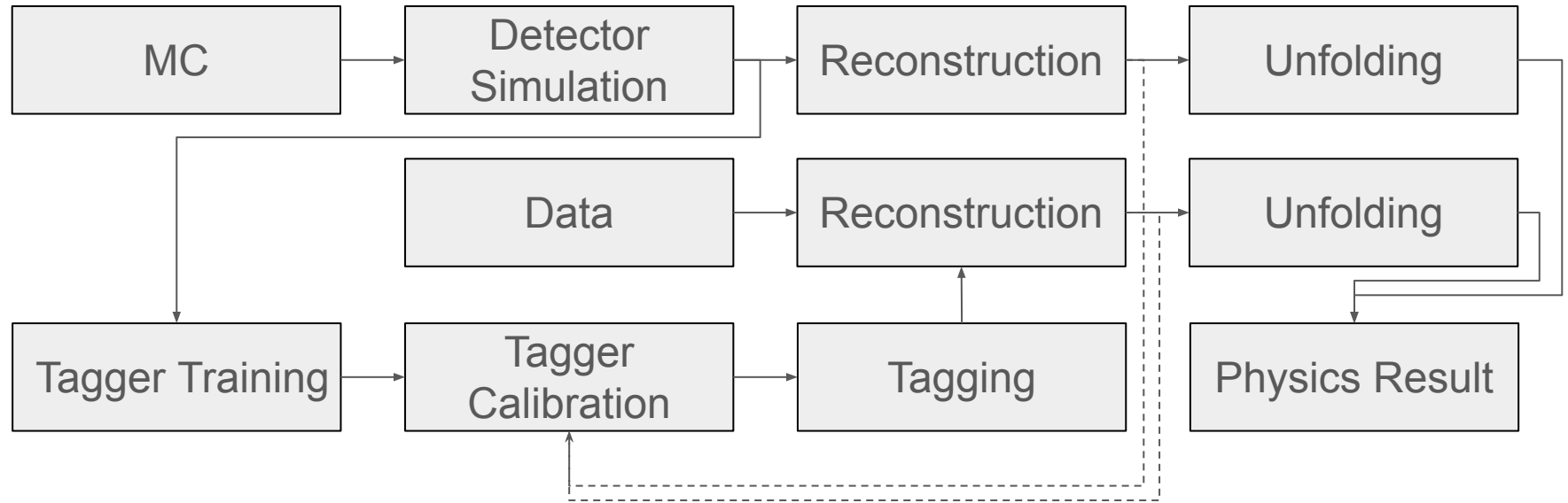5. How calibration methods lead to process dependance

# Flavor Tagging Overview

- Discriminating jets based on initial parton is necessary for many analyses
- Flavor taggers give jets light (l), charm (c), bottom (b) scores
- Modern flavor taggers use ML models trained on MC
- MC mismodelling can carry over into tagger training
- Charlie gave a great overview a couple weeks ago



$$D_b = \log \left[ \frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_u} \right]$$

Ref  3

# How Tagging Calibration fits into General Analysis Workflow

| | | | |
|---|---|---|---|
| MC | Detector Simulation | Reconstruction | Unfolding |
| | Data | Reconstruction | Unfolding |
| Tagger Training | Tagger Calibration | Tagging | Physics Result |

- Basic goal of tagger calibration is to understand how MC mismodelling affects tagging and determine uncertainties

# Why is Calibration Necessary?

- Analyses need to understand tagger efficiency to correctly extract numbers of events
- Tagger efficiency may vary when applied to MC or data
- Tagger efficiency may be process dependant
- Some jet parameters are known to be generally modeled poorly
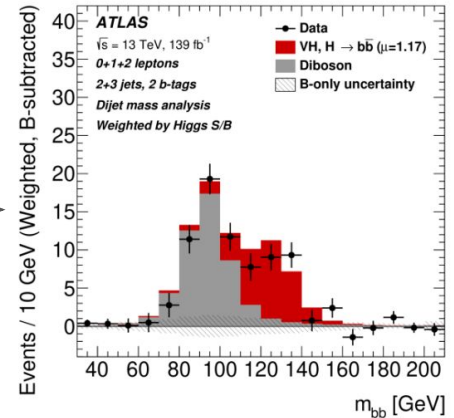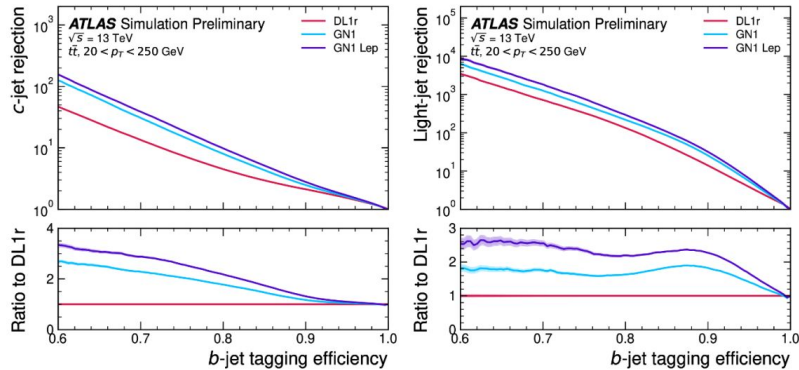
# How do we Calibrate?

- Need to determine efficiencies
- MC efficiency is easy, data efficiency is harder
- Extract SFs from 'well known' sample dominated by f-jets

$$SF^f = \frac{\varepsilon^f_{\text{data}}(p_T, \ldots)}{\varepsilon^f_{\text{MC}}(p_T, \ldots)}$$

f = flavor, process

# Example: b-tagger Calibration Overview

- Need to understand efficiencies of l, c, b-tagging on data
- Need samples enriched in jet flavor of interest with well known composition
- Samples need to be determined without use of tagger
- Separate MC-MC comparisons can be used to try to remove process dependance
- Will go through a handful of different SF analyses
- Potential drawback of this presentation strategy is I am not knowledgeable enough on this topic to judge the quality of these analyses…

# Example: b-tagger l-mistag efficiency

- Use Z+jets
- b-taggers are too good at l-rejection, too few l-jets left after use of b-tagger
- Use 'flipped' b-tagger to suppress b,c-tag efficiency without changing l-tag efficiency
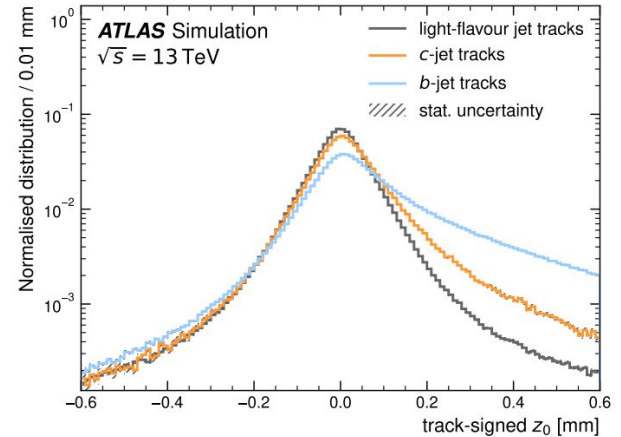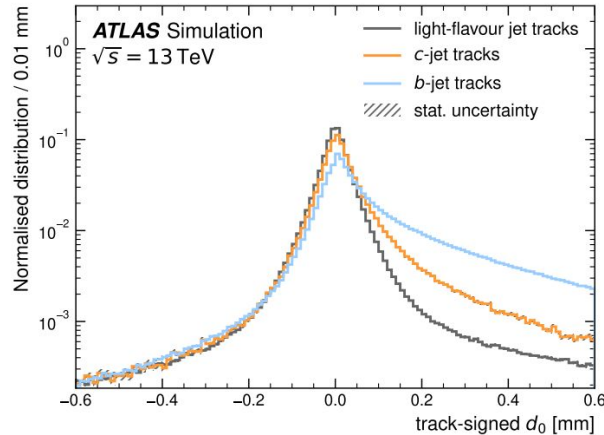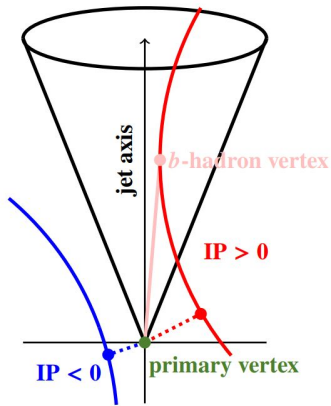
# Example: b-tagger l-mistag efficiency

- Perform fit in tagger discriminant, and SV mass
- Extract SFs
- A bit confusing to me to use MC to characterize MC mismodelling
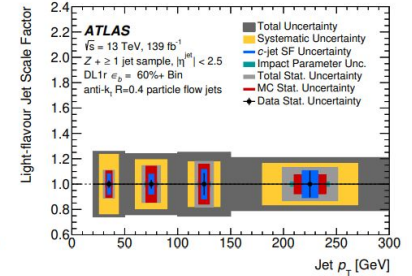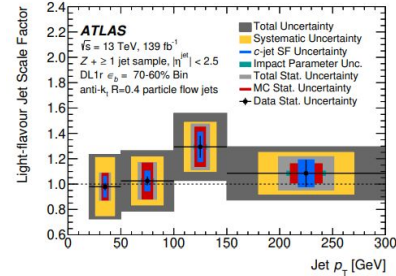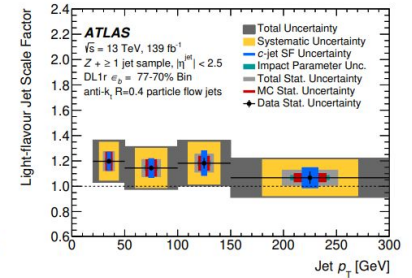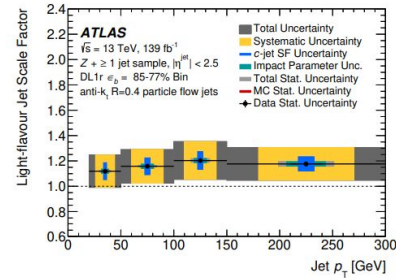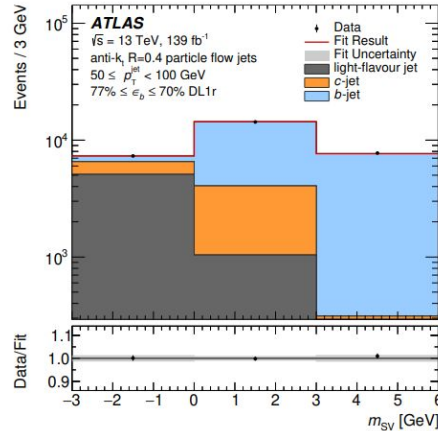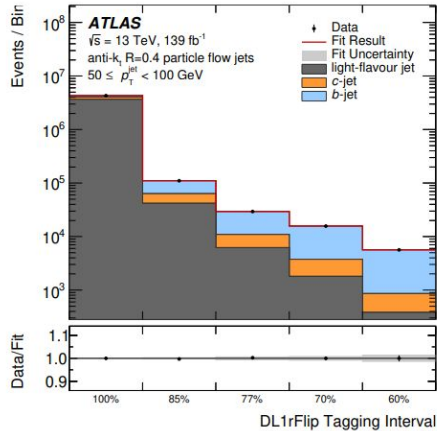
# Example: b-tagger c-mistag efficiency

- Historically has mainly used single lepton ttbar, but W+c has also been used
- Use log likelihood fit to reconstruct ttbar and tag jets as originating from t or W
- Applies b-tagging to t-jets
- W branching ratio is well known, so W jet sample composition should be too
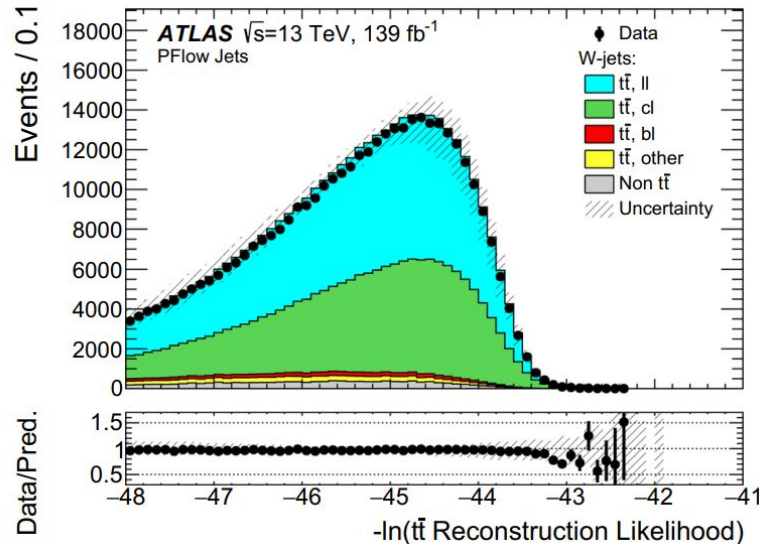
# Example: b-tagger c-mistag efficiency

- $\chi^2$ fit done in bins of W-jet $p_T$ and tagging interval
- SFs are left as a free parameter and extracted from the fit
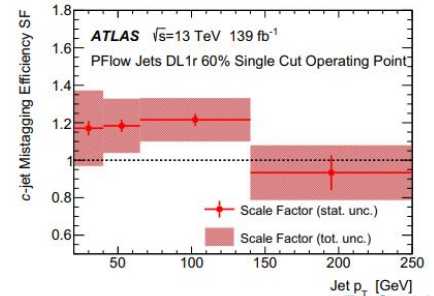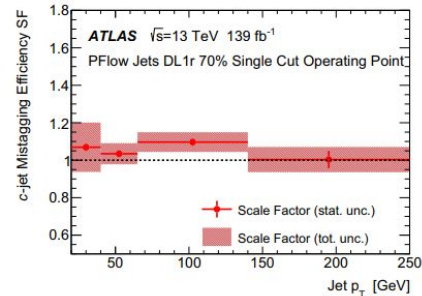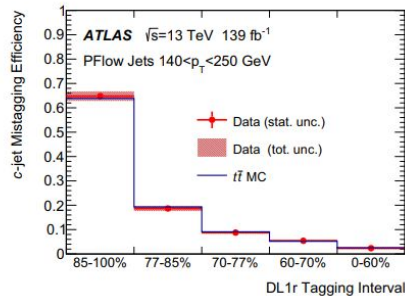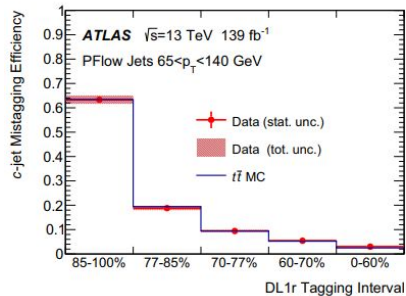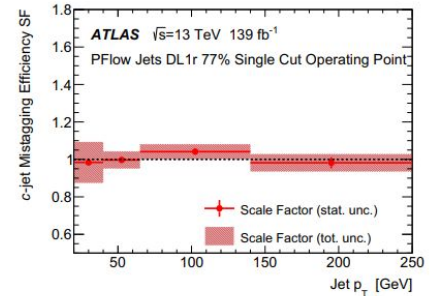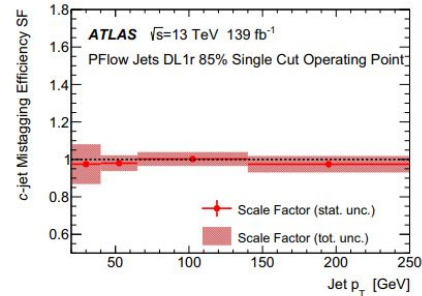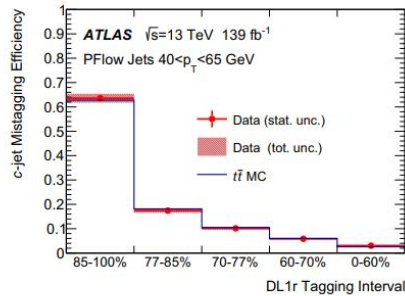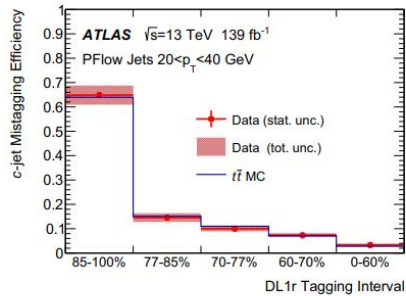
# Example: b-tagger b-tag efficiency

- Use ttbar eμ dilepton dijet channel, with lepton $p_T \geq 50$ GeV to reduce τ background
- Jets and leptons are paired into t candidates by minimizing $m^2_{j1+l1}+m^2_{j2+l2}$ to reward reconstructing objects of similar mass
- Main remaining background is ttbar ISR (FSR), and denoted as bl (lb)
- Kinematics allows events to be split into bb, bl, lb, ll enriched regions based on $m_{j1,l}$, $m_{j2,l}$
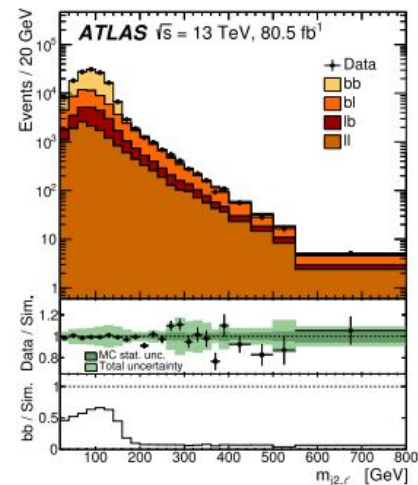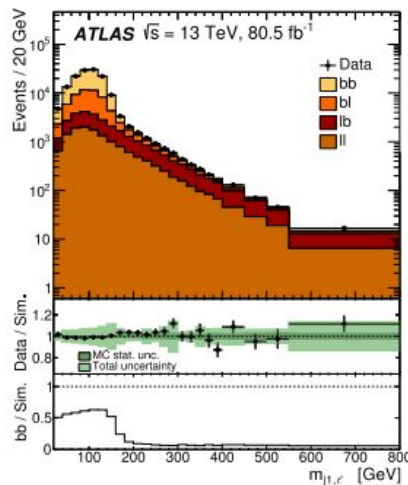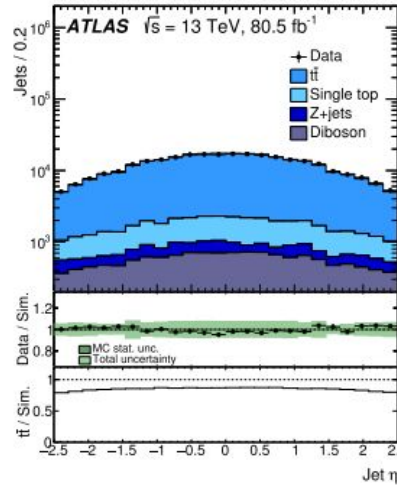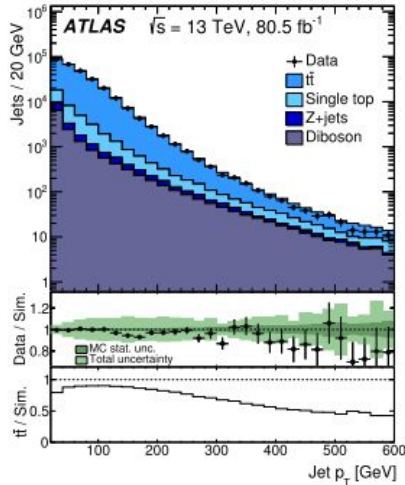
# Example: b-tagger b-tag efficiency

- SR classified via ABCD method for each bin of leading, subleading jet $p_T$
- Log likelihood fit is performed, with probability of b-tag as a nuisance parameter
- SFs are extracted as a function of jet $p_T$



13
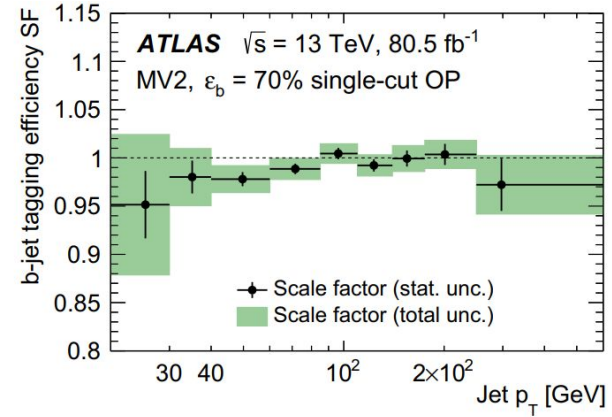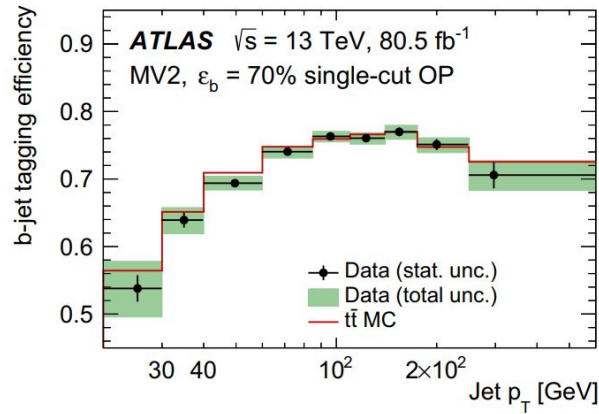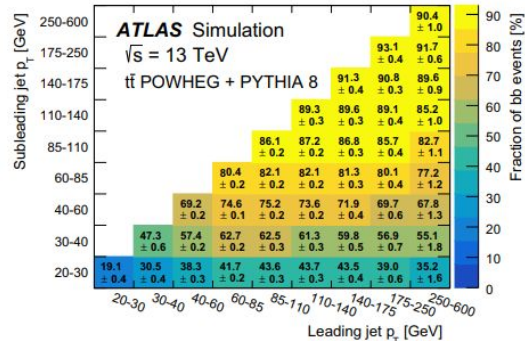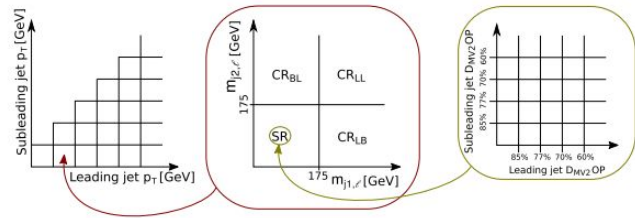
# Process Dependance

- The methods described for determining SFs are naturally optimized for the samples they use
- MC to MC SFs can be determined, but checking against data is superior
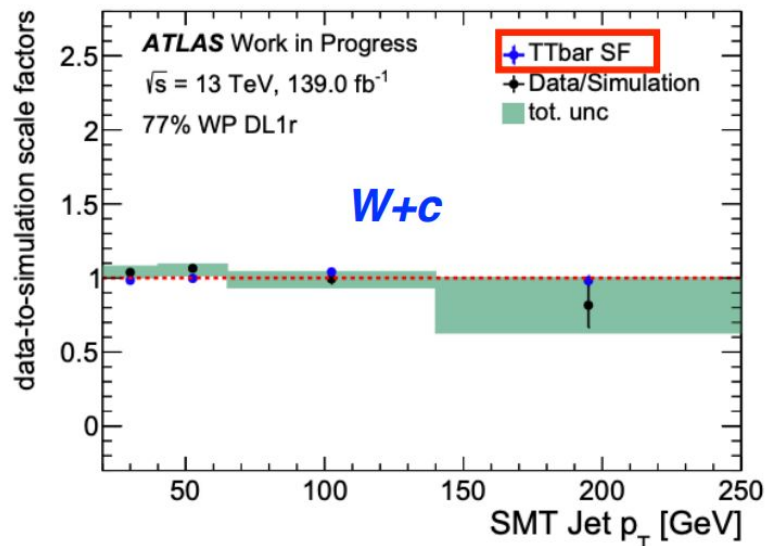
# Summary

We've gone over

1. What SFs are
2. Why SFs are important
3. How SFs are generally determined
4. How SFs tend to be process dependant

# A much better talk by Valerio Dao

# Thanks! Questions?