Progress towards a more sensitive CWoLa hunt with the ATLAS detector

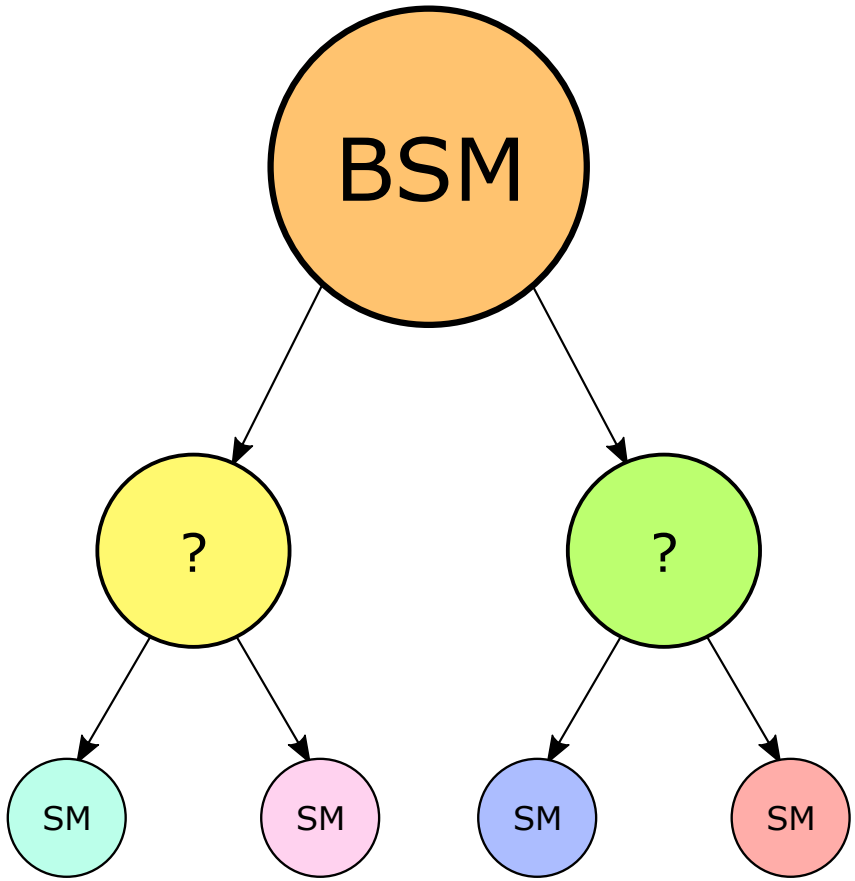
Kees Benkendorfer (Reed College)

DPF2021

12 July 2021

Model-agnostic searches

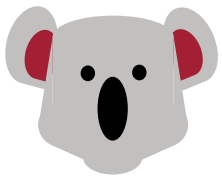
- Wide range of model-specific searches
- Even wider range of possible signal models
- Model-agnostic searches can dramatically expand search 'width'



Data used

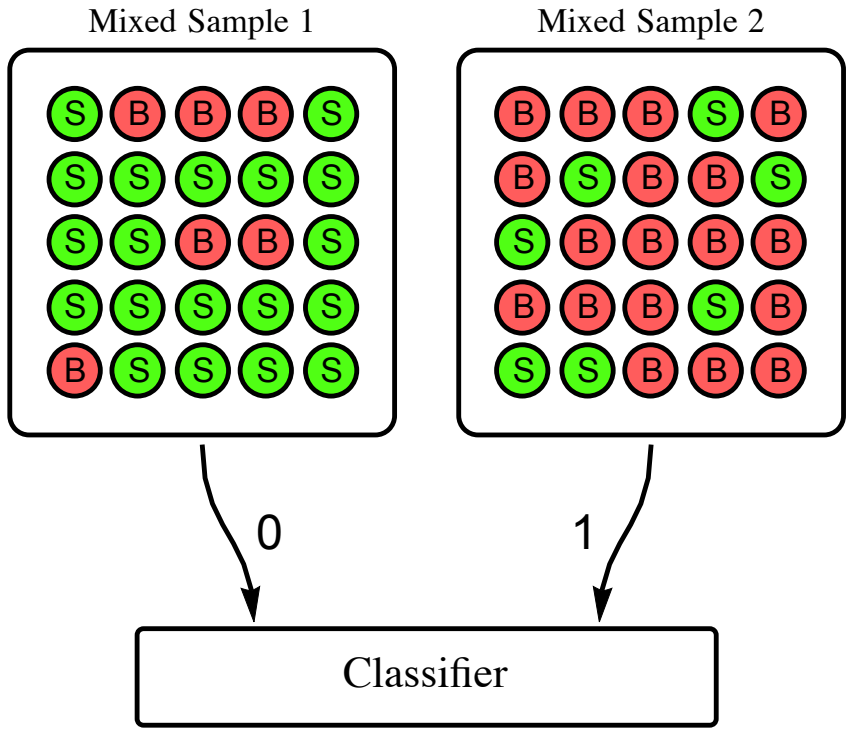
- High p_T dijets from ATLAS Run 2 data
  - 139 fb^{-1}
- Blinding:
  - Considering events with $|\Delta\eta| > 1.2$
  - Only looking at 10% of the above
- Simulation is Pythia

CWoLa overview



CWoLa

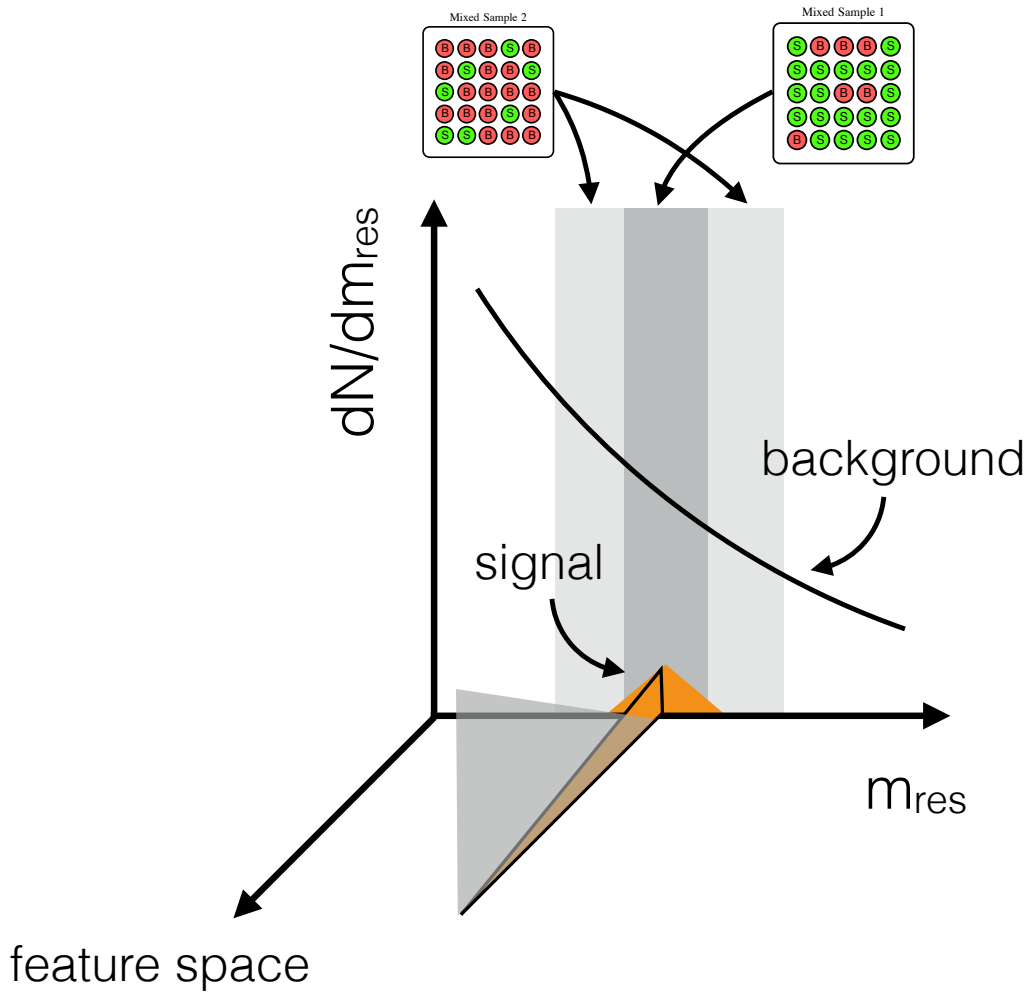
Classification Without Labels



- Classifier learns likelihood ratio
- See arXiv: 1708.02949, 1902.02634

Standard CWoLa hunt

- Previously performed with ATLAS Run 2 data
  - Looking at dijet events
- **Resonant feature**: dijet invariant mass
- **Classification feature**: jet masses
- See arXiv: 2005.02983

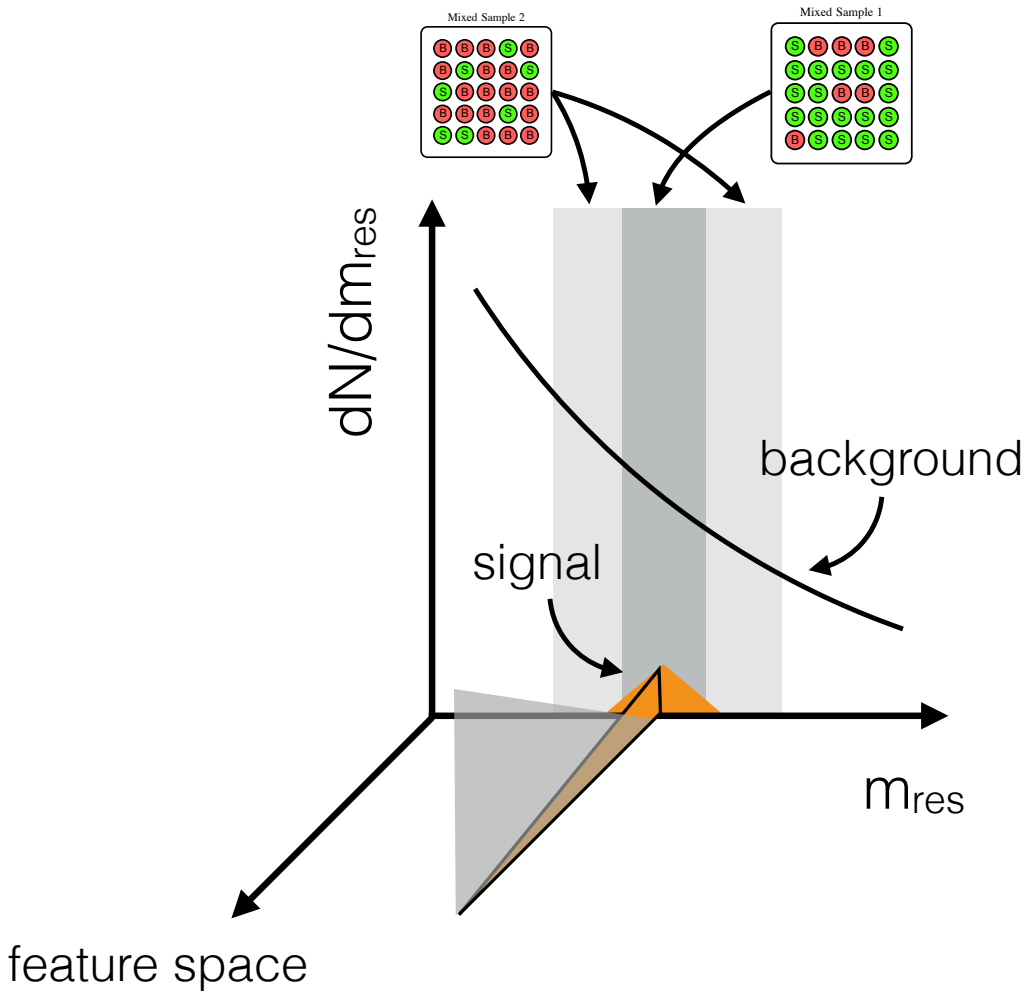


Standard CWoLa hunt

- Previously performed with ATLAS Run 2 data
  - Looking at dijet events
- **Resonant feature**: dijet invariant mass
- **Classification feature**: jet masses
- See arXiv: 2005.02983

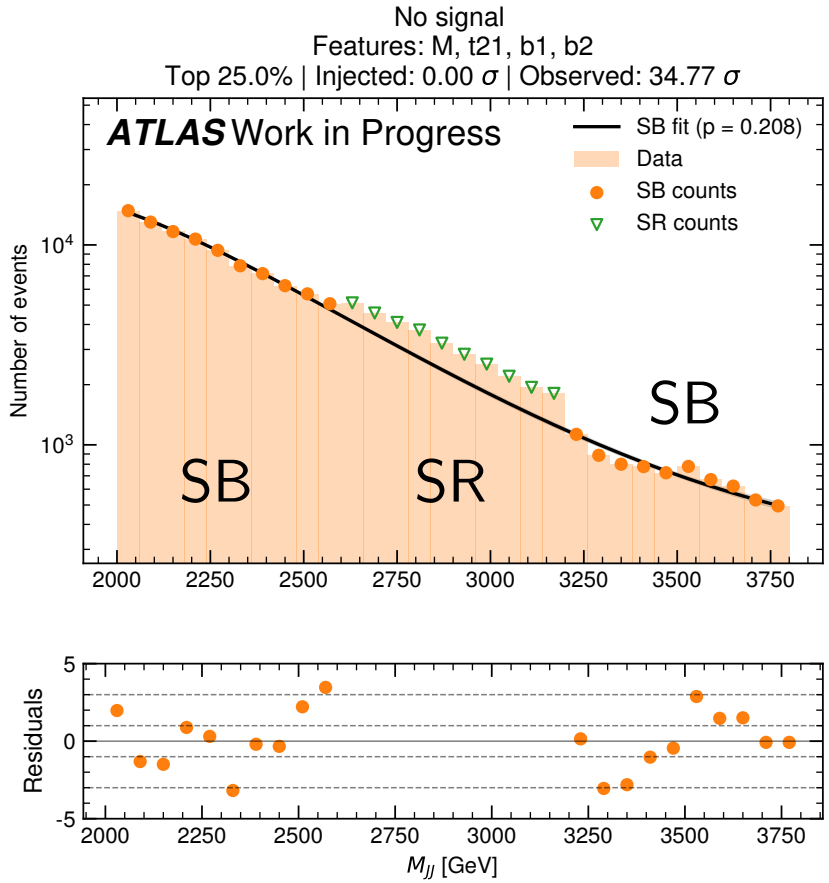
Our Goal

Achieve higher sensitivity with more classification features

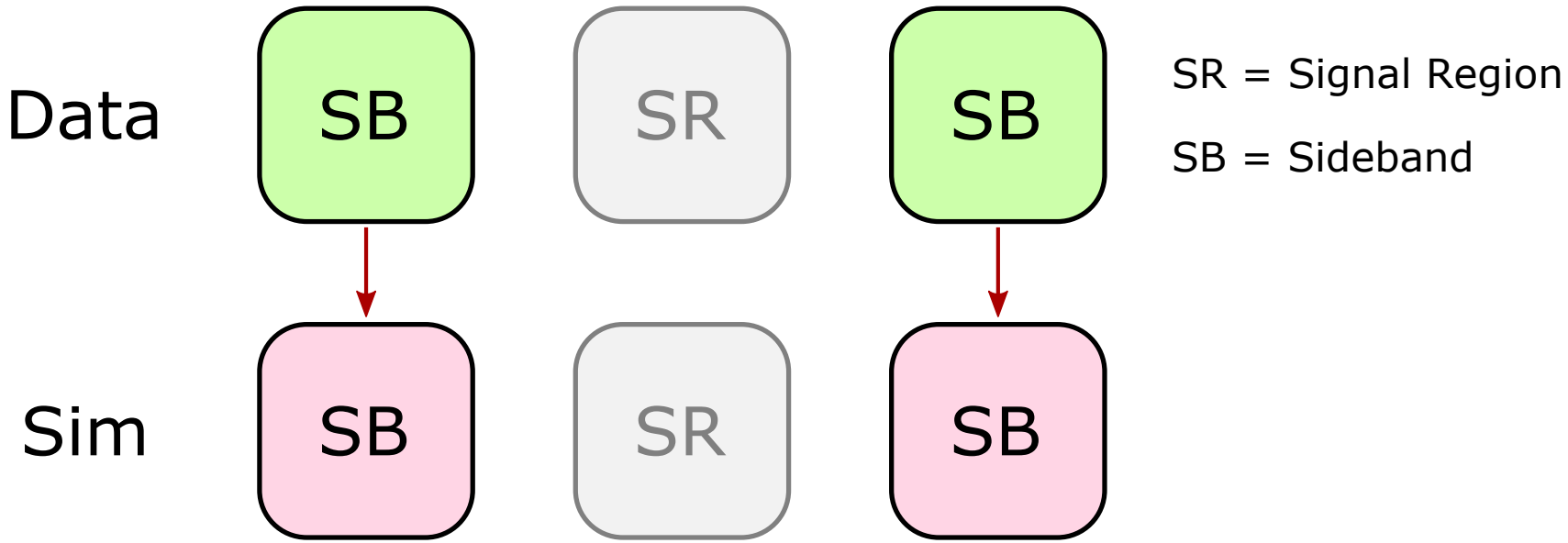


Problems with CWoLa

- Signal region (SR) and sideband (SB) have different M_{JJ}
- More classification features $\rightarrow$ CWoLa infers M_{JJ}
- CWoLa tags entire signal region as signal
- **Result**: large false positive and low sensitivity to real signal
  - Previously avoided with statistical decorrelation

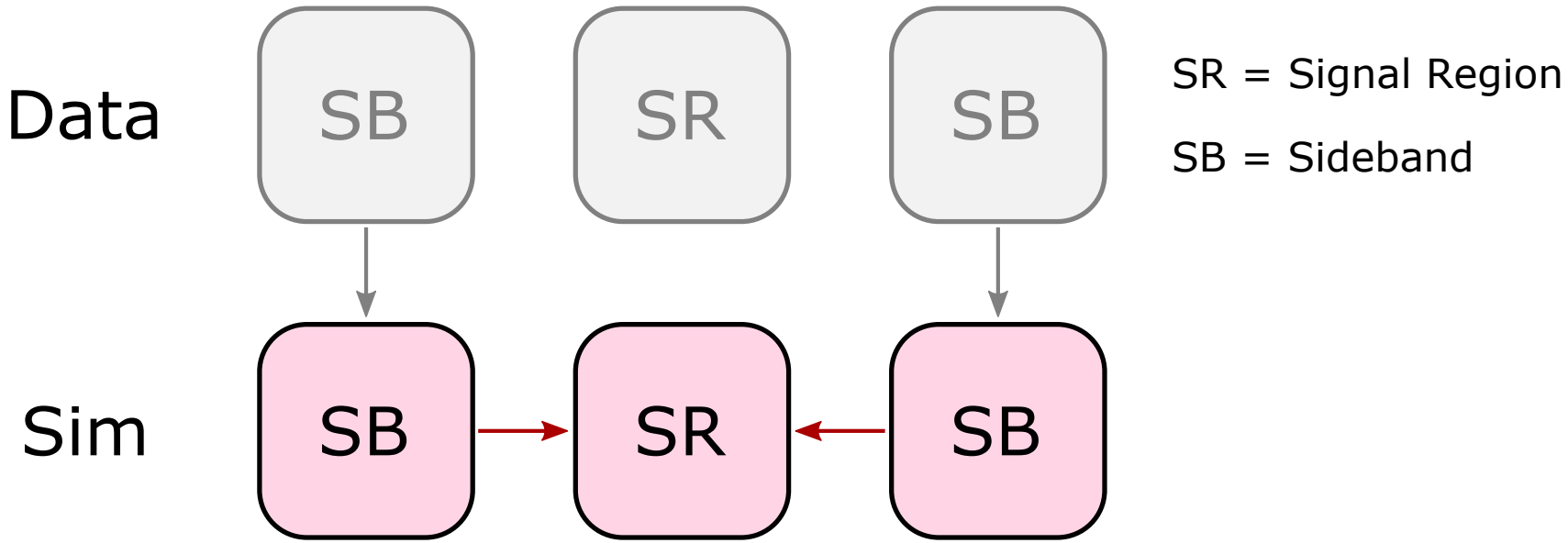


Step 1: learn imperfections in simulation



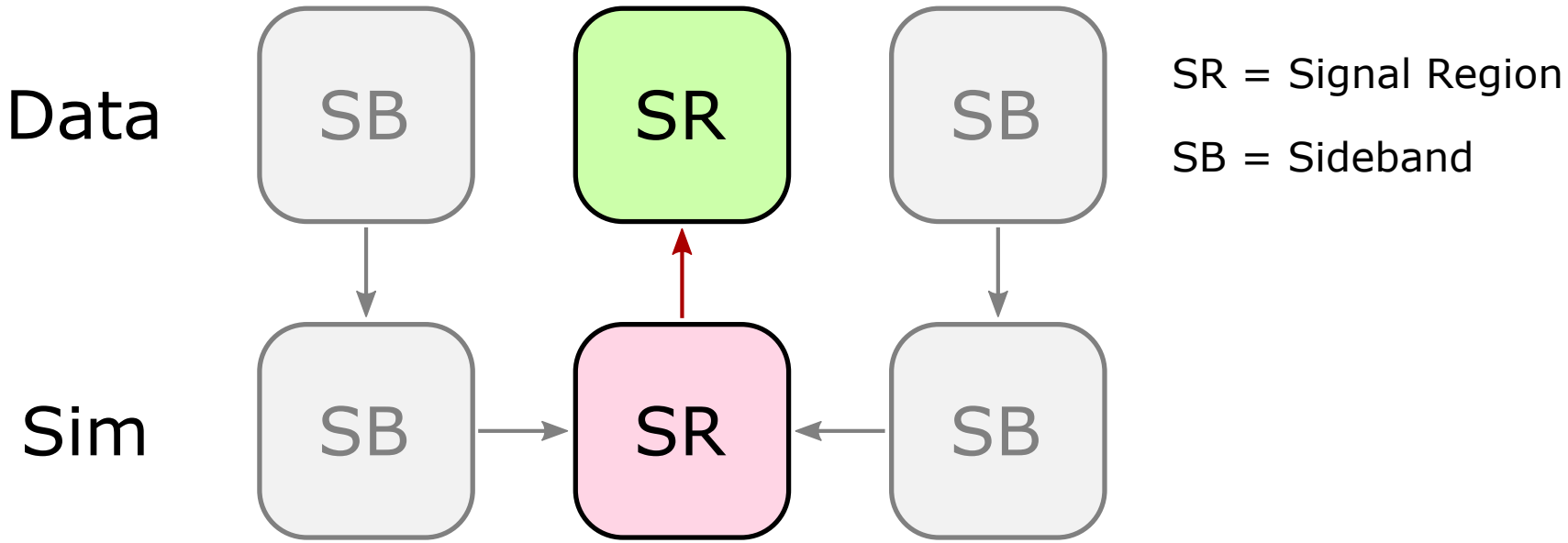
- **SALAD**: **S**imulation **A**ssisted **L**ikelihood-free **A**nomaly **D**etection
- See arXiv: 2001.05001

Step 2: reweight SR in simulation



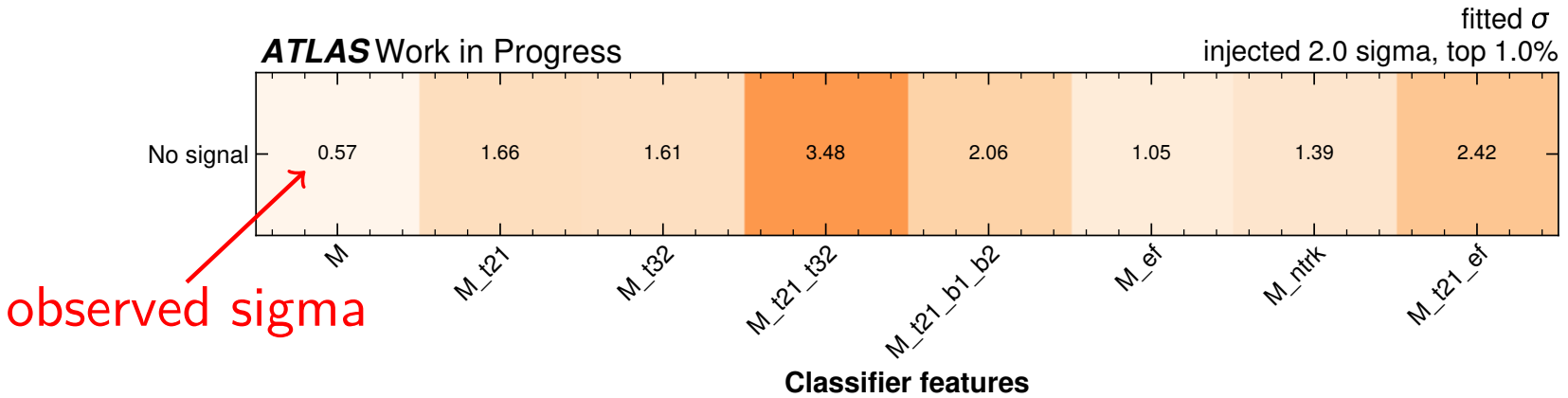
- **SALAD**: **S**imulation **A**ssisted **L**ikelihood-free **A**nomaly **D**etection
- See arXiv: 2001.05001

Step 3: learn simulation vs data in SR



- **SALAD**: **S**imulation **A**ssisted **L**ikelihood-free **A**nomaly **D**etection
- See arXiv: 2001.05001

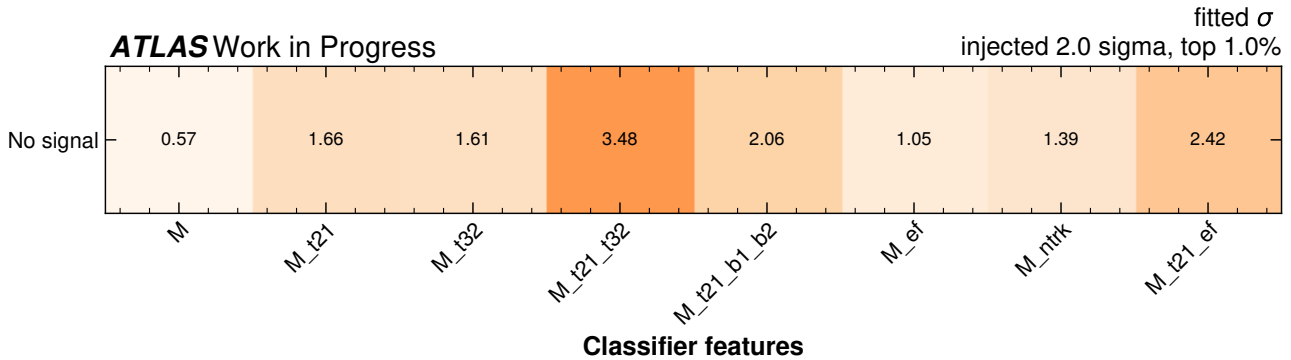
Results without signal



- Colors and values are fitted significance
  - Fit in sideband according to $p_1(1-m)^{p_2} m^{-p_3}$
  - Interpolate to signal region
- Features used:
  - M: jet mass
  - τ_{21}: N-subjettiness ratio τ_2/τ_1
  - τ_{32}: N-subjettiness τ_3/τ_2
  - b_1, b_2: b tagging on first and second subjets
  - EF: fraction of jet energy lost to EM calorimeter
  - N tracks: number of jet tracks

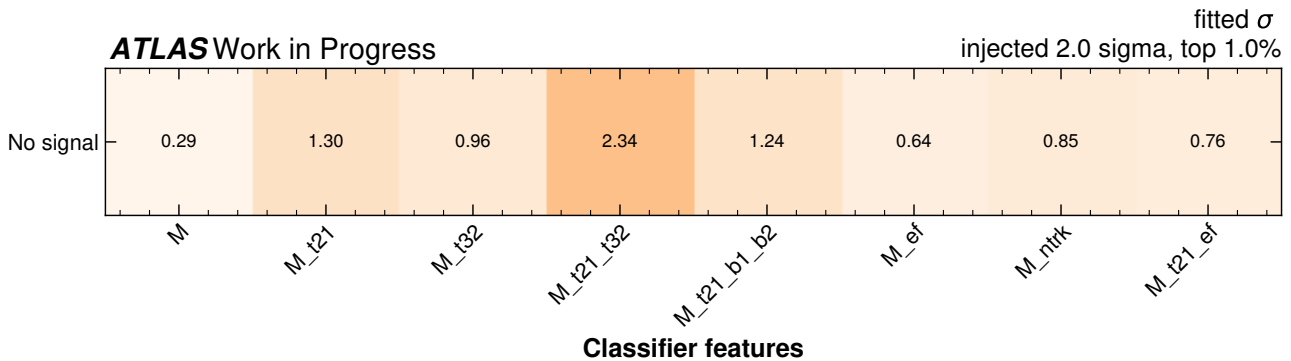
Strategic confusion

- Standard SALAD without signal:



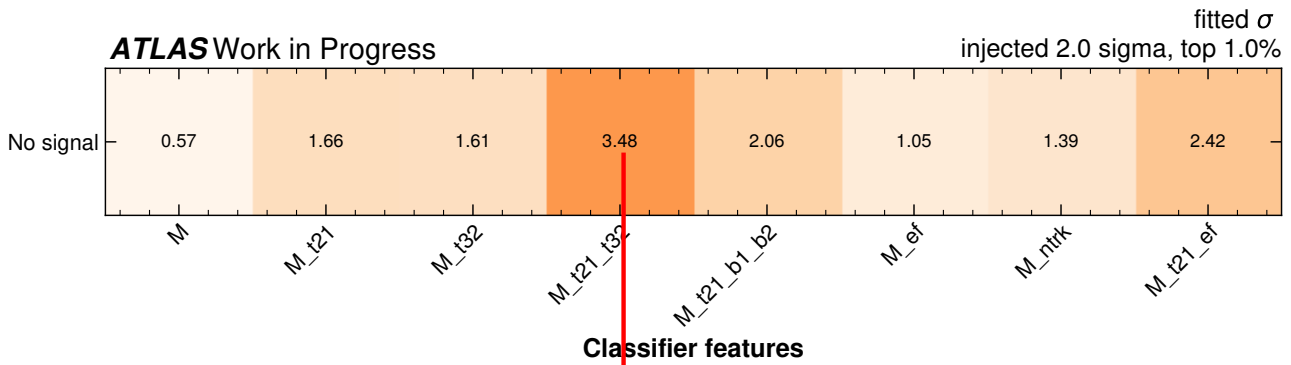
- Injecting some sideband events into signal region can reduce false positive significance

- With 25% of sideband in SR:



Strategic confusion

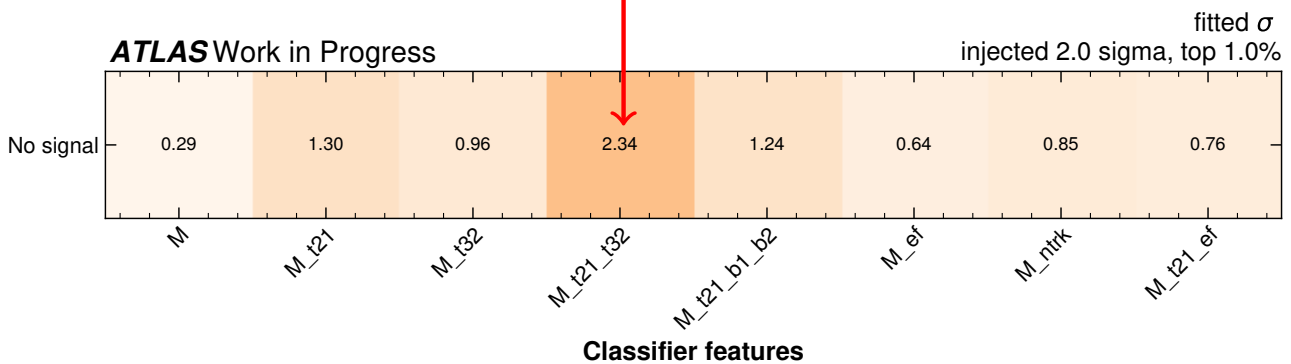
- Standard SALAD without signal:



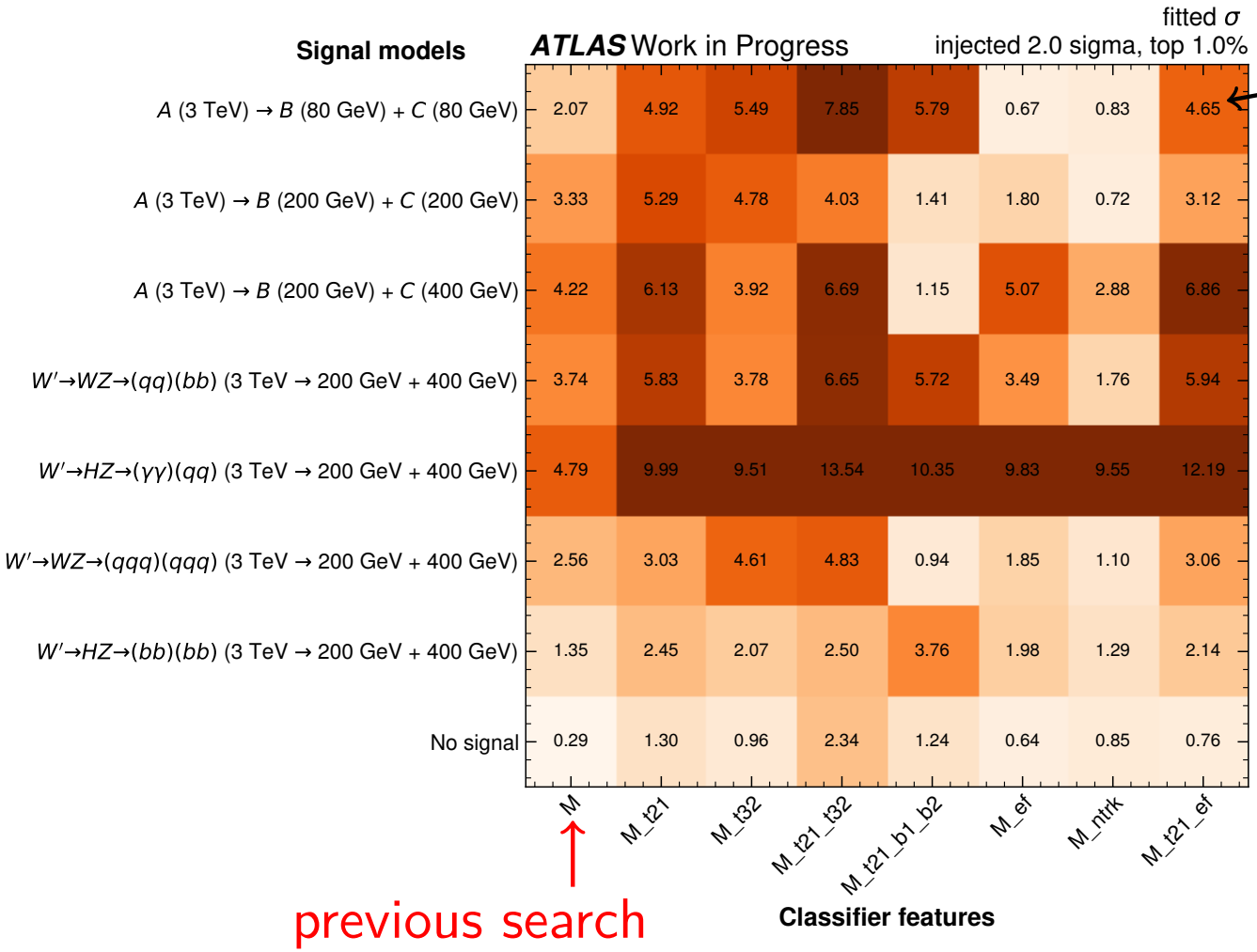
- Injecting some sideband events into signal region can reduce false positive significance

reduced false positive

- With 25% of sideband in SR:



Sensitivity to 2σ signal



- Search strategy can take 2σ excess to $>5\sigma$ result

Conclusion

- Making progress towards a more sensitive model-agnostic search in ATLAS Run 2 data
- More classifier features $\rightarrow$ wider search
- Simulation assistance (SALAD) $\rightarrow$ more robust search
- **Next steps**: need to correct for classification bias

- **Acknowledgements**: I am very grateful to Benjamin Nachman and Luc Le Pottier for their support, guidance, and collaboration