Machine learning application for background modelling

ATLAS Higgsino search in $hh \rightarrow 4b$

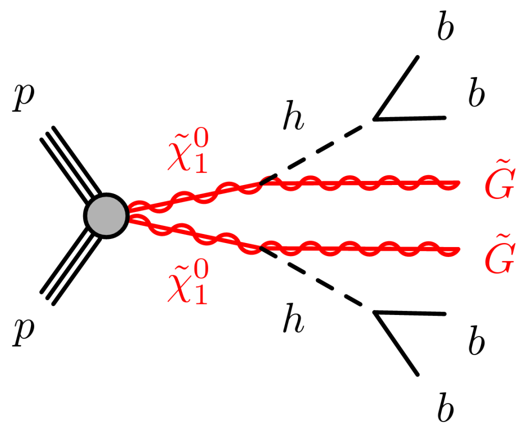
ML4BSM
Durham, UK

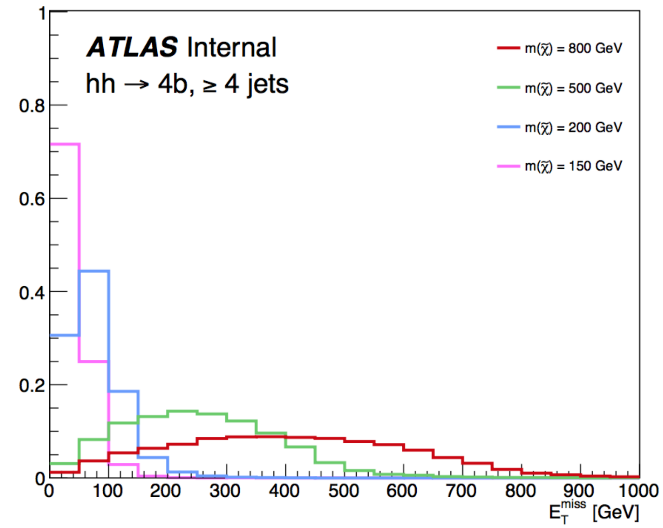
April 5th 2018

Kristian Gregersen (UCL / Lund)

- **BDT Reweighting for background modelling – in the context of a Higgsino search in ATLAS**

- **Target Signal:**
  - hh(→4b) + MET final state





- **Two analyses**
  - Low/High Higgsino mass → Low/High missing transverse energy

Today →

Low-mass
- b-jet triggers
- Targets low signal masses

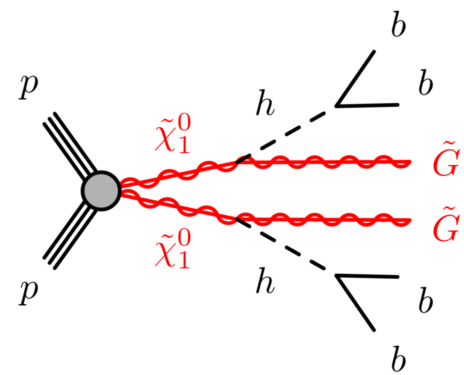
High-mass
- MET trigger
- Targets high signal masses

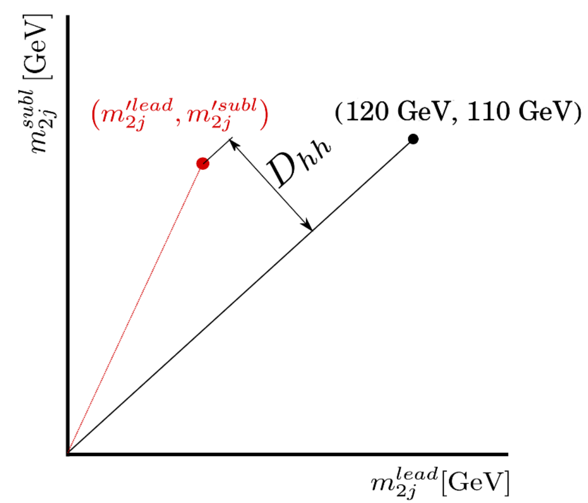
- **Conference Note for SUSY17** : ATLAS-CONF-2017-081

Backgrounds & Event Selection (briefly)

- **Backgrounds**
  - ~95% QCD
  - ~5% leptonic ttbar (→ 50% for MET ~ 150 GeV)
- **Jets**
  - >= 4 jets with pT > 40 GeV
- **Event categories :**
  - 4-tag events : >= 4 jets are b-tagged
  - 2-tag events : == 2 jets are b-tagged
- **Di-Higgs candidates:**
  - Use *Dhh* to resolve pairing ambiguity (3 combinations)
  - Shortest distance to line from (0,0) to (120,110) GeV in the mass plane of the lead/subl pT Higgses.
- **Top Veto:**
  - Leptonic top :
    - Events with electron or muon
  - Hadronic top :
    - Reconstruct top and W boson candidates from three jets
    - 'B-jet' from top decay must be a Higgs candidate jet
    - Any other two jets for W boson candidate
    - Veto events with X_{Wt} < 1.5

$$X_{Wt} = \sqrt{\left(\frac{m_W - 80.4\ \text{GeV}}{0.1 m_W}\right)^2 + \left(\frac{m_t - 172.5\ \text{GeV}}{0.1 m_t}\right)^2}$$





Regions

- **2D plane of higgs masses is used to define regions**
  - Leading pT Higgs vs Subleading pT Higgs

Signal region (SR)

$$X_{hh} = \sqrt{\left(\frac{m_{2j}^{\text{lead}} - 120}{0.1 \times m_{2j}^{\text{lead}}}\right)^2 + \left(\frac{m_{2j}^{\text{subl}} - 110}{0.1 \times m_{2j}^{\text{subl}}}\right)^2} < 1.6$$

Control region (CR)

$$X_{hh} = \sqrt{\left(\frac{m_{2j}^{\text{lead}} - 120}{0.1 \times m_{2j}^{\text{lead}}}\right)^2 + \left(\frac{m_{2j}^{\text{subl}} - 110}{0.1 \times m_{2j}^{\text{subl}}}\right)^2} > 1.6$$

$$R_{hh}^{\text{CR}} \equiv \sqrt{(m_{2j}^{\text{lead}} - 126.0)^2 + (m_{2j}^{\text{subl}} - 115.5)^2} < 55 \text{ GeV}$$

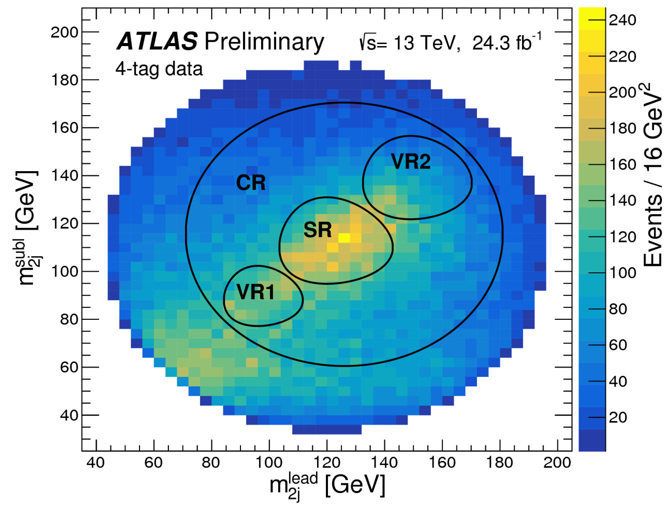
Validation region 1 (VR1)

$$X_{hh}^{\text{VR1}} \equiv \sqrt{\left(\frac{m_{2j}^{\text{lead}} - 96}{0.1 \times m_{2j}^{\text{lead}}}\right)^2 + \left(\frac{m_{2j}^{\text{subl}} - 88}{0.1 \times m_{2j}^{\text{subl}}}\right)^2} < 1.4$$

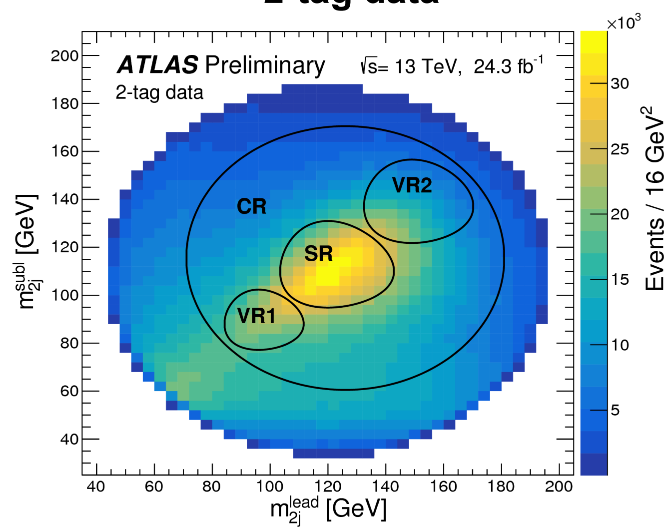
Validation region 2 (VR2)

$$X_{hh}^{\text{VR2}} \equiv \sqrt{\left(\frac{m_{2j}^{\text{lead}} - 149}{0.1 \times m_{2j}^{\text{lead}}}\right)^2 + \left(\frac{m_{2j}^{\text{subl}} - 137}{0.1 \times m_{2j}^{\text{subl}}}\right)^2} < 1.25$$

4-tag data



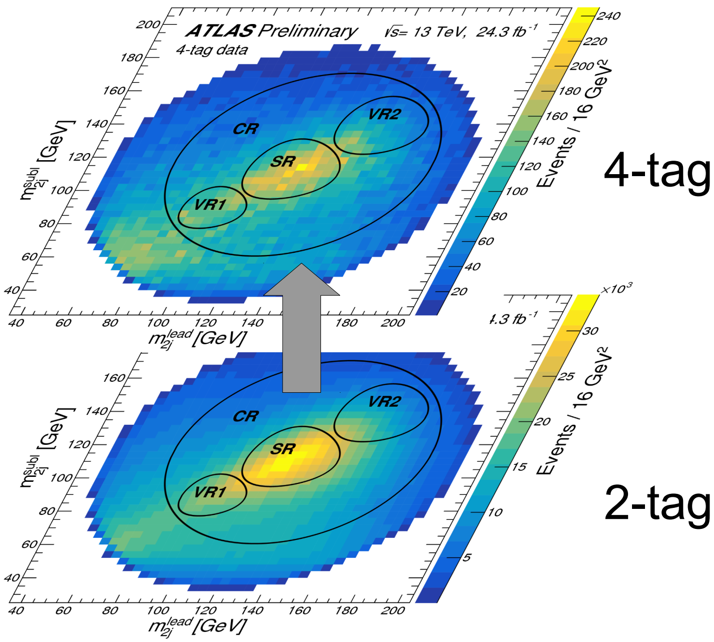
2-tag data



Strategy

- **Background estimation** relies on use of 2-tagged events
- **Reweighting:** Reweight 2-tagged events to look like 4-tagged events

Reweight function: f : 2-tag → 4-tag



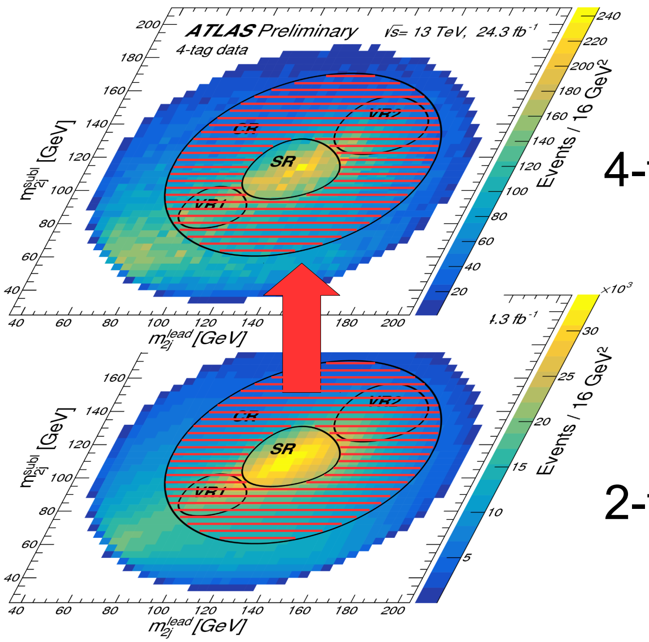
Background estimation

Strategy

- **Background estimation** relies on use of 2-tagged events
- **Reweighting:** Reweight 2-tagged events to look like 4-tagged events

Reweight function: f : 2-tag → 4-tag

Derive on CR data $f_{CR} := (\text{4-tag} / \text{2-tag})_{CR}$



4-tag

2-tag

Background estimation

Strategy

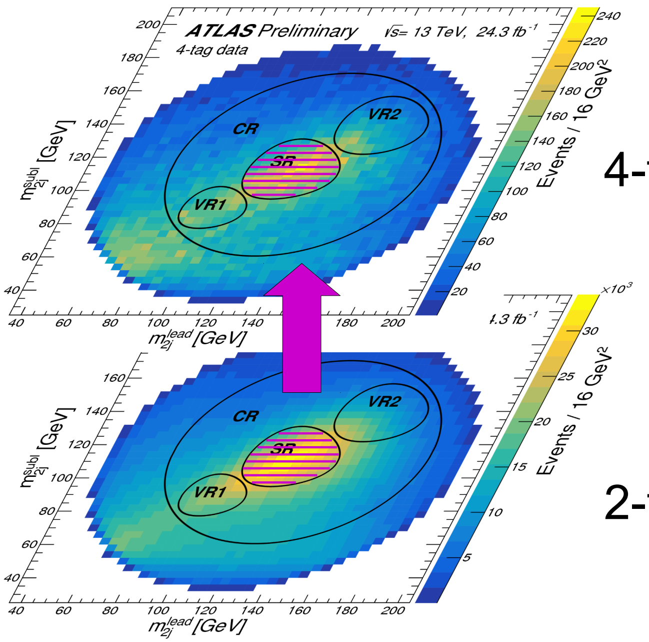
- **Background estimation** relies on use of 2-tagged events
- **Reweighting:** Reweight 2-tagged events to look like 4-tagged events

Reweight function: f : 2-tag → 4-tag

Derive on CR data $f_{CR} = (\text{4-tag} / \text{2-tag})_{CR}$

Apply on SR data $f_{CR}(\text{2-tag}_{SR}) = \text{4-tag}_{SR}$ estimate

This is the BG estimate!



- **Reweight function split into two pieces:**
  - **Normalisation** (about a factor of 200 more 2-tag events than 4-tag events)
  - **Kinematic reweighting** based on **BDTs** derived to correct residual mis-modelling

BDT–based Reweighting

- **Kinematic reweighting** of 2-tagged data → 4-tagged data

- **BDT** has capability to reweight events based on **multiple dimensions**.
  - Splits space of variables in motivated/clever way to capture differences between datasets.
  - Doesn't suffer from 'curse of dimensionalty', always good statistical precision on weights.
  - Divides space of variables into O(30) regions.

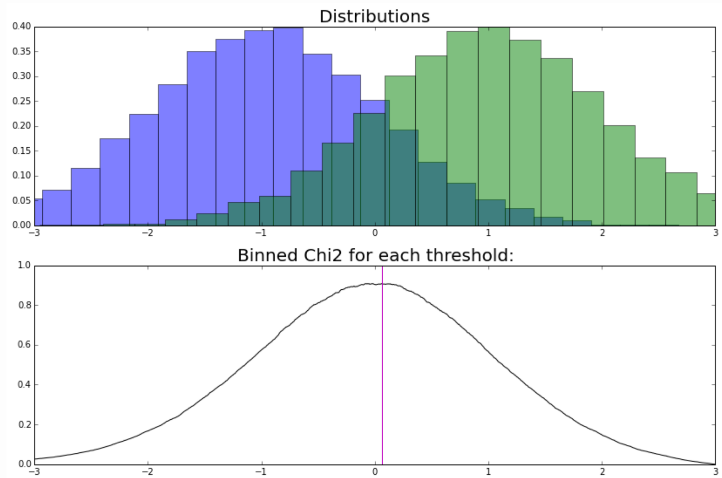
- **Procedure :**
  - For each variable, determine the cut that divides the distribution into two bins which maximises the chi2.
  - Divide sample into two by cutting on the variable with maximum chi2.
  - For each resulting sub-sample, repeat first two steps until some stop-criteria is reached.
  - Each event on a given "leaf" gets the weight :

$$\frac{w_{\text{leaf, target}}}{w_{\text{leaf, original}}}$$

  - This defines one tree. Before repeating whole procedure, apply the weights to the sample.
  - This way, build a forest of trees.
  - The final weight for a given event is the product of contributions from the individual trees.

$$\chi^2 = \sum_{\text{bin}} \frac{(w_{\text{bin, original}} - w_{\text{bin, target}})^2}{w_{\text{bin, original}} + w_{\text{bin, target}}}$$

w	=	sum of event weights
original	=	2-tagged
target	=	4-tagged



See full discription here: link

BDT–based Reweighting

- Variables :
  - Determined by adding one variable at the time, in a “greedy” fashion.
  - The single best was added first, by checking how well a BDT could distinguish the reweighted 2-tagged data from the 4-tagged data.
  - This was repeated, addding more variables, until convergence was reached.
  - Variables:

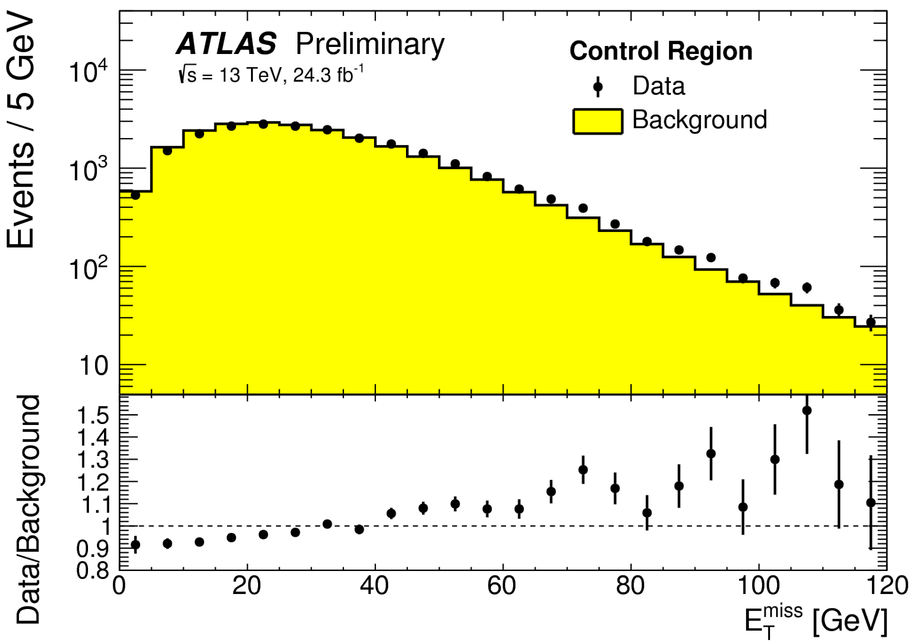
    - The number of jets in the event.
    - p_{T} of each of the four jets in the Higgs candidates.
    - η of each of the four jets in the Higgs candidates.
    - The ΔR separation of the closest two of the four jets in the Higgs candidates.
    - The ΔR separation of the other two of the four jets in the Higgs candidates.
    - For each of the four jets in the Higgs candidates, the scalar sum of track-jet mass over the scalar sum of track-jet p_{T}, for track-jets matched to the jet.
    - Which of each Higgs candidate’s jets are b-tagged. (For the four-tag selected events, two of the tagged jets are randomly identified as "untagged".)
    - The number of track-jets matched to each Higgs candidate.
    - ΔR separation of the jets in each Higgs candidate.
    - p_{T} of each Higgs candidate.
    - Mass of di-Higgs system.
    - $\Delta\eta$ separation of the two Higgs candidates.
    - $E_{\mathrm{T}}^{\mathrm{miss}}$.
    - Variable sensitive to all-hadronic top decays: X_t.

BDT–based Reweighting

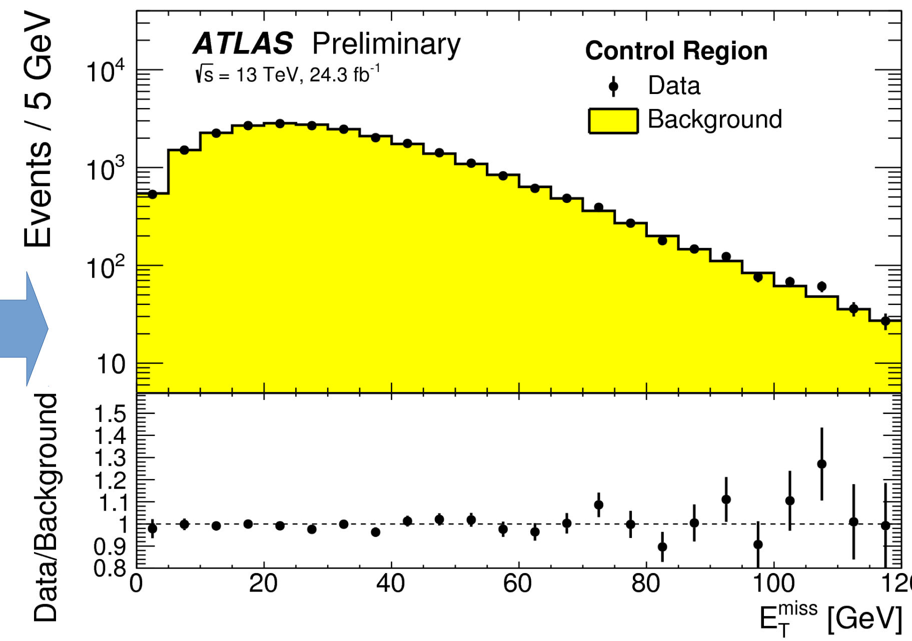
- Hyperparameters :
  - Optimised in a way similar to identifying the variables.
  - A BDT was trained to distinguish the reweighted 2-tag data from the 4-tag data.
  - The configuration of hyperparameters for which this BDT did worst, was chosen.
  - Very similar performance for most configurations, no clear optimal choice.
  - Parameters:
    - Minimum #events on leafs : 250
    - Maximum #layers : 5
    - Number of trees : 30
    - Learning rate : 0.7
    - Sampling fraction : 0.7

- **Control Region Plots**
  - Missing transverse momentum before/after BDT reweighting
  - “Data” is 4-tag events
  - “Background” is reweighted 2-tag events

Before BDT

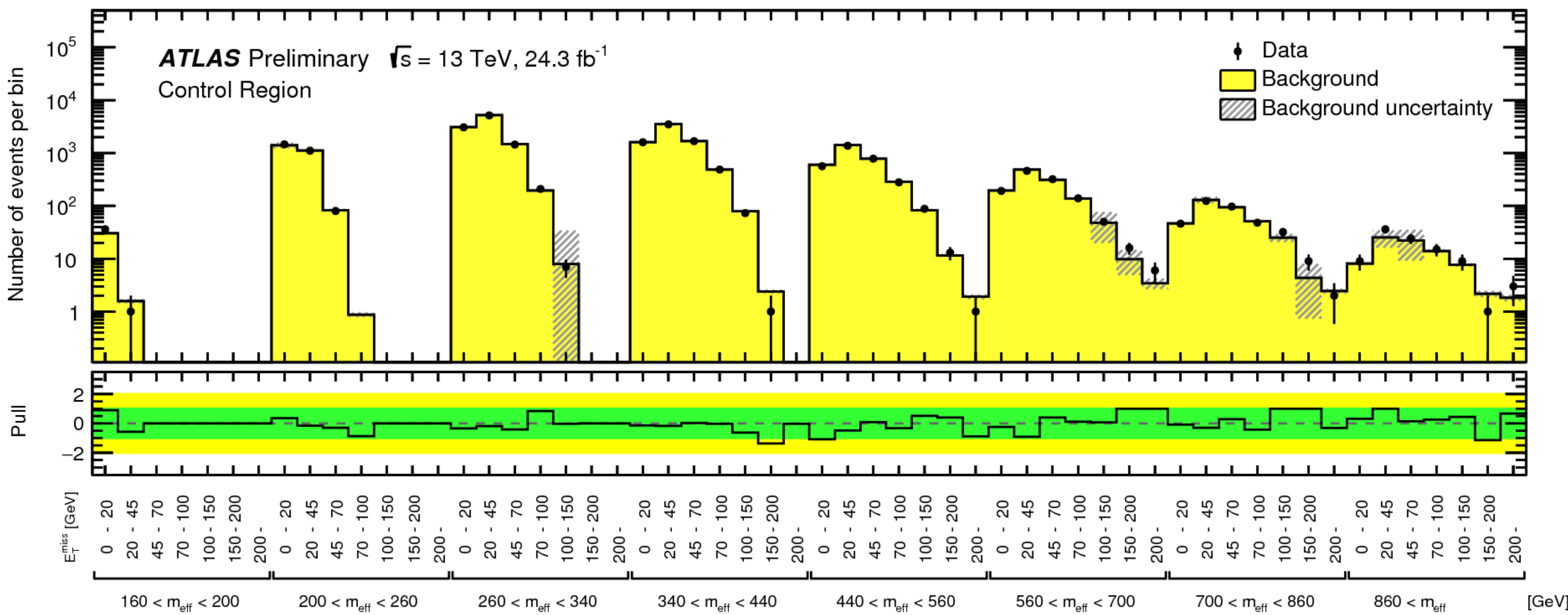


After BDT



- **Control Region Plots**
  - 'Unrolled' 2D distribution of missing transverse momentum vs. the effective mass
  - (Discriminant in statistical analysis)

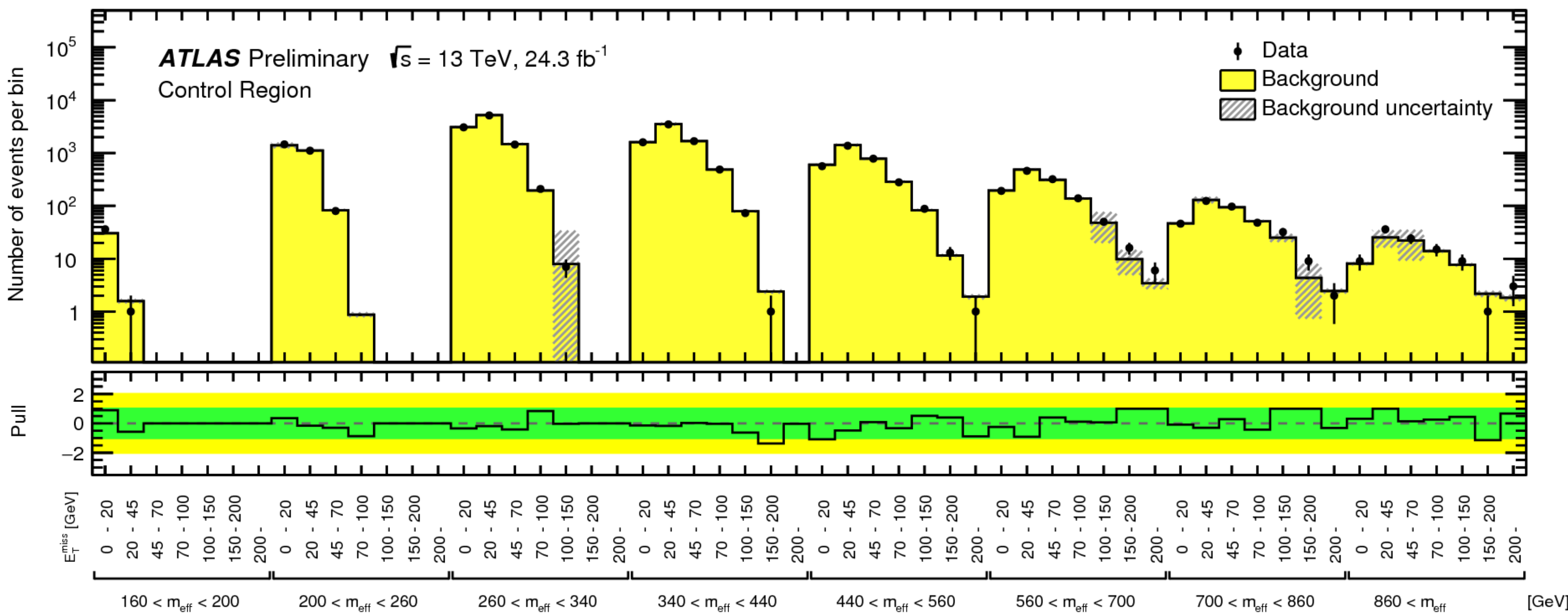
$$m_{\mathrm{eff}}^{4b\mathrm{-jets}} = \sum_{i=1,..,4} p_{\mathrm{T}}^{j_i} + E_{\mathrm{T}}^{\mathrm{miss}}.$$



Background estimation

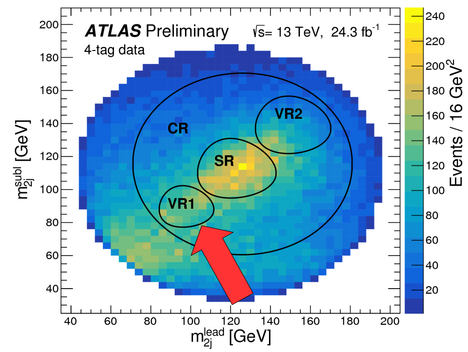
- **Background modelling uncertainties**
  - Non-closure
  - Transfer of weights
  - Statistical uncertainty in 2-tag SR

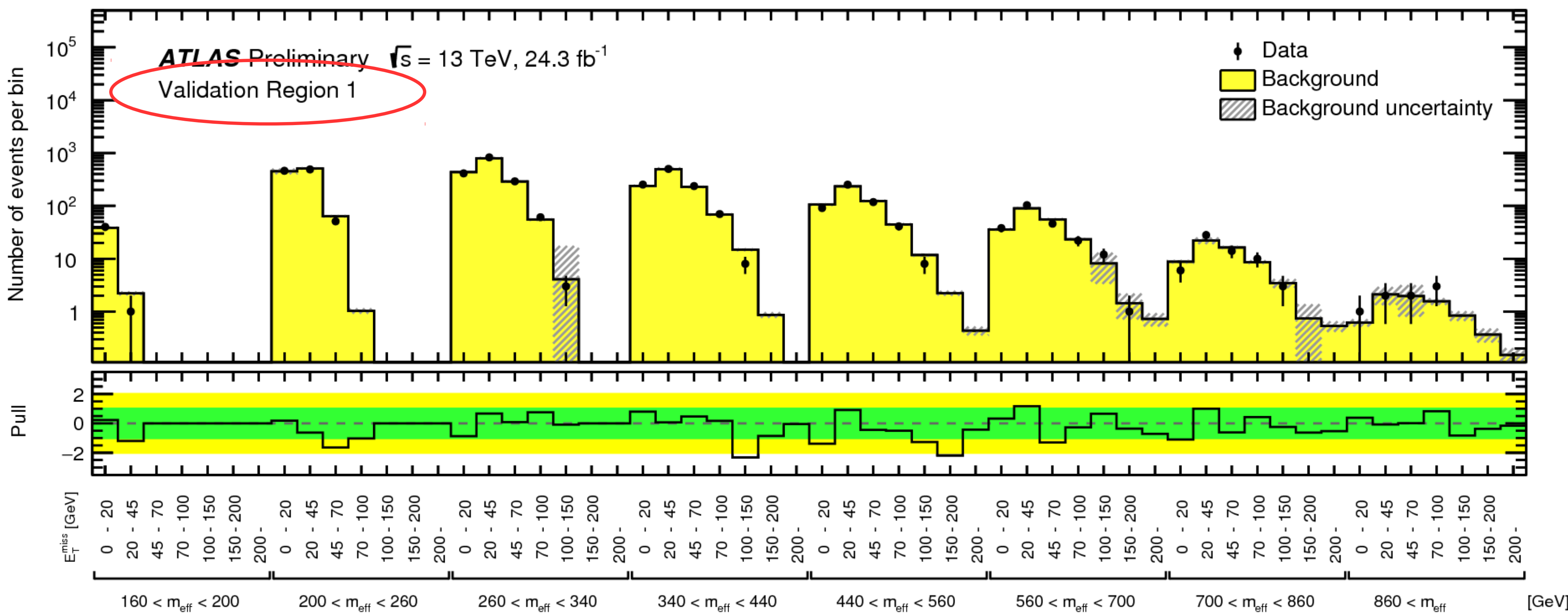
- **Background modelling uncertainties**
  - **Non-closure** → bin-by-bin differences between data and background in CR
  - Transfer of weights
  - Statistical uncertainty in 2-tag SR



Background estimation

- **Background modelling uncertainties**
  - Non-closure
  - **Transfer of weights** → norm./shape differences in the validation regions
  - Statistical uncertainty in 2-tag SR



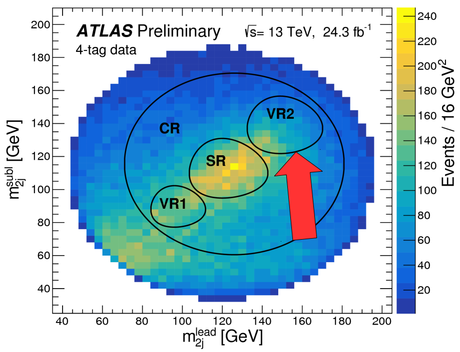


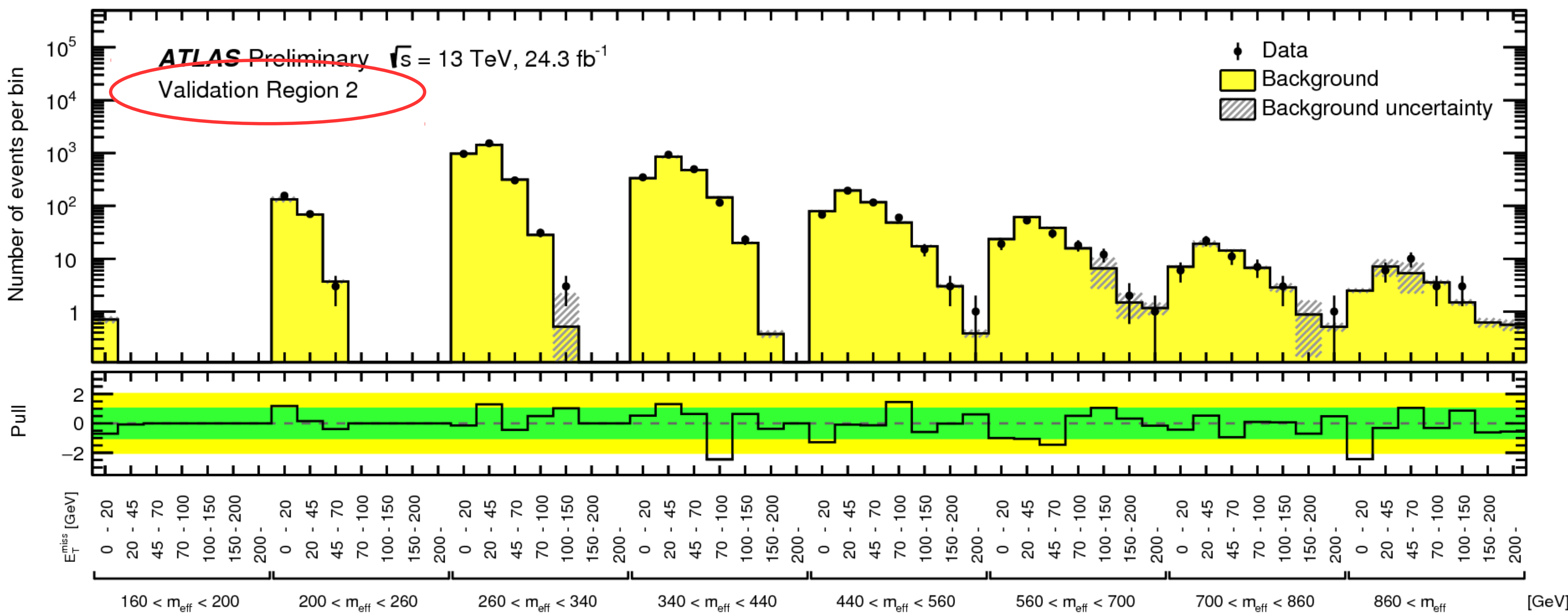
- New reweighting functions were derived for this check, excl. the VR under consideration:

  $f_{CR \setminus \{VRi\}}(\text{2-tag}_{VRi}) = \text{4-tag}_{VRi}$ estimate

Background estimation

- **Background modelling uncertainties**
  - Non-closure
  - **Transfer of weights** → norm./shape differences in the validation regions
  - Statistical uncertainty in 2-tag SR





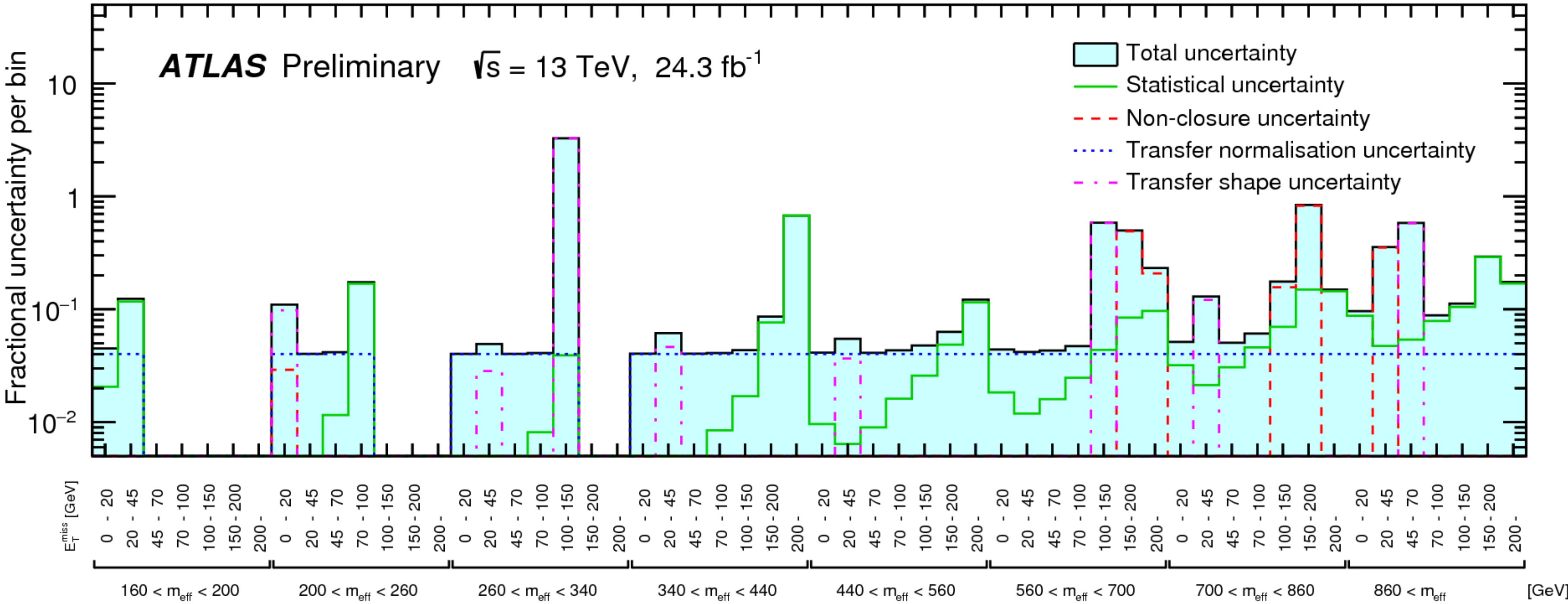
- New reweighting functions were derived for this check, excl. the VR under consideration:

  $f_{CR \setminus \{VRi\}}(2\text{-tag}_{VRi}) = 4\text{-tag}_{VRi}$ estimate

Background estimation

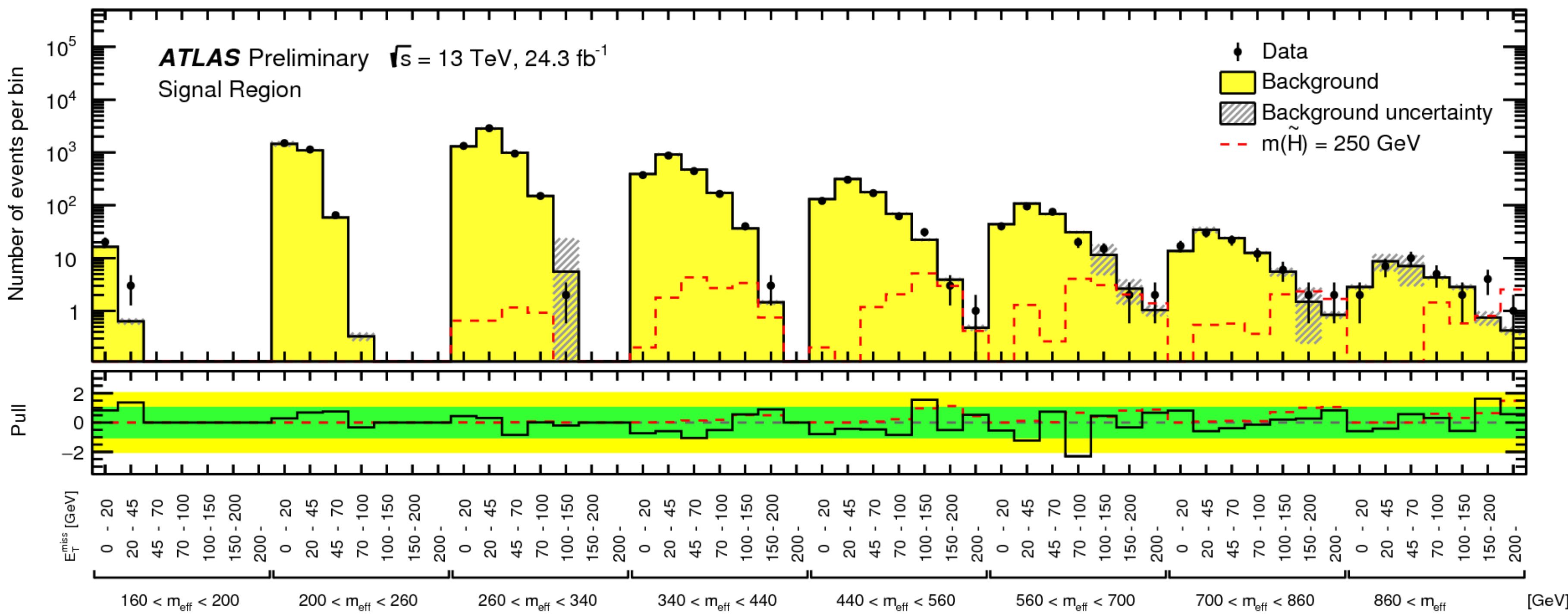
- **Background modelling uncertainties**
  - Non-closure
  - Transfer of weights
  - **Statistical uncertainty in 2-tag SR**

Total uncertainty



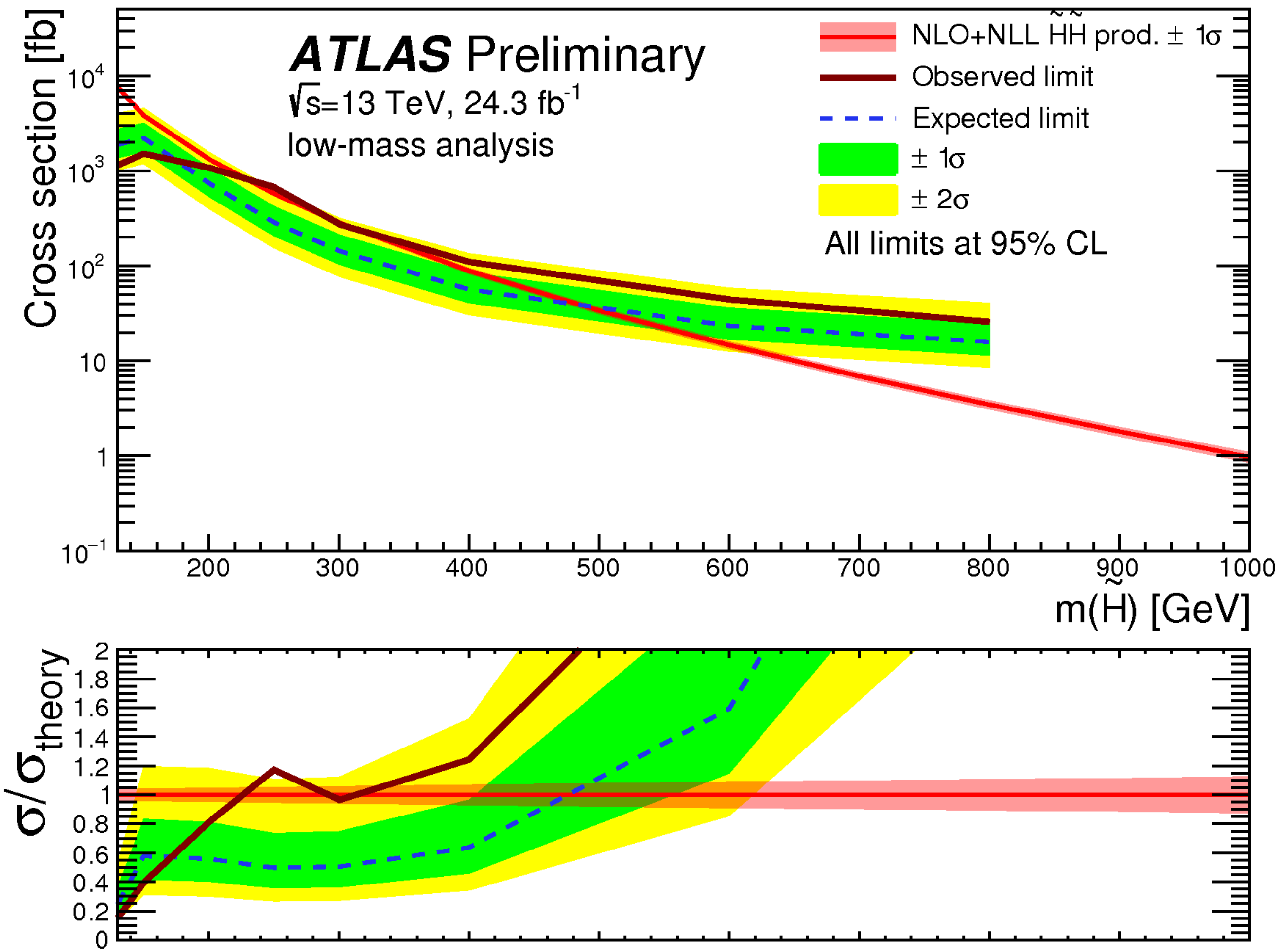
- **Signal Region Plot**
  - 'Unrolled' 2D distribution of missing transverse momentum vs. the effective mass
  - (Discriminant in statistical analysis)

$$m_{\mathrm{eff}}^{\mathrm{4b-jets}} = \sum_{i=1,..,4} p_{\mathrm{T}}^{j_i} + E_{\mathrm{T}}^{\mathrm{miss}}.$$

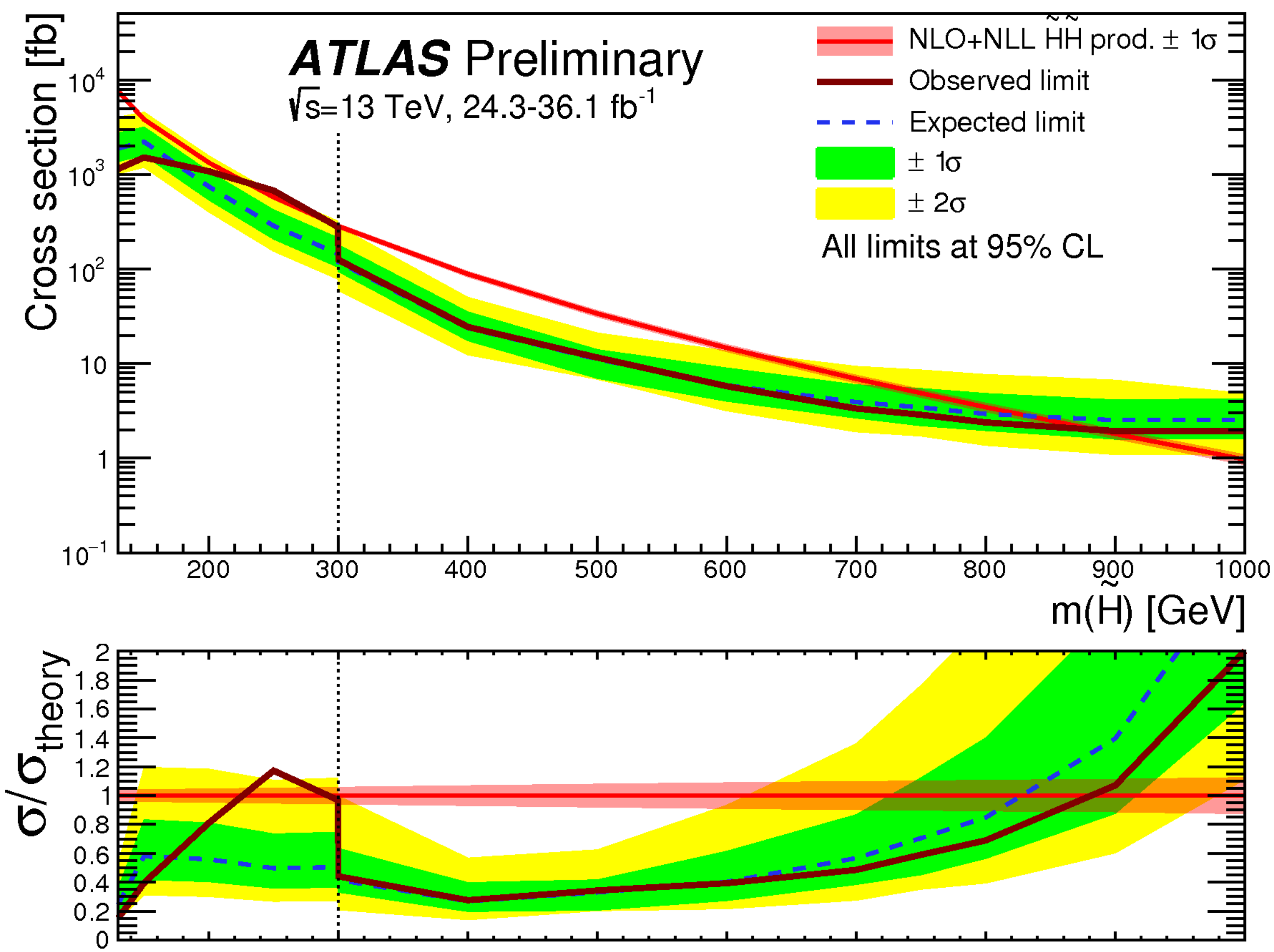


No signs of new physics :-(

- **Exclusion limits – low-mass analysis**

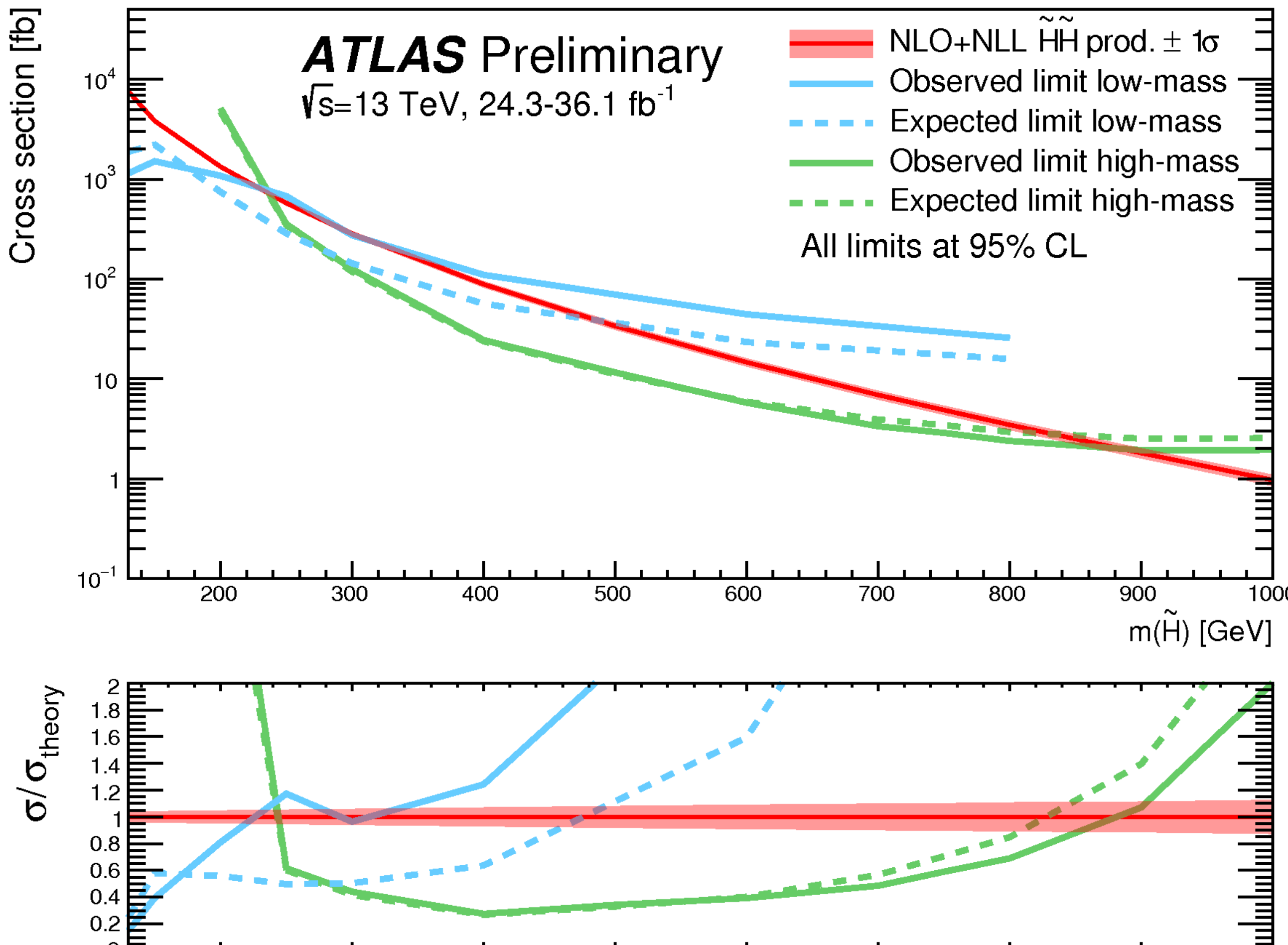


- **Exclusion limits – combined with high-mass analysis**



Interpretation

- **Exclusion limits – separately for high/low-mass analyses**



Ideas for next iteration

- **Can we explore machine learning techniques further for reweighting ?**

  - **Bootstrap Aggregation (bagging)**
    - Build multiple models based on sub-set of the data
      - Randomly draw percentage of data (with replacement), typically 60-70%
      - Train algorithm on each “bag” of data separately
    - Final model is the average of the individual models
      - Each event weight is the average of the weights from all models
      - Gives a handle on the uncertainty in the final weight → the spread in weights!

  - **Random Forests**
    - Based on Decision Trees and Bagging
    - Trees independent of each other
      - Each tree uses random sub-set of data (bagging)
      - AND random sub-set of variables (feature-bagging)
    - Final weight is the average of the weights from all trees
    - Can be grown in parallel! (Contrary to BDTs)

  - **Extremely Randomised Trees (ExtraTrees)**
    - Same as Random Forest, except
      - Nodes are split by randomly choosing variable and splitting point
      - This is contrary to Random Forests and BDTs which split the nodes in a “greedy” fashion.

All of the above are currently under investigation – stay tuned for more !!

Back ups