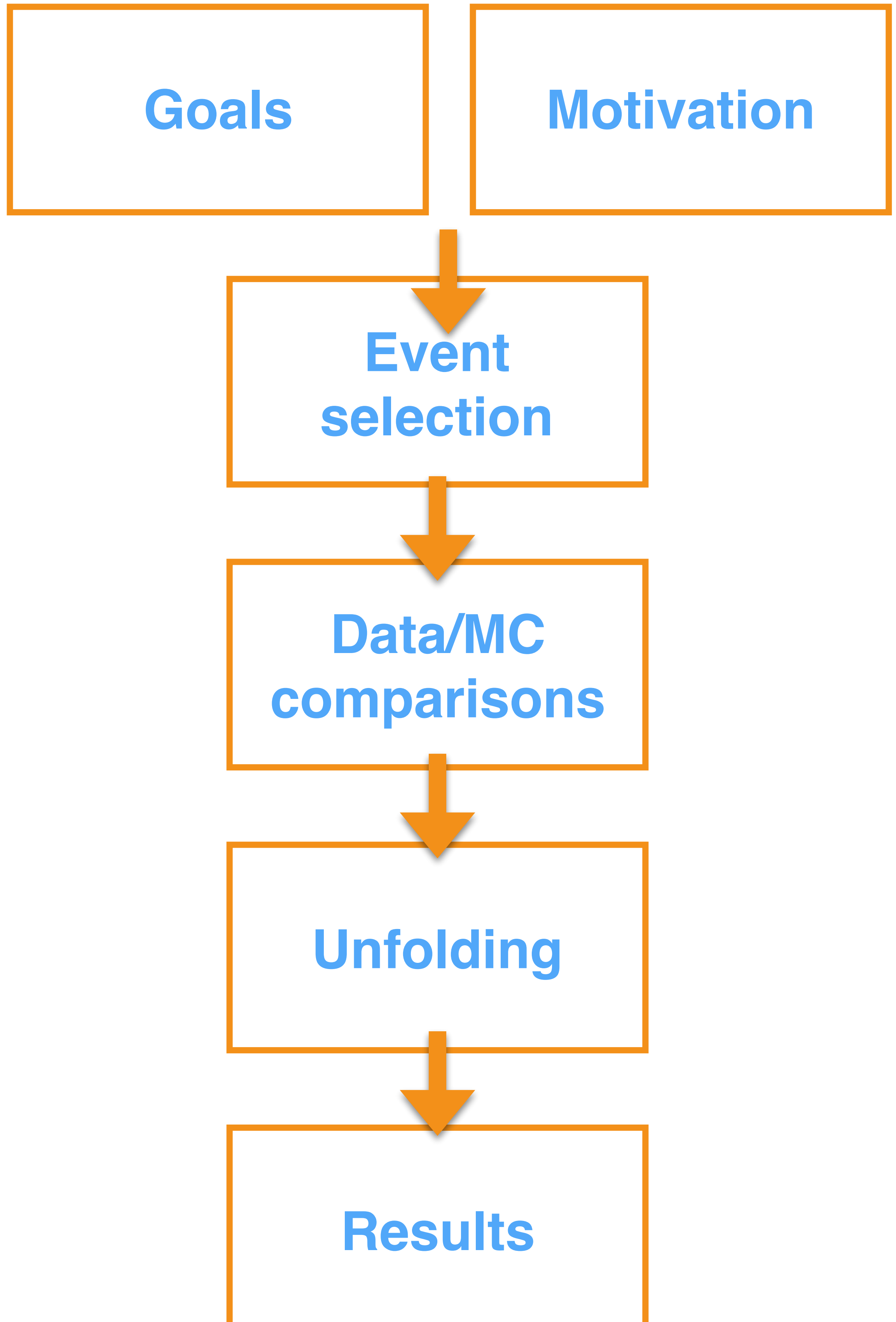
Boosted Z+bb Analysis

Chloe Gray
University of Glasgow

IOP APP and HEPP conference
University of Bristol

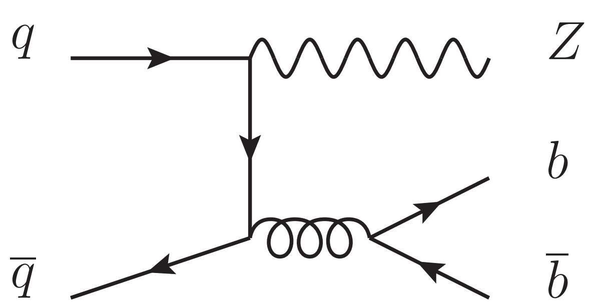
26th March 2018

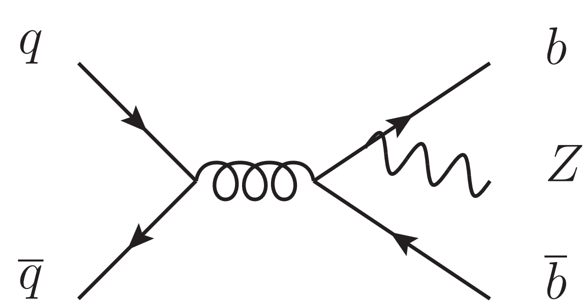
Outline


Goals
Motivation
Event selection
Data/MC comparisons
Unfolding
Results

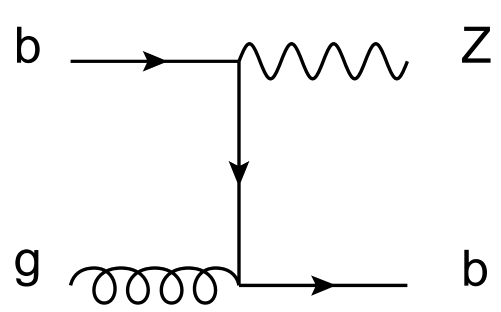
Motivation

- **Hasn't been measured** in the boosted phase space
- **Important background** for VH(bb), ttH(bb), exotics/resonances searches and measurements
- Sensitive to the **b-flavour component of PDFs**
  - Two different schemes can be used for heavy-flavour calculations:
    - 4 Flavour (4F) scheme: no b-quark in PDF, b-quark in shower



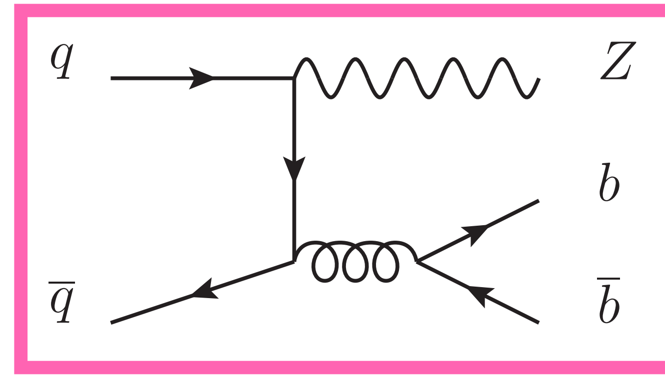


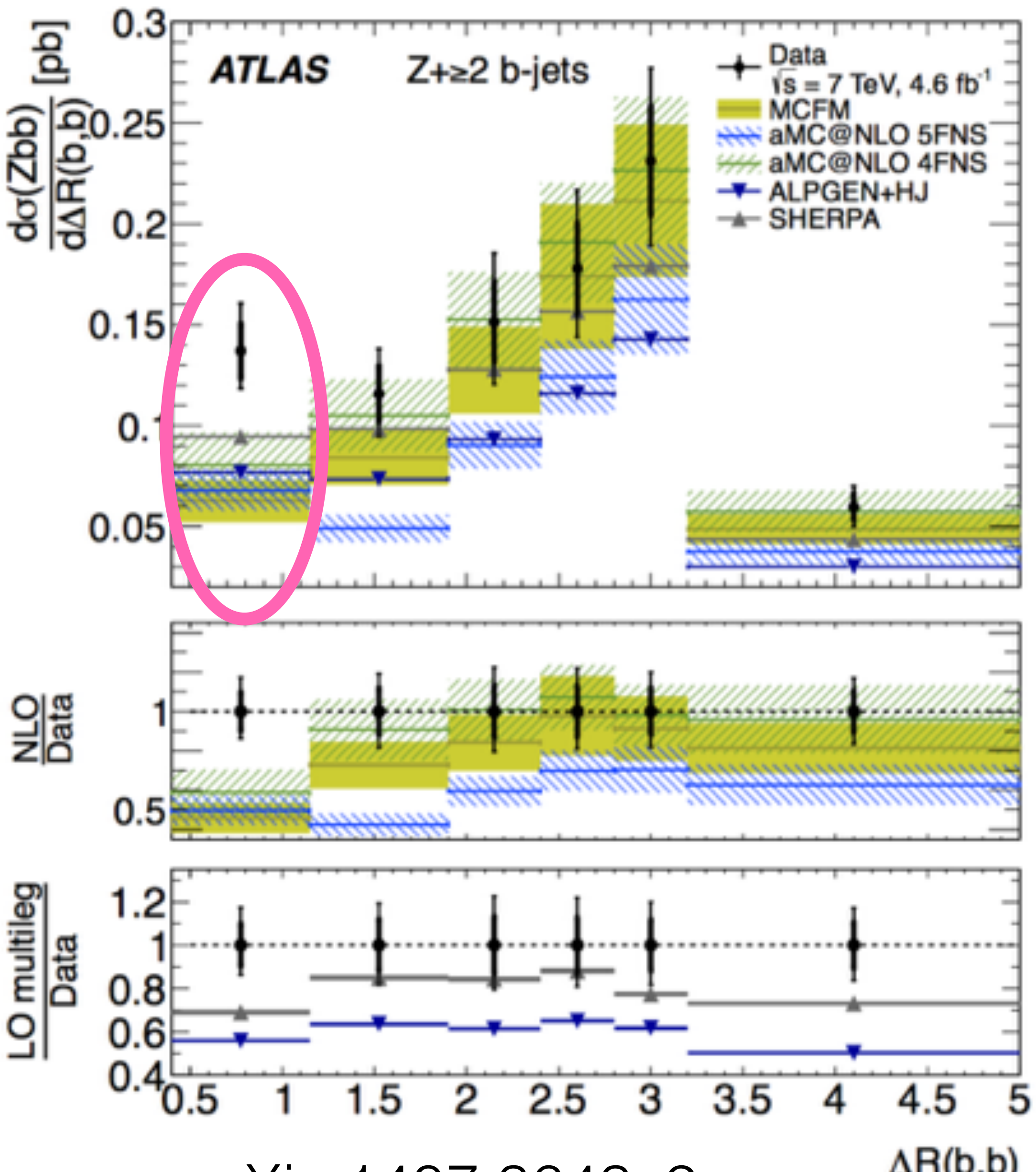
    - 5 Flavour (5F) scheme: b-quark in PDF and shower



Motivation

- The opportunity to study **g->bb splitting** helps with parton-shower modelling:



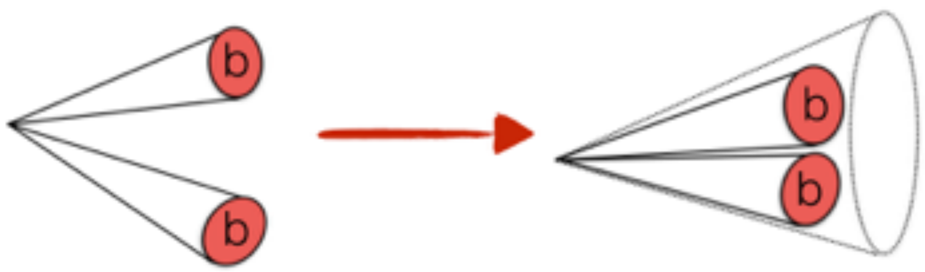


arXiv:1407.3643v2

- The low ΔR region corresponds to correlated b's which is typical of gluon splitting
- Found to be badly modelled in the ATLAS Run-1 Z+bb measurement
- We will be more sensitive to this region and we will be able to access smaller ΔR

Goals

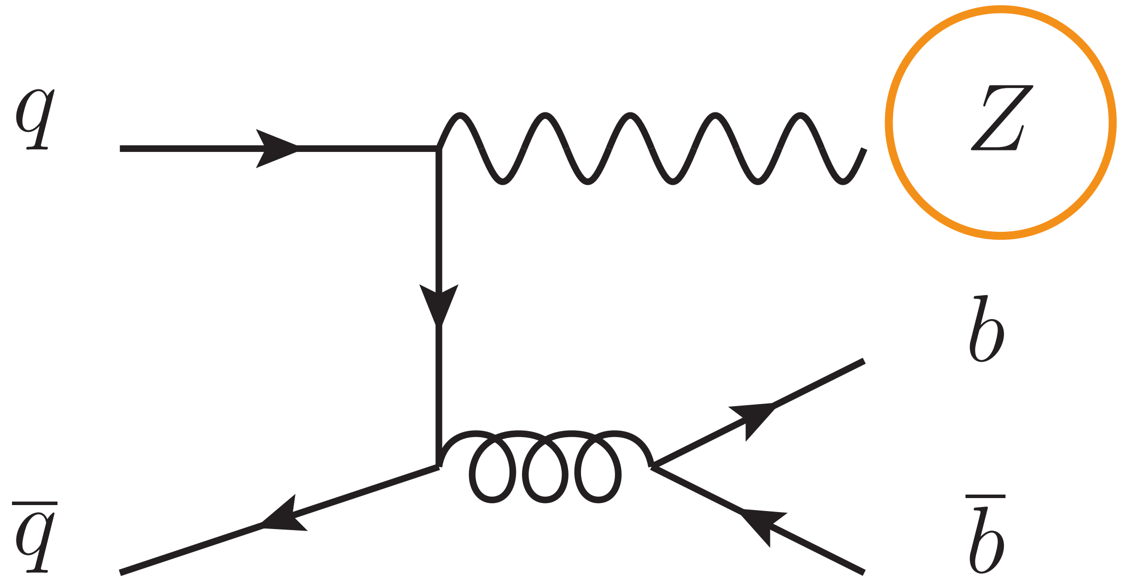
- Measure **differential fiducial cross sections** of large-*R* jets and tagged sub-jet variables in boosted Z+bb events



- **Unfold the data** and compare differential cross sections to different MC predictions
- **Primary observables:**
  - ΔR(b,b)
  - Large-R jet mass and pT in the inclusive (no tagging requirements) and 2 b-tag regions

Event selection

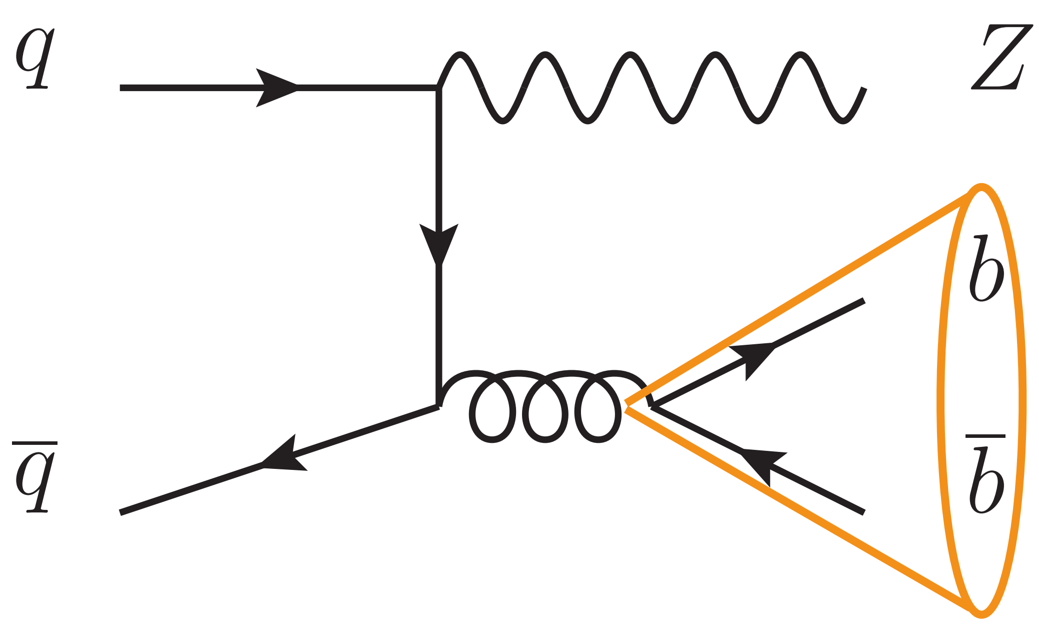
- To select the **Z+bb signal events**, we require:



  - 2 leptons: electrons or muons
  - $71 < m_{\ell\ell} < 111$ GeV

Event selection

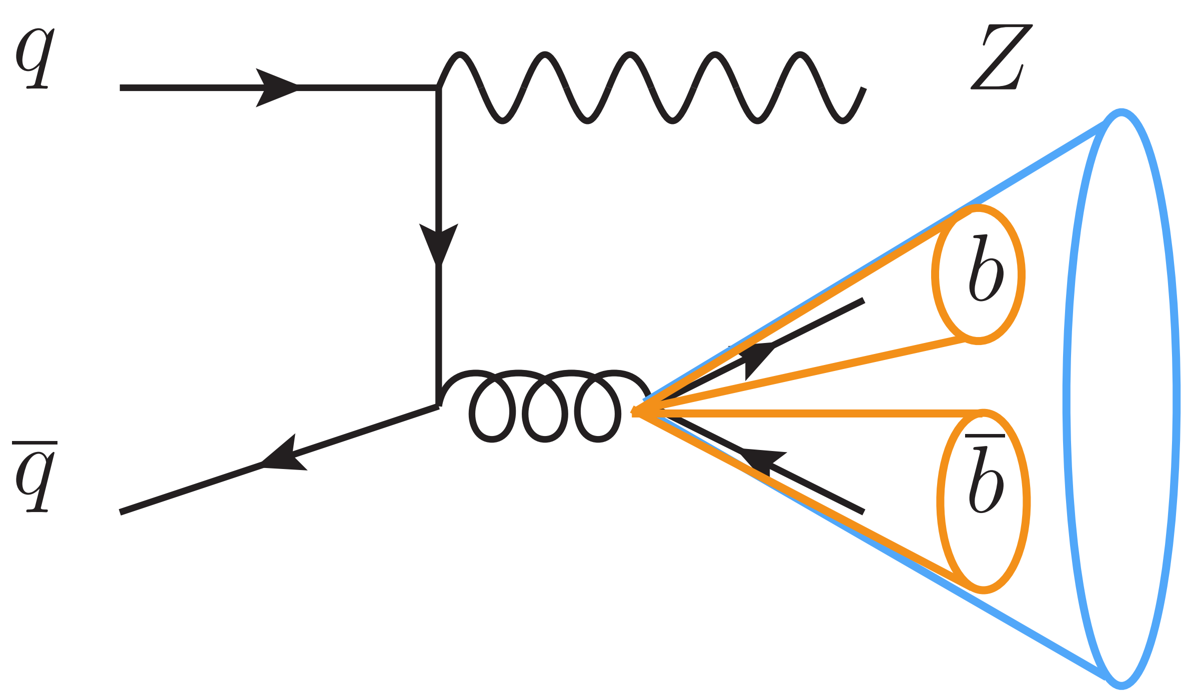
- To select the **Z+bb signal events**, we require:



  - 2 leptons: electrons or muons
  - $71 < m_{\ell\ell} < 111$ GeV
  - 1 large-radius (R = 1.0) jet with pT > 200 GeV

Event selection

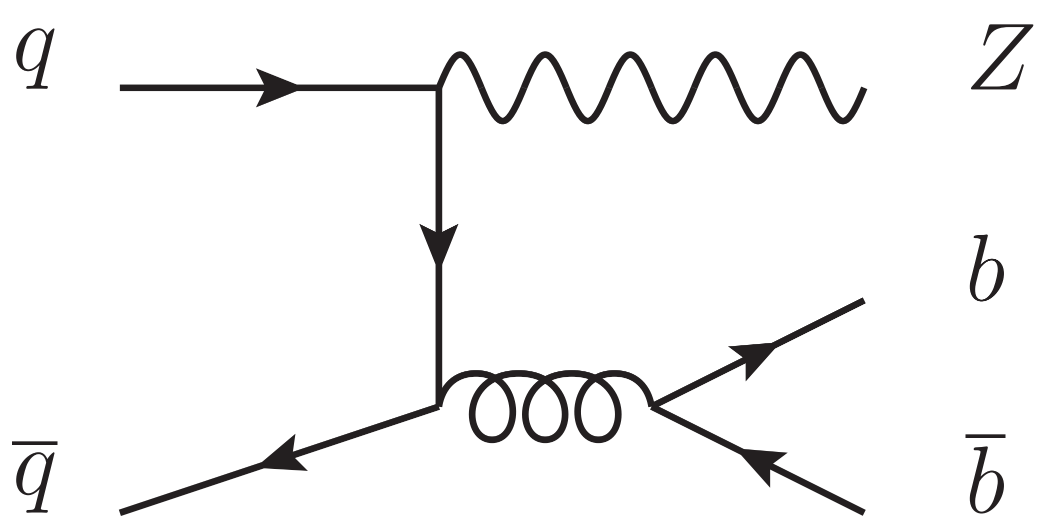
- To select the **Z+bb signal events**, we require:

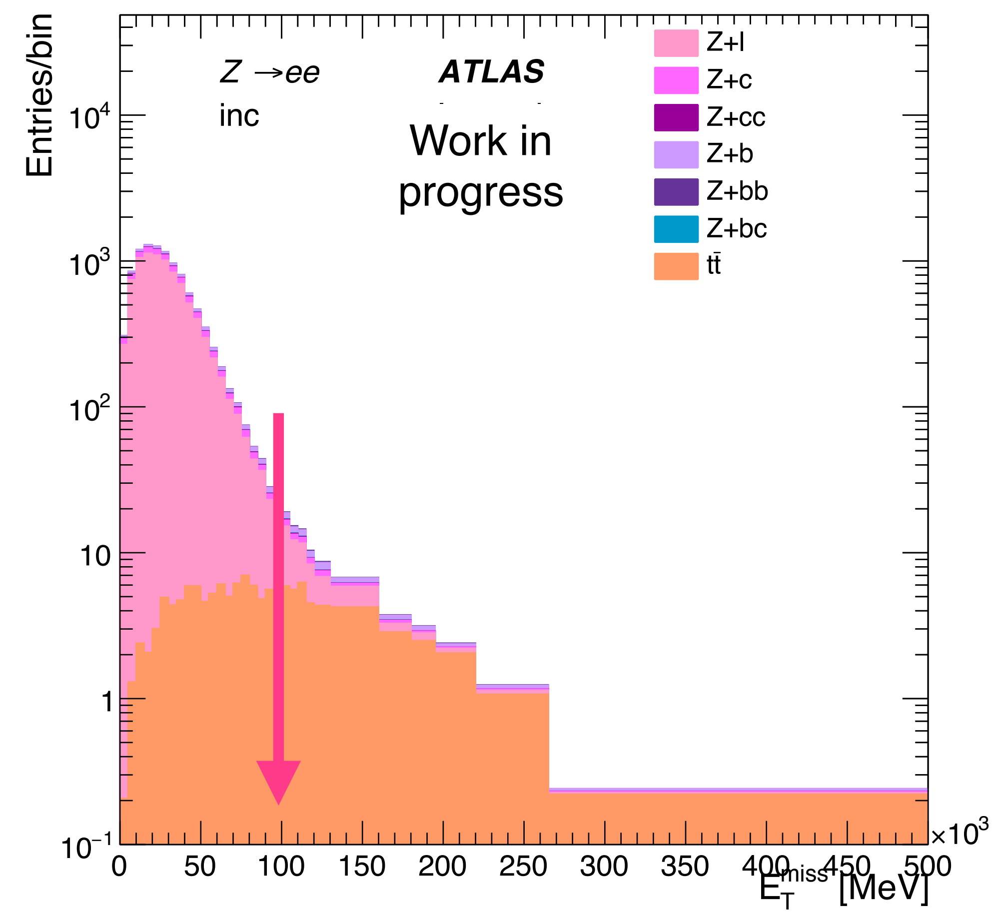


  - 2 leptons: electrons or muons
  - $71 < m_{\ell\ell} < 111$ GeV
  - 1 large-radius (R = 1.0) jet with pT > 200 GeV
  - Look at jets inclusively (no tag requirement) and with 2 b-tags

Event selection

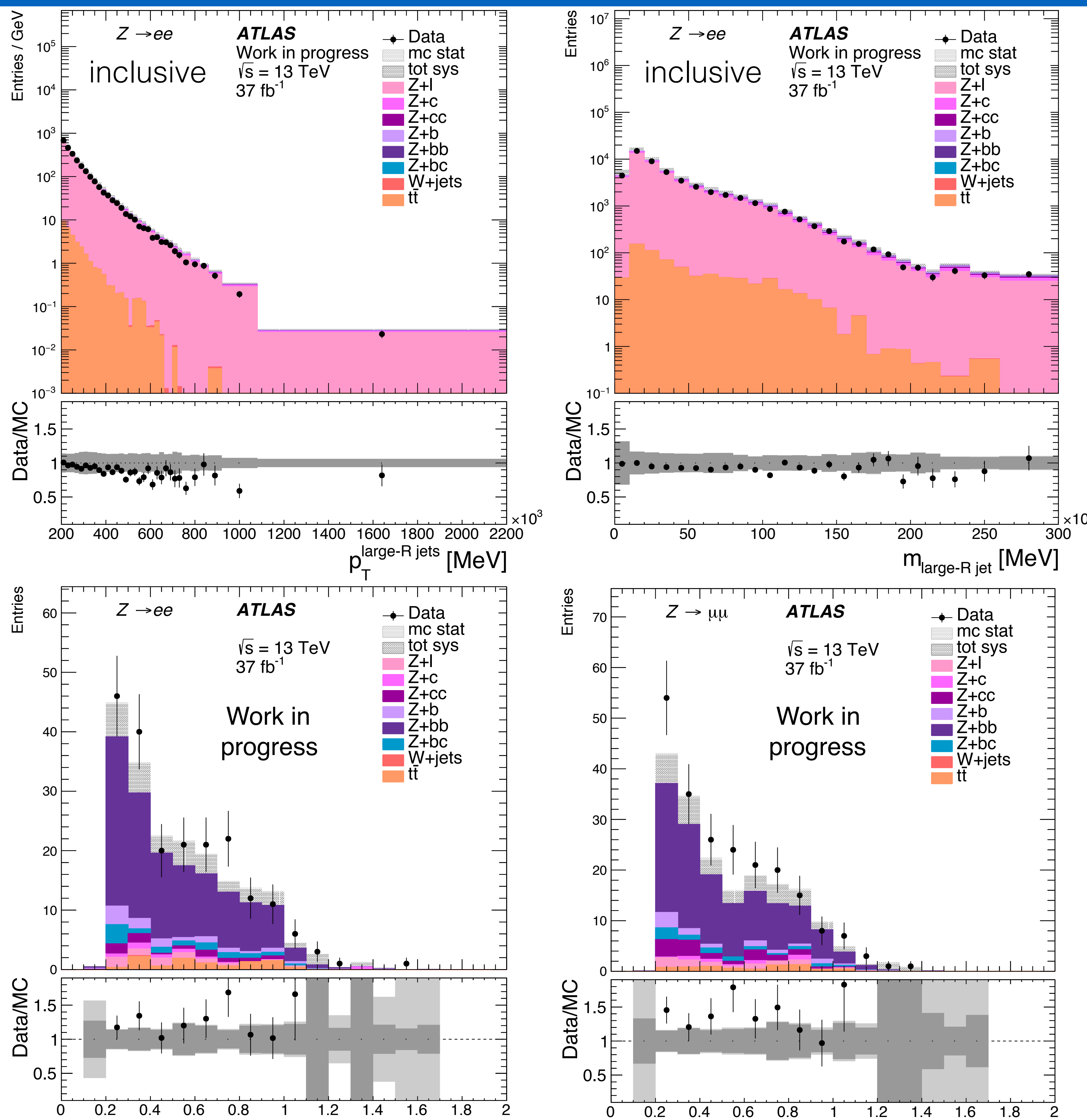
- To select the **Z+bb signal events**, we require:
  - 2 leptons: electrons or muons
  - $71 < m_{\ell\ell} < 111$ GeV
  - 1 large-radius (R = 1.0) jet with pT > 200 GeV
  - Look at jets inclusively (no tag requirement) and with 2 b-tags





- What about **backgrounds**?
  - Main background is **tt̄**
  - Apply MET < 100 GeV to reduce this
  - Z-mass window cut also helps

Data/MC comparisons



- Generally good modelling of inclusive variables, with the data undershooting the MC for pT
- ~20% difference in the 2-tag variables
- Systematic band contains detector systematics, signal-modelling and top-modelling uncertainties
- Dominant uncertainties come from large-R jet energy scale and b-tagging

Unfolding overview

- We unfold the data to particle level to compare to predictions
- We are using the **Fully Bayesian Unfolding (FBU)** method (arXiv:1201.4612)

Basic principle:

- Compute the likelihood of the data,*d,* given the signal cross sections, ***σ***, and nuisance parameters, *Λ*:

$$\mathcal{L}(d|\sigma,\Lambda) = \prod_{i \,\in\, \text{recobins}} \text{Poiss}(d_i|x_i(\sigma,\Lambda))$$

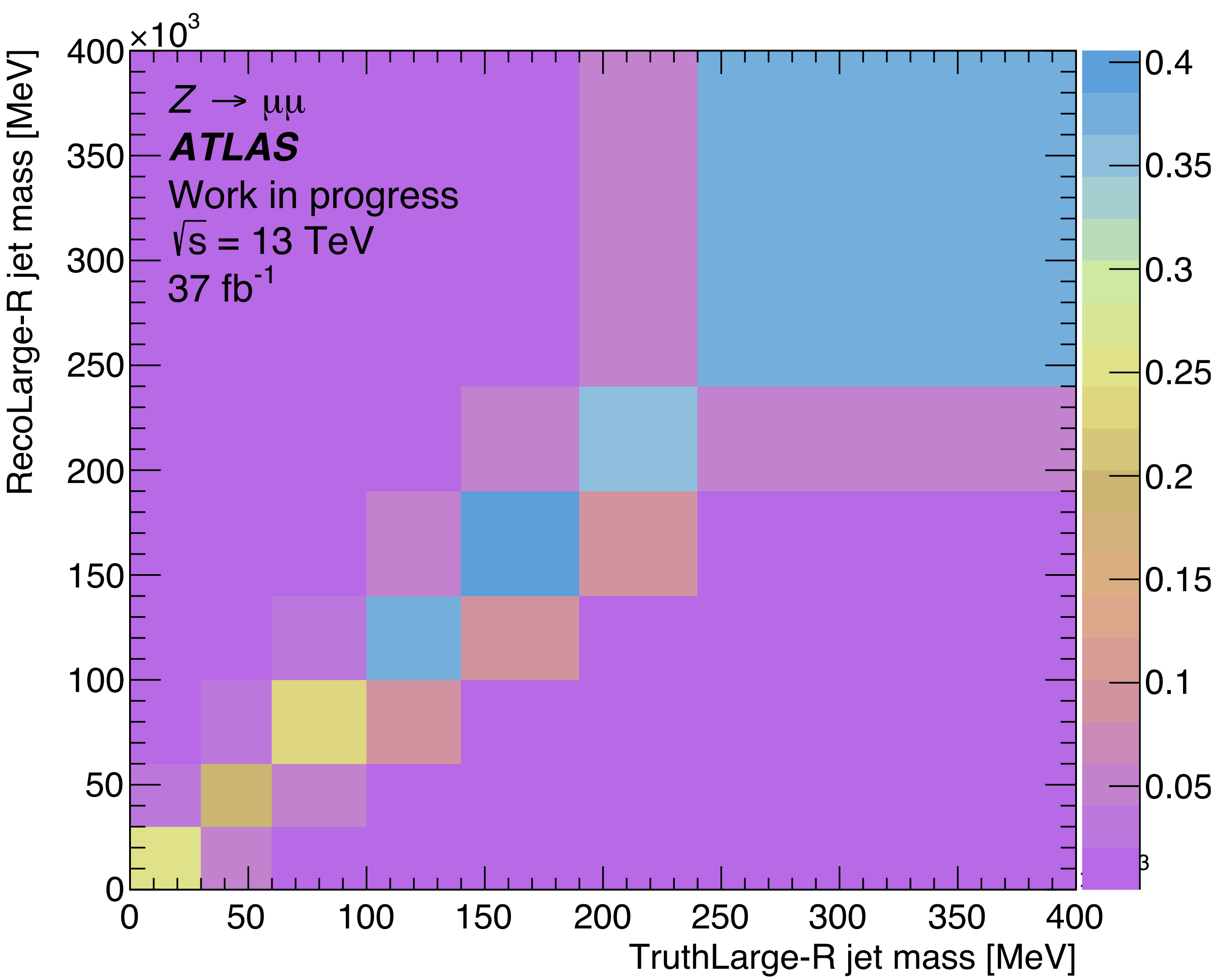
$$x_i(\sigma,\Lambda) = L(\Lambda)\times(b_i(\Lambda)+M_{ij}(\Lambda)\,\sigma_j)$$

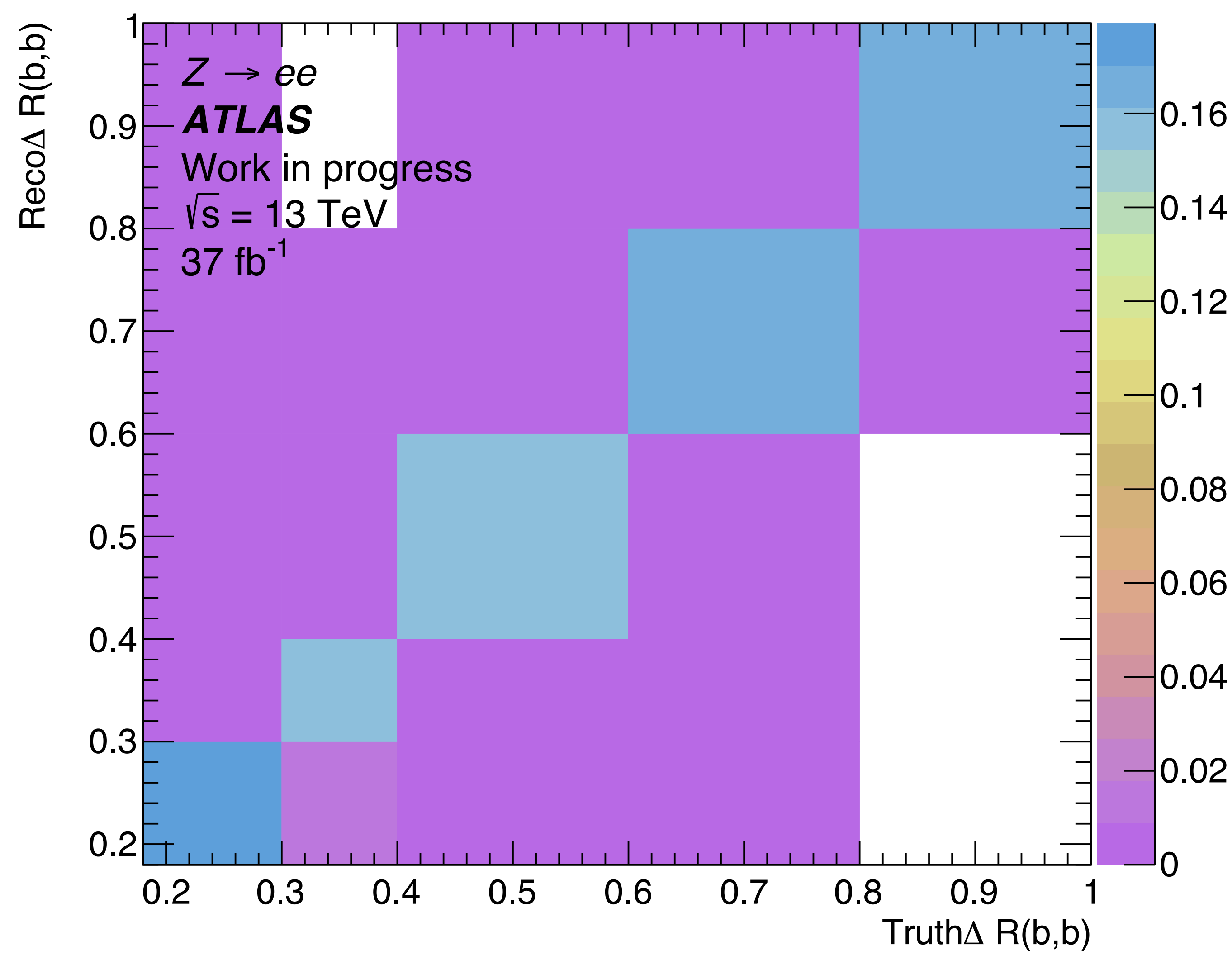
luminosity — background — migration matrix

- Posterior probabilities are then extracted by sampling the full *(σ,Λ)* space

Response matrices

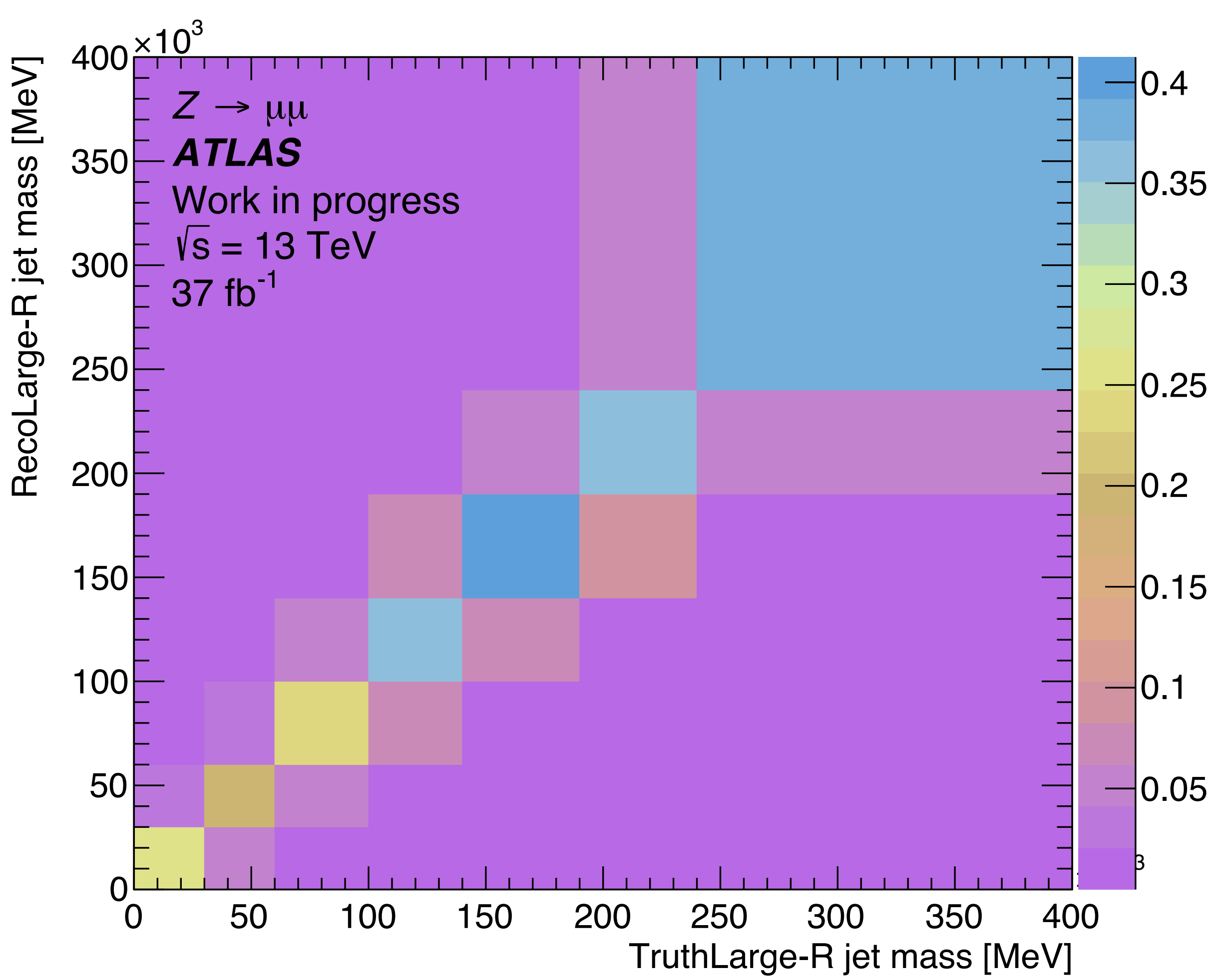
- Examples of nominal response matrices
- Events must fulfil particle and reco-level event definitions
- Both fairly diagonal

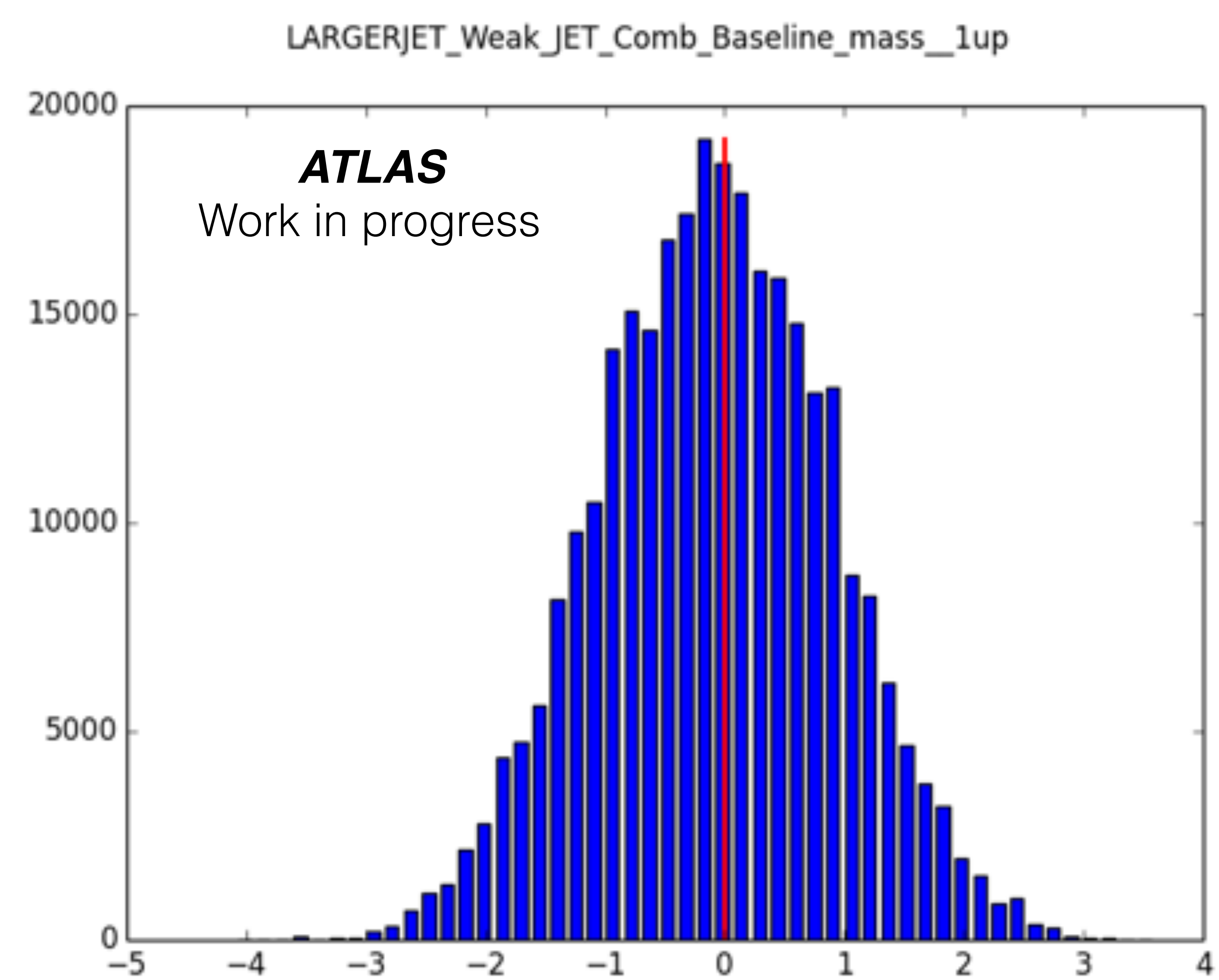




Backgrounds + uncertainties

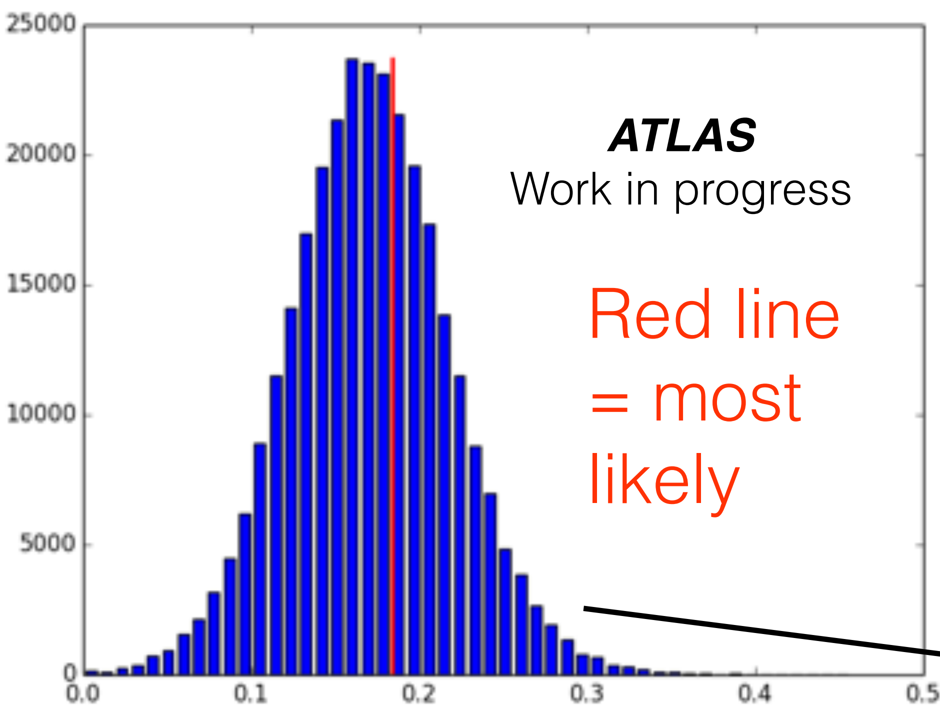
- Systematics and backgrounds handled using **nuisance parameters**
- Each systematic has a corresponding response matrix and background prediction
- Response matrices and backgrounds can be **smoothly varied** between the nominal and systematic
- Allows the unfolding to ‘wander around’ in the space of predictions

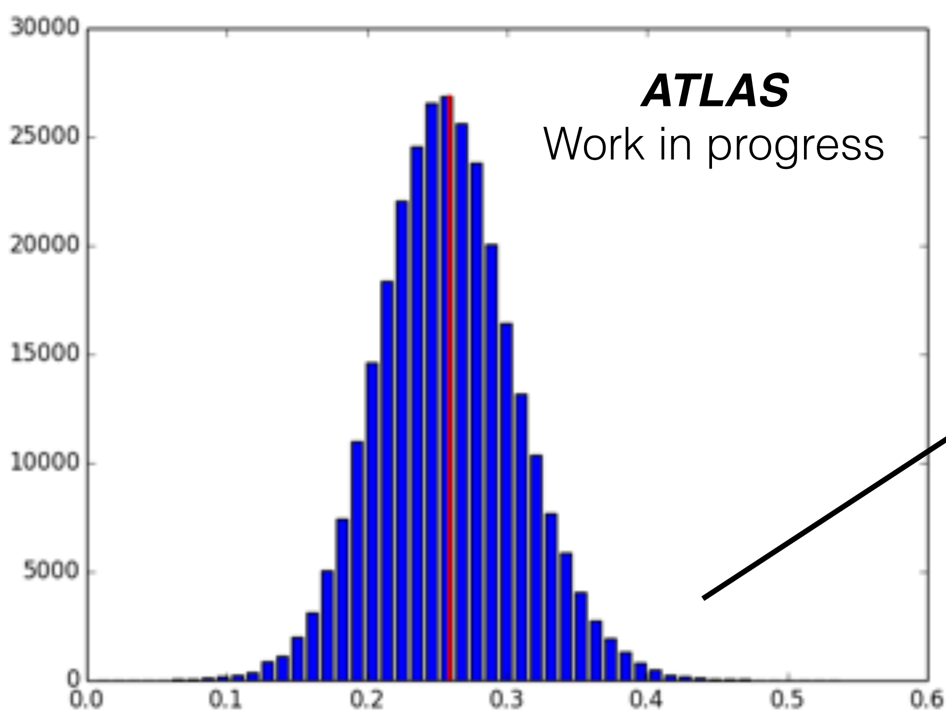




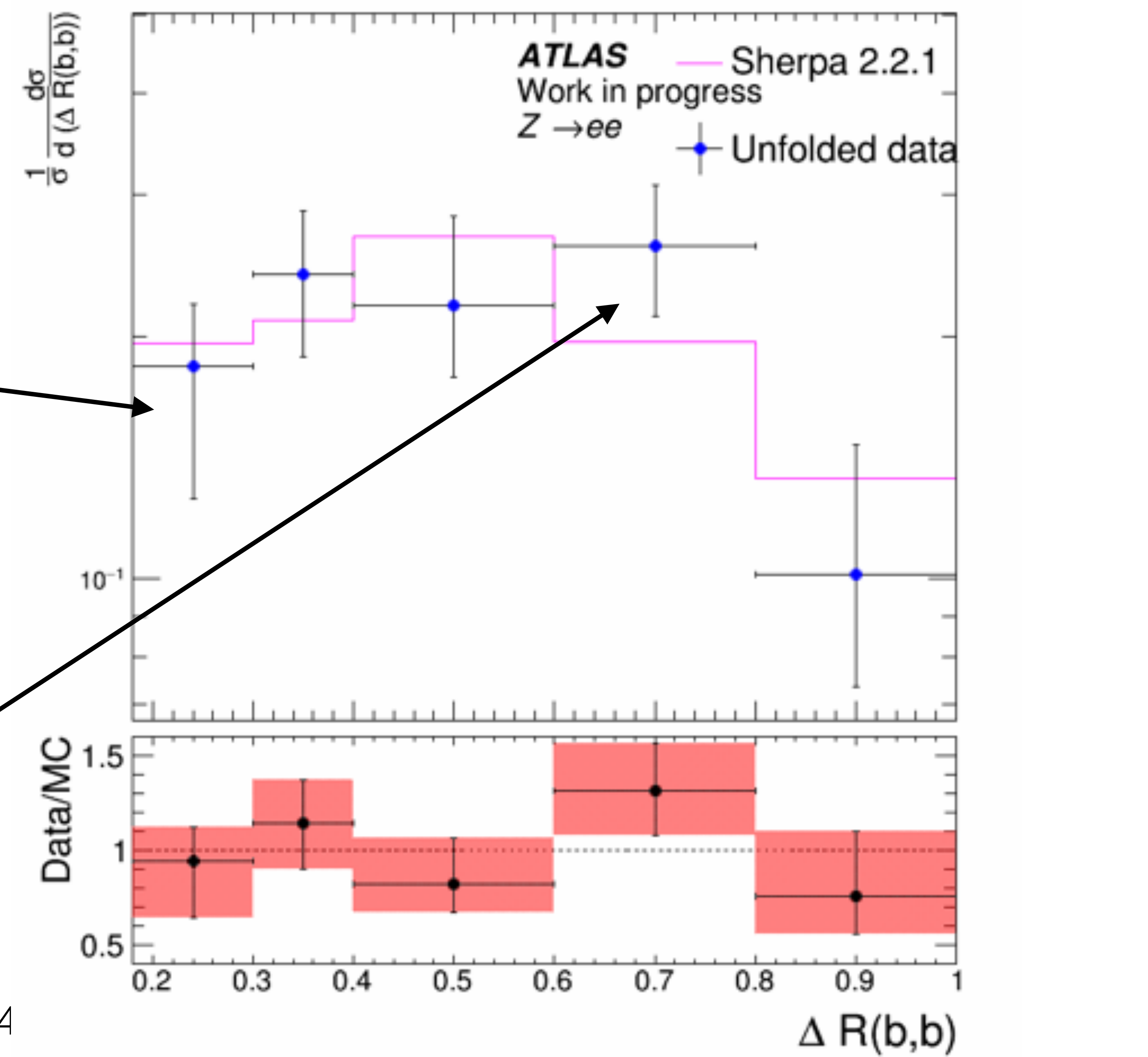
Posteriors

- The result is a set of posterior probability distributions

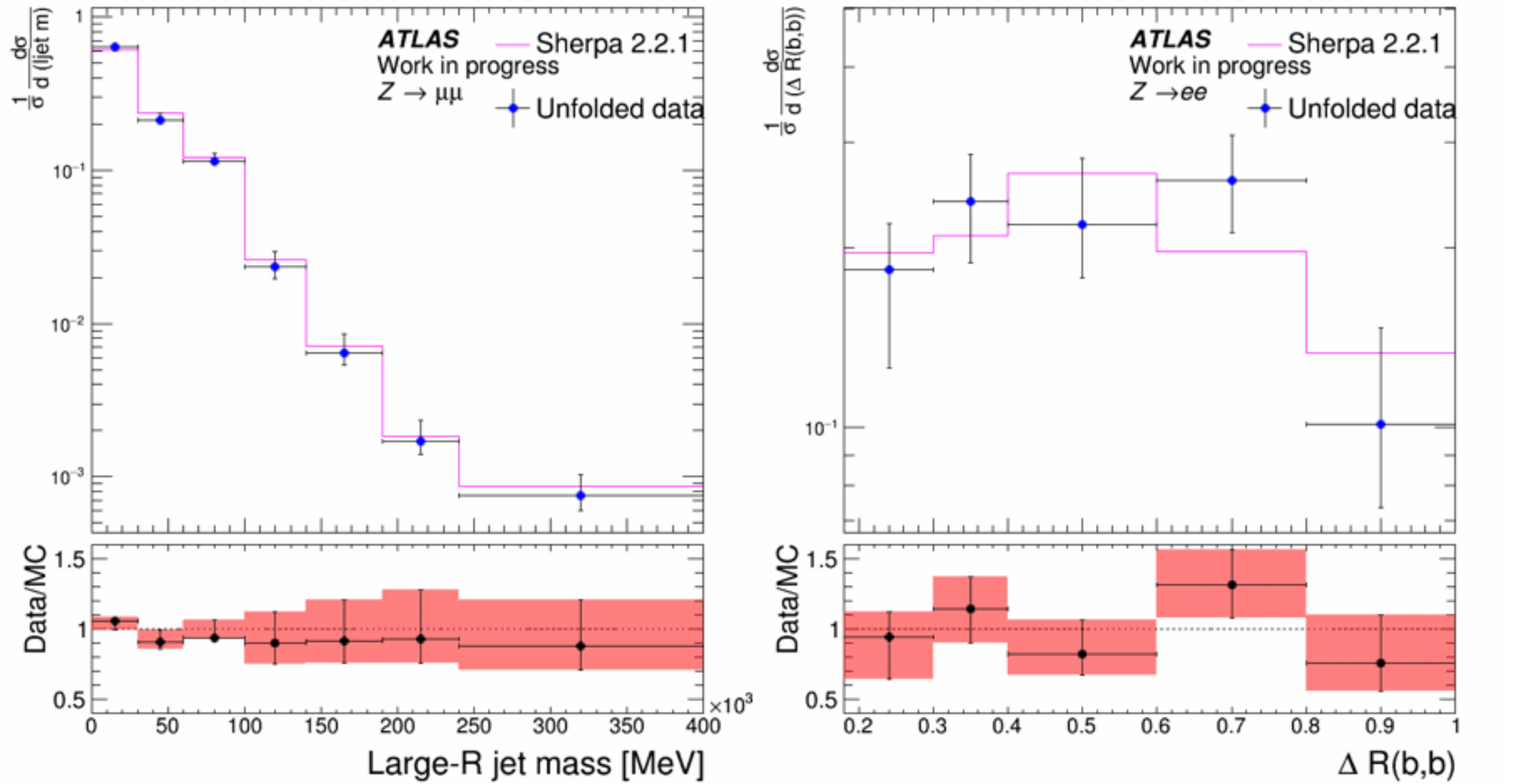




- Error band is defined by central region containing 68% of the probability



Results



- Error band includes systematic uncertainties
- No strong disagreement with respect to sherpa 2.2.1 for large-R jet mass
- Some disagreement in 0.6-0.7 ΔR region

Summary + outlook

- Analysis methodology of the measurement of differential fiducial cross sections in Z+bb events has been presented
- The use of the Fully-bayesian unfolding method is discussed
- Some unfolded results using the method are shown
- No strong disagreement between the data and Sherpa 2.2.1 is observed so far
- We would like to compare to other predictions and consider systematic uncertainties on the truth prediction

Backup

Object definitions

*If more than 1 large-R jet survives OR, throw event away

Electron
- **ID:** MediumLH
- **Isolation:** Gradient
- pT > 28 GeV

Muon
- **ID:** Medium
- **Isolation:** Gradient
- pT > 28 GeV

MET
- TST MET

Lepton - large-R jet OR
- if ΔR(ljet,lepton) < 1.2 , throw jet away *

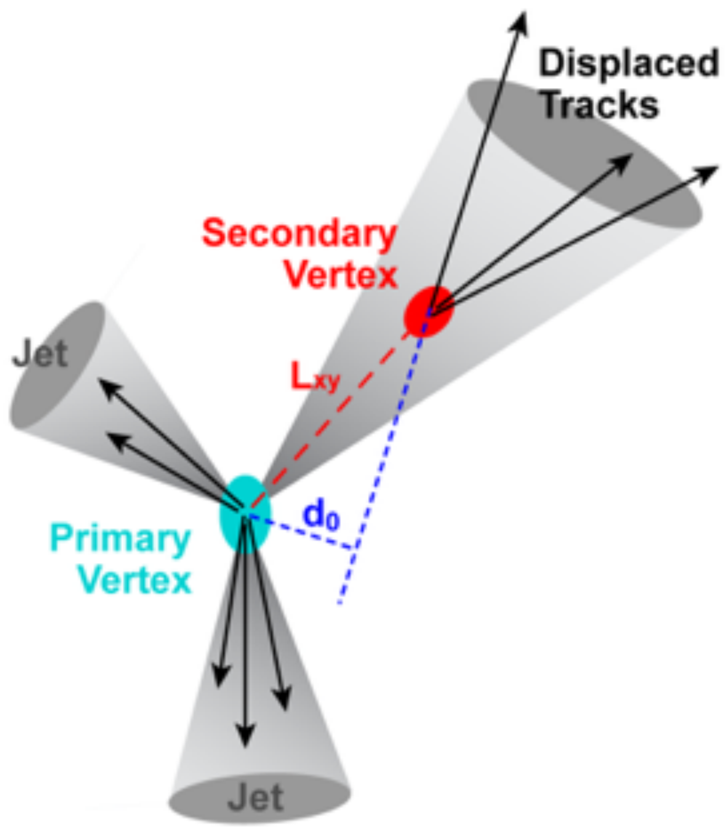
Large-R Jet
- AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets:
  - ➡ Anti Kt
  - ➡ Radius = 1
  - ➡ Trimming
- pT >= 200 GeV
- |eta| < 2

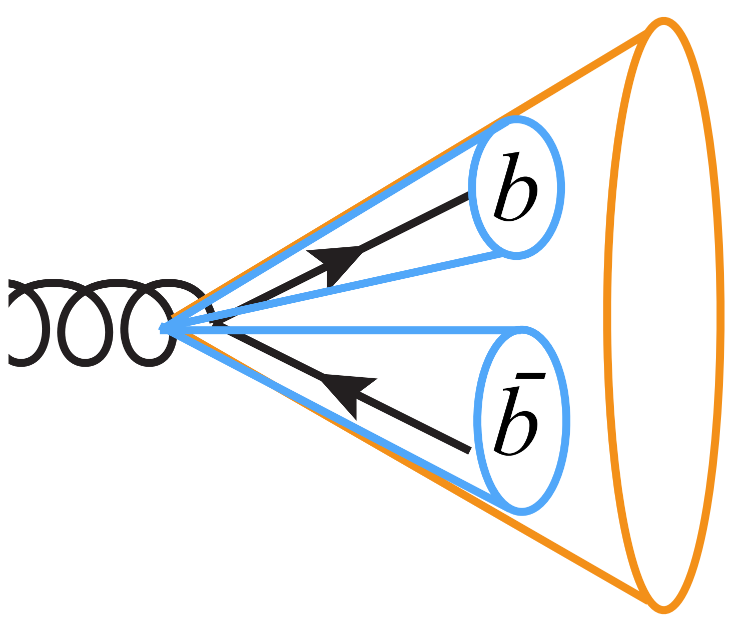
Track Jets
- AntiKt2PV0TrackJets
  - ➡ Anti Kt
  - ➡ Radius = 0.2
- pT > 10 GeV
- |eta| < 2.5
- N_{tracks} >= 2

B-tagging
- Track jets are ghost-associated to Large-R jet and b-tagging is applied to the track-jets
- mv2c10 > 0.6455 (70% working point)

How is the b-tagging performed?

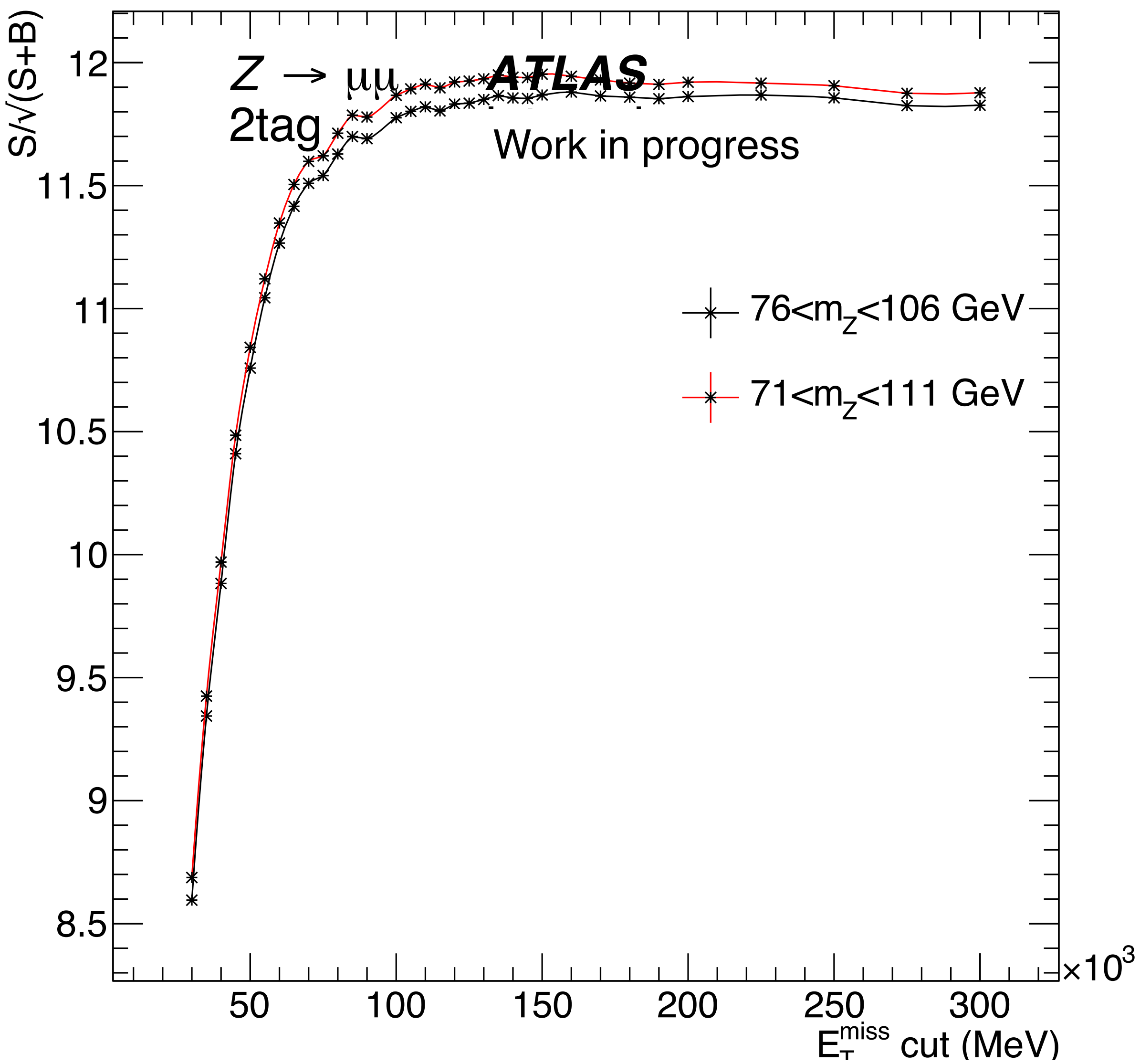
- Track-jets are considered b-tagged if they pass the 70% efficiency working point of the MV2c10 algorithm
- Properties of the b-hadron decay are used in the MV2c10 algorithm:
  - Secondary vertex displaced from primary vertex
  - Impact parameter (d0)
  - Decay length
- MV2c10 is an MVA (BDT specifically) which combines these properties

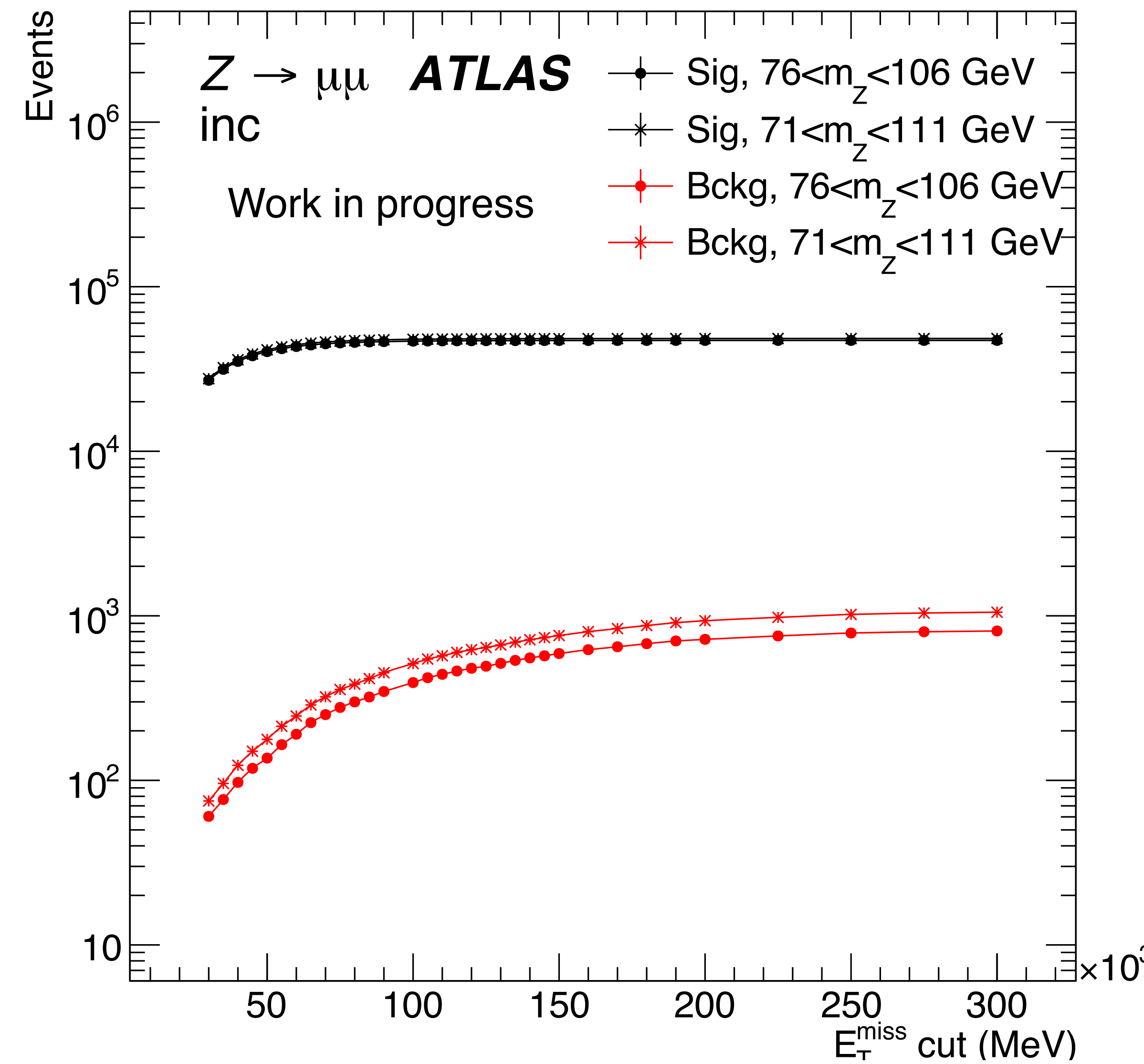




- In this analysis, we tag small-radius (R = 0.2) track jets
- These are ghost-associated to the large-radius jet in the event
- This allows us to classify the large-R jet into tag regions

Why <100 GeV met cut?





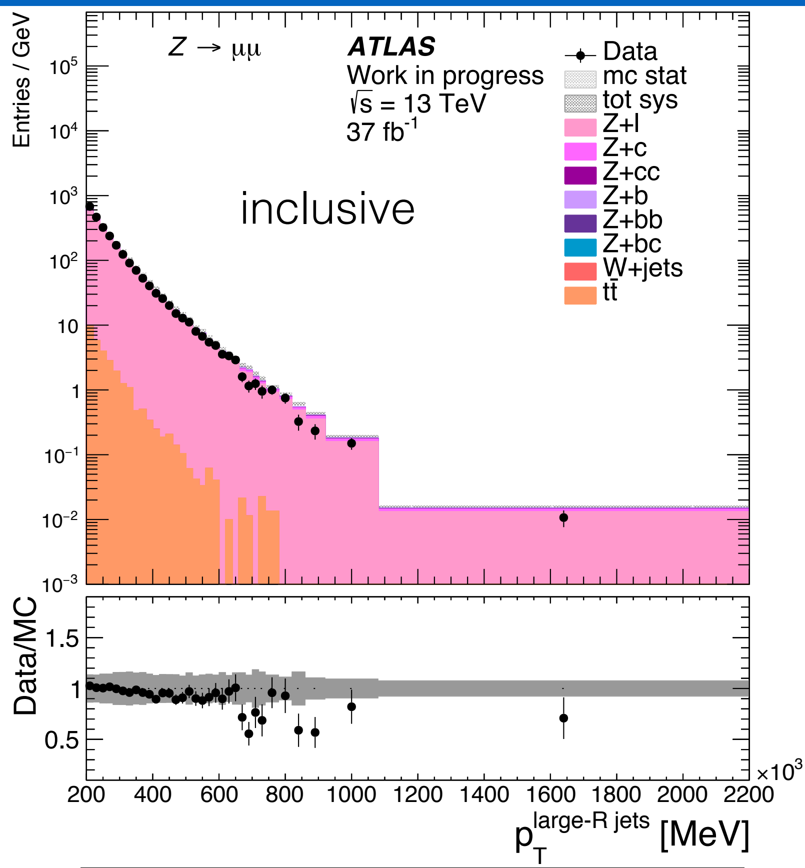
- The significance plateaus at around 100 GeV
- Left plot shows signal and background yields as a function of the met cut, and different Z-mass window cuts
- If we apply a cut of 100 GeV, we can cut the background by a factor of two, whilst losing almost no signal

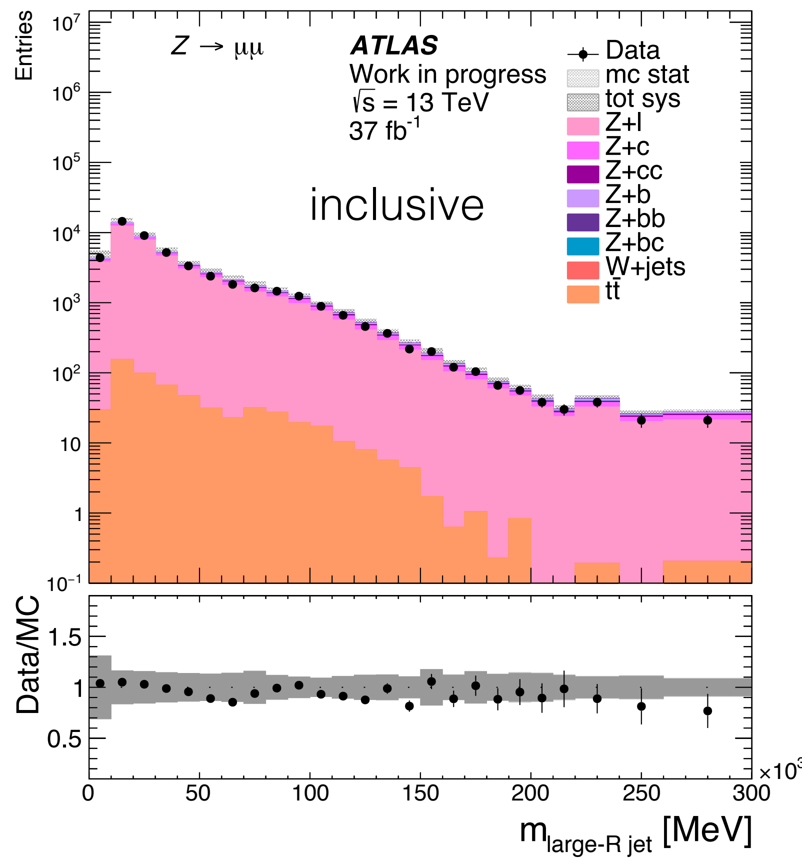
Systematics

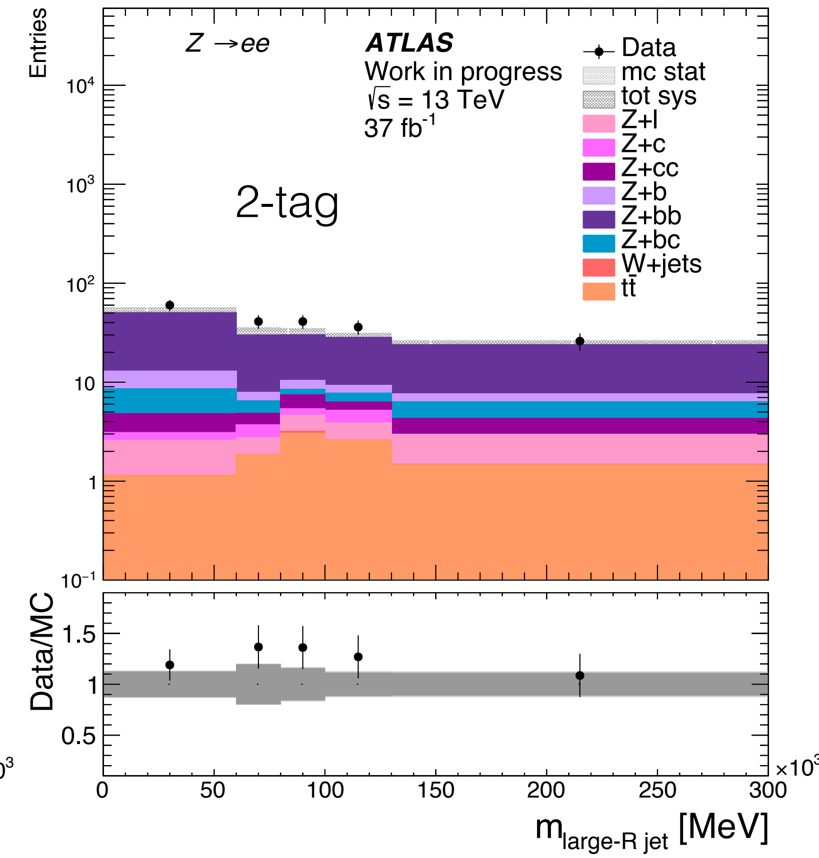
- **Detector-modelling systematics:**
  - B-tagging efficiency
  - Large-R jet energy/mass scale and resolution
  - Lepton-related (ID, reconstruction ….)
  - Met-related
- **Signal-modelling systematics** considered:
  - Scale variations (factorisation and renormalisation)
  - PDF uncertainty
  - CKKW matching scale
    - Only truth-level samples exist for this variation so we will compute this using Rivet
- **Top-modelling systematics** considered:

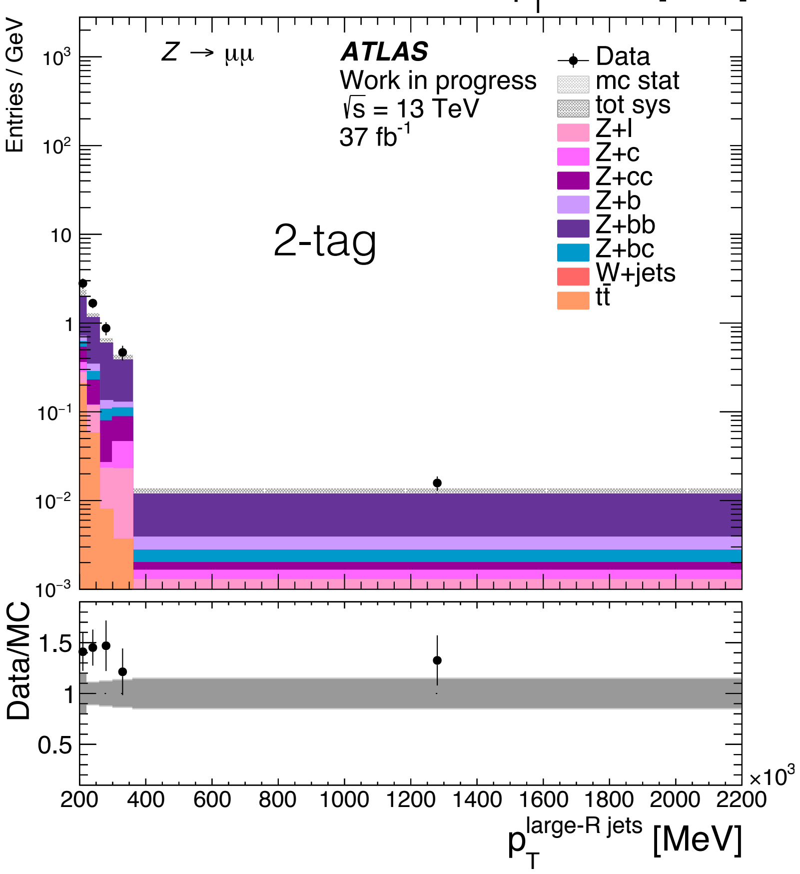
  - Using the usual samples and prescription from TopWG:
  - Rad Hi/Lo
  - Hard scatter generation (aMC@NLO vs Powheg)
  - Parton shower (Pythia 8 vs Herwig++)

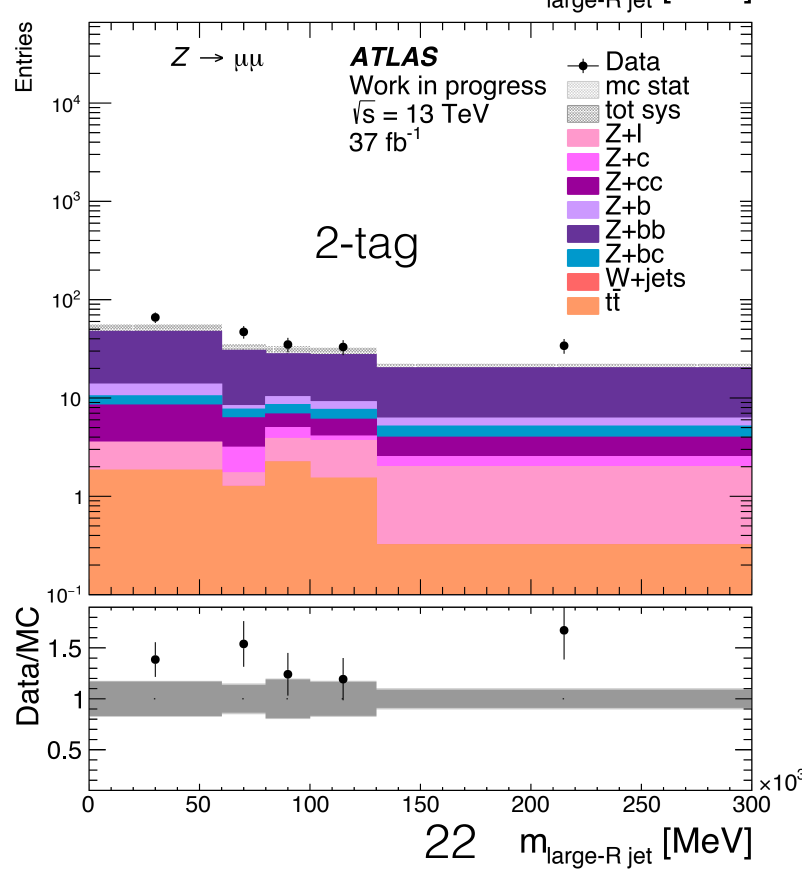
Other variables/channel

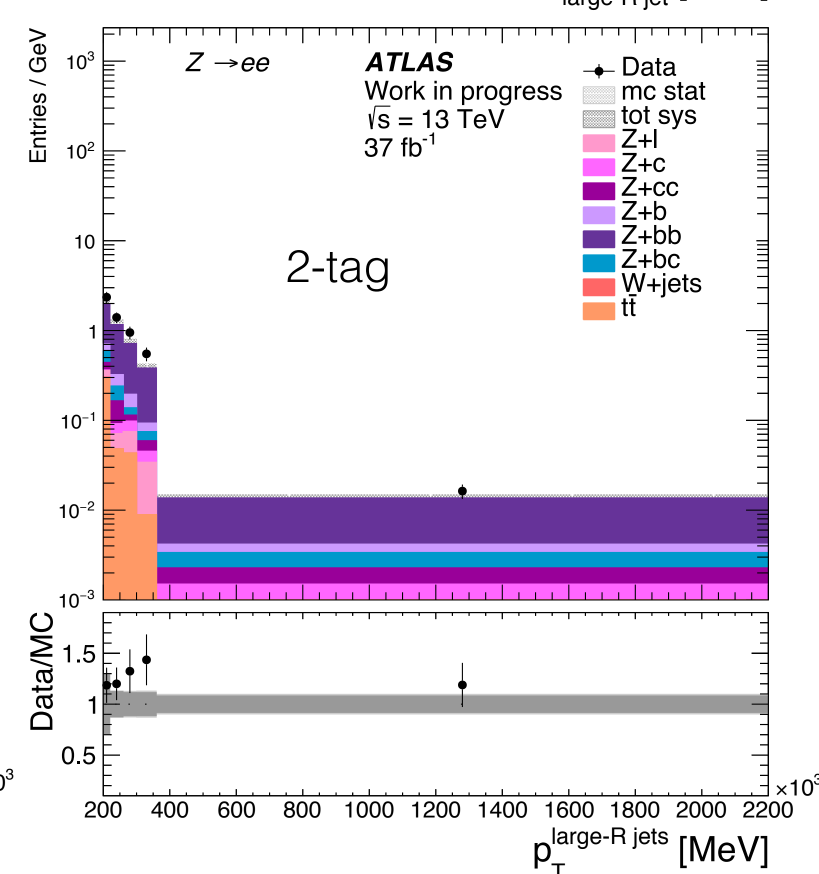












Fully-bayesian unfolding

- Bayes' rule: $$P(\sigma, \Lambda | d) \propto \mathcal{L}(d | \sigma, \Lambda)\ \pi(\Lambda)$$
- here ***Λ*** encodes "nuisance parameters" (e.g. systematic uncertainties) and is subject to our prior beliefs; ***d*** and ***σ*** are the data yields and signal cross sections, respectively.
- The likelihood of the data given a signal spectrum and ***Λ*** is then

$$\mathcal{L}(d | \sigma, \Lambda) = \prod_{i \in \text{recobins}} \text{Poiss}(d_i | x_i(\sigma, \Lambda))$$

$$x_i(\sigma, \Lambda) = L(\Lambda) \times (b_i(\Lambda) + M_{ij}(\Lambda)\ \sigma_j)$$

- where ***x*** is the total number of predicted events in each reco bin, ***L*** is the luminosity, ***b*** is the number of background events, and ***M*** is the migration matrix.
- we then extract the posterior probability of a signal spectrum given the data by sampling points in (***σ***, ***Λ***) space.