$WW \rightarrow 2l2\nu$

CROSS-SECTION ANALYSIS WITH RANDOM FOREST

THOTH GUNTER, MICHAEL SCHMITT

NORTHWESTERN UNIVERSITY

APR. 8, 2018

ABOUT ME

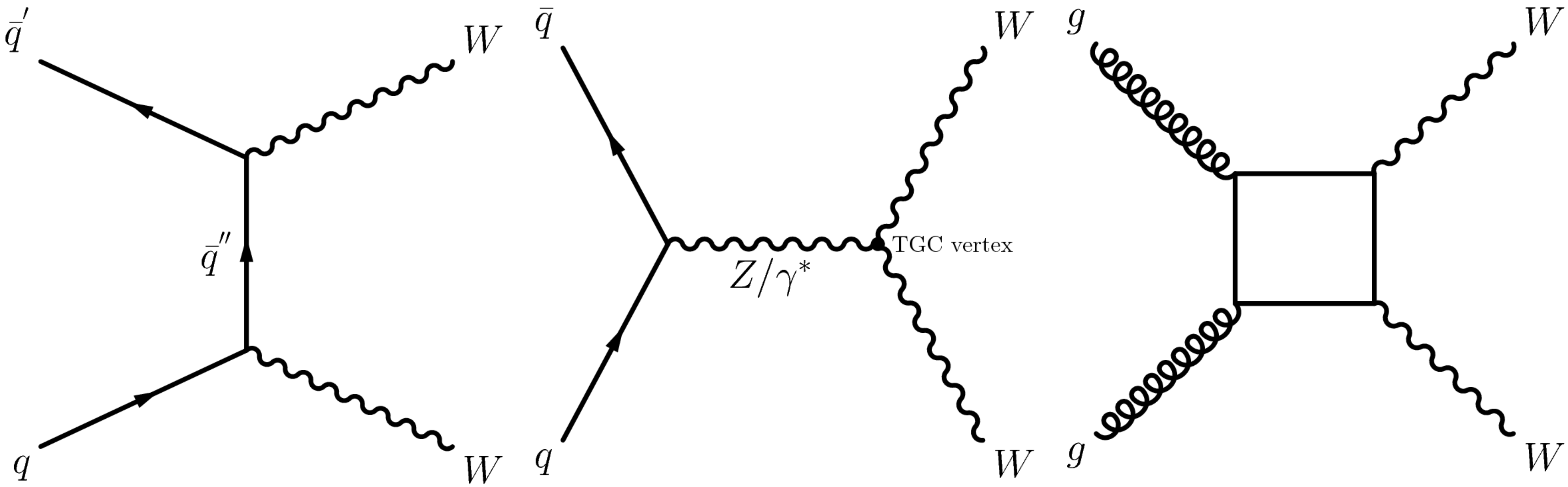
THOTH GUNTER
NORTHWESTERN PHYSICS GRADUATE STUDENT
MEMBER CMS COLLABORATION

You can reach me at thothgunter@live.com

You can find me at @TKGgunter

You can find these slides at
https://github.com/TKGgunter/presentations

WHY STUDY WW?



The WW processes allows us to probe anomalous triple gauge coupling. It is also the main irreducible background for the Higgs to dilepton decay. Unlike other diboson channels there has been discrepancy between the theory and experimental measurement of the WW cross section.

WW TROUBLE

Over the last decade WW cross section measurements and predictions have been in disagreement.

Experiment	Energy(TeV)	cross section(pb)	Theory(pb)	σ/σ_{th}
CMS[6]	7	52.4 ± 5.1	47.0 ± 2.0	1.11
ATLAS[3]	7	51.9 ± 4.8	44.7 ± 2.0	1.16
CMS[4]	8	60.1 ± 4.9	59.8 ± 1.2	1.01
ATLAS[5]	8	71.1 ± 5.6	63.2 ± 1.9	1.13

What is the cause?

- WW pt distribution.
- BSM effect
- Strict cut on jets

We are specifically studying the WW → 2l2ν. Which accounts for ~10% of the total cross section.

CROSS SECTIONS

$$\sigma = \frac{N_{data} - N_{MC_{bkg}}}{\mathscr{L} \; \epsilon \; Br \; A}$$

Measurement variables

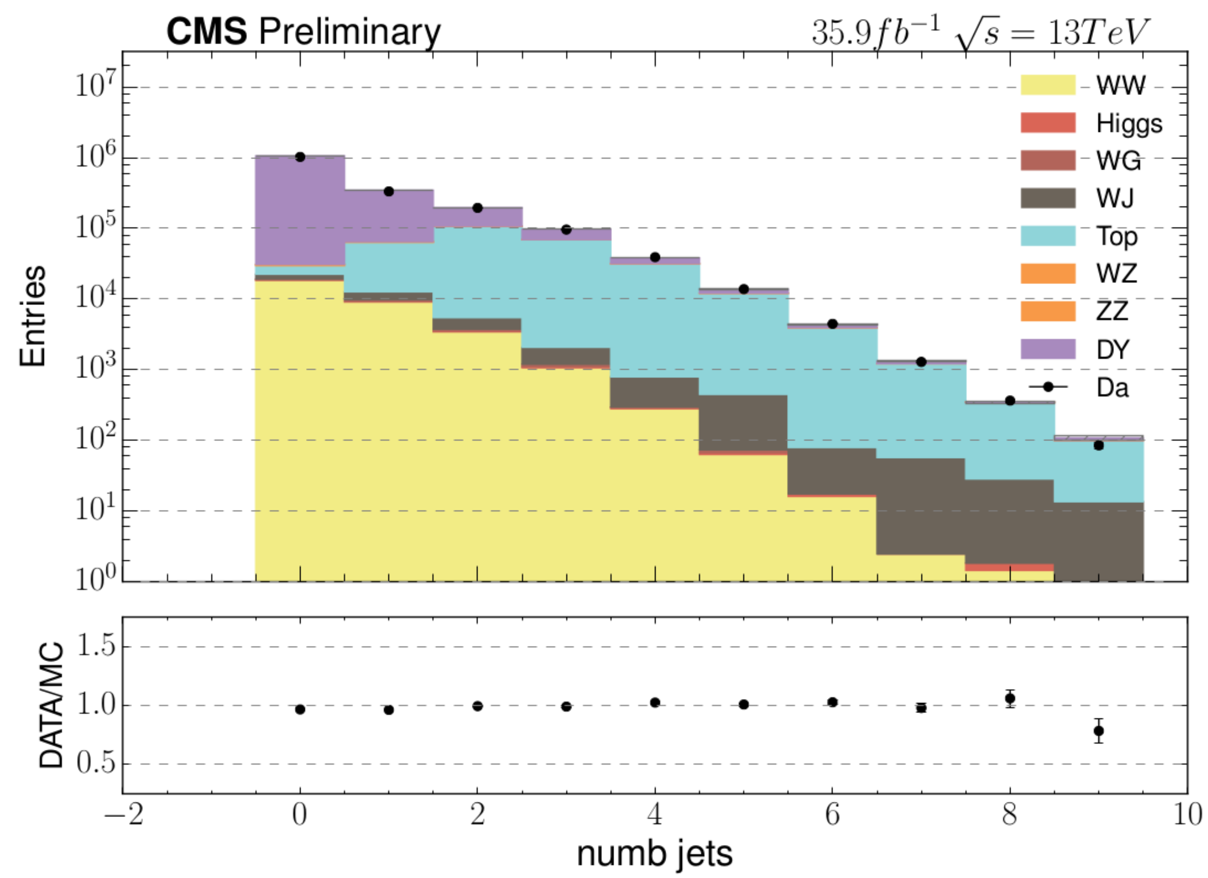
- N_{data} and $N_{MC_{bkg}}$ depend on our lepton and jet selection criteria.
- ϵ is the fraction of WW we expect to remain after event selection.
- $\mathscr{L}$ as measured CMS' luminometers.

Constants and Detector Constraints

- Br is the WW branching ratio.
- A is the percent of events that we expect to see given detector geometry, sensitivity and other.

WHAT ARE THE GOALS?

- Reduce background.
- Increase/maintain efficiency compared to standard analysis.
- Maintain as much of the jet multiplicity distribution as possible.



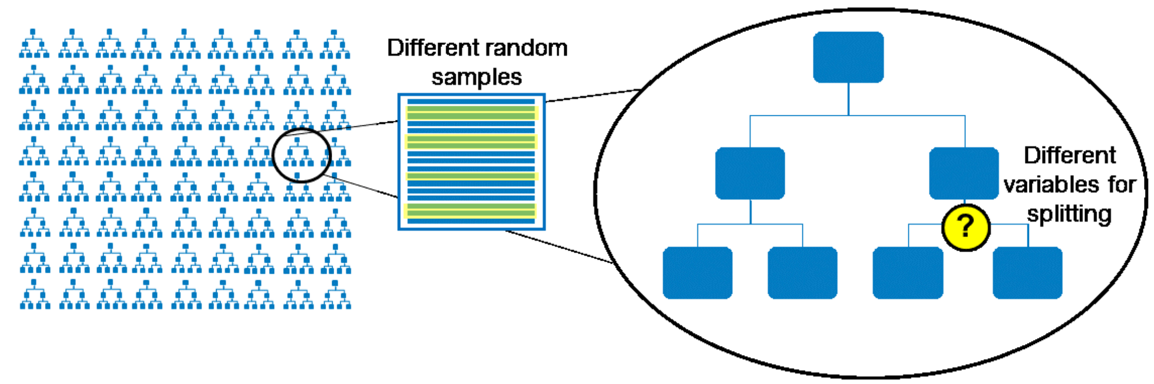
ENTER THE RANDOM FOREST

What is the random forest? Why not other classifiers?

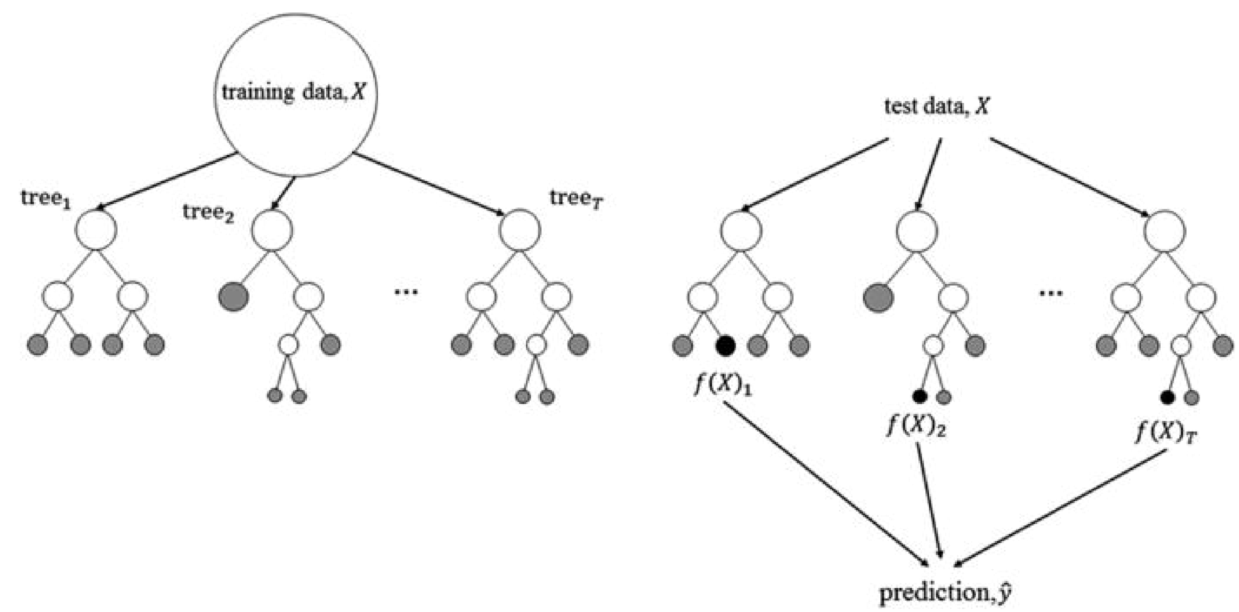
Random Forest is a multivariate classification and regression technique. Its strengths include, parallelizable training, control over-fitting, strong predictive accuracy.

How does the random forest works?

The random forest the aggregation of tree estimators. Each tree is trained on a random subset of data using a random subset of features. These splits are determined by impurity of the samples that survive/fail the cut. We derive a result by averaging the probabilistic prediction of each tree.



[1]



[2]

THE RANDOM FOREST IN PRACTICE

We wish to apply the random forest to classify events as WW events. To do this we train our forest on Monte Carlo events. Drell Yan and TTbar represent our largest backgrounds therefore are selected to be trained against.

ALL ABOUT THE FEATURES

DRELL YAN

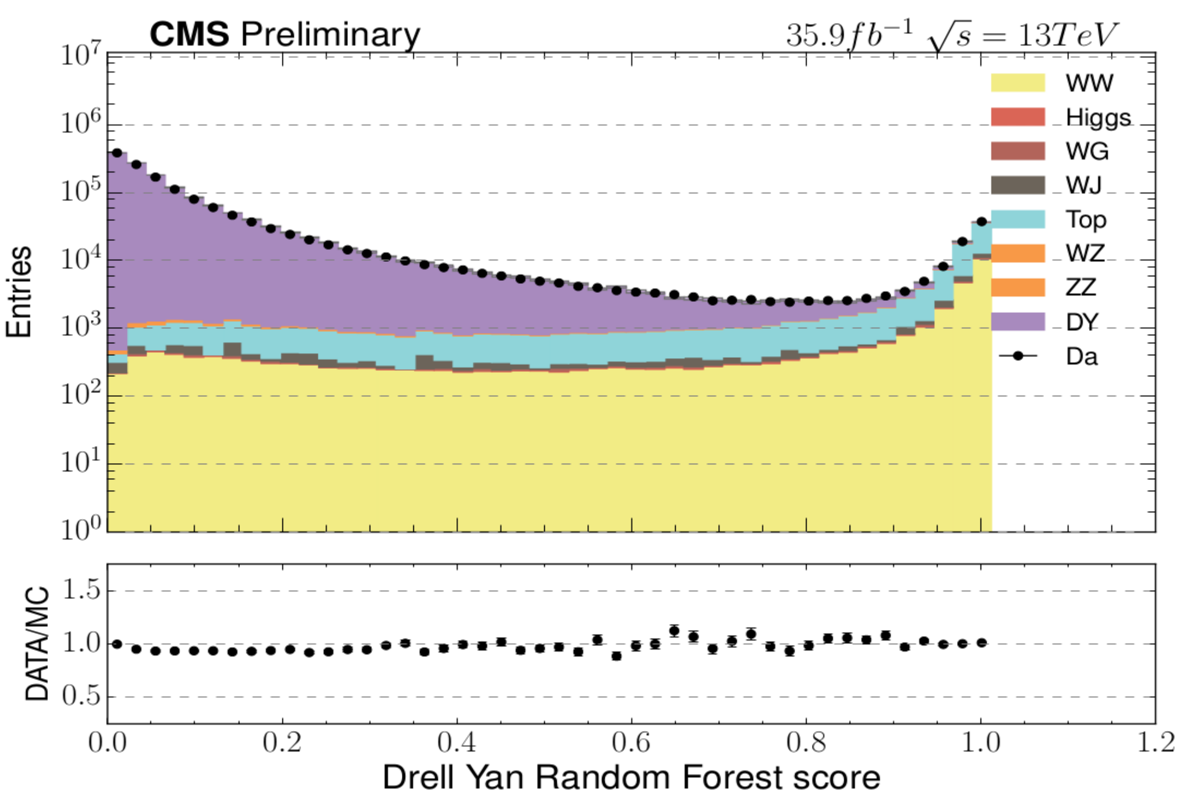
- invariant mass
- Flavor of the lepton pair
- Missing Energy
- ...

TOP QUARK

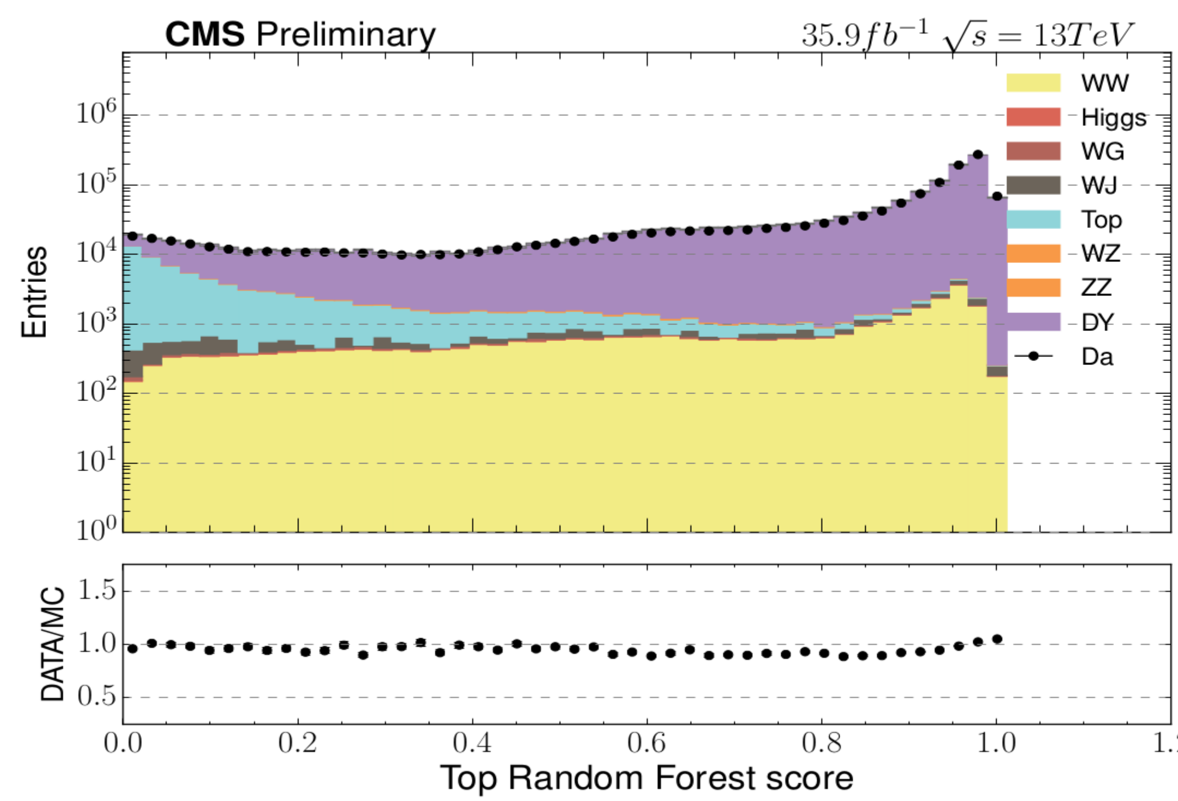
- Number of jets
- Scalar sum of jet momenta
- ...

RANDOM FOREST TRAINING RESULTS

DRELL YAN DISCRIMINATOR



TTBAR DISCRIMINATOR



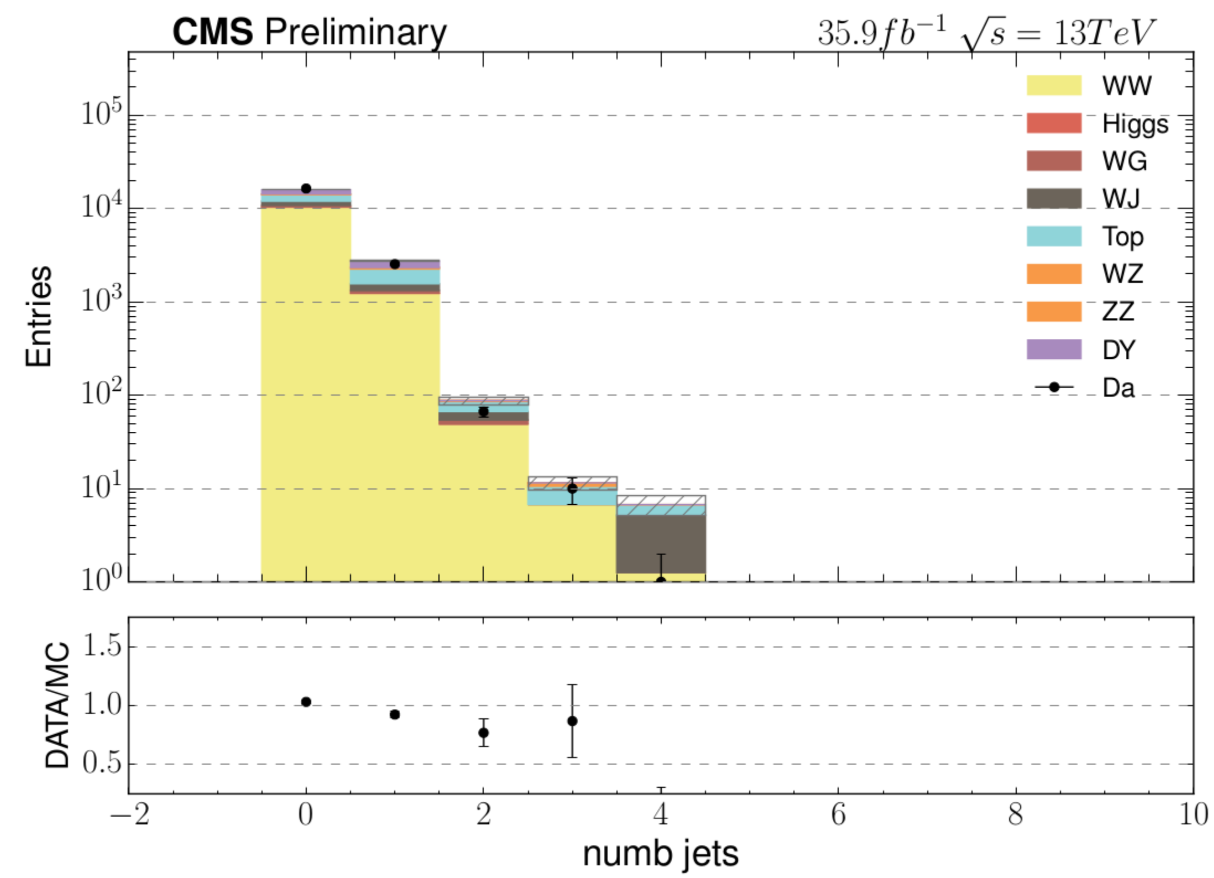
We select our cut by maximizing purity and minimizing uncertainty.

RESULTS

DID WE HIT OUR GOALS?

- Purity random forest: 0.64
- Purity standard cut based analysis: 0.50*
- Efficiency random forest: 0.025
- Efficiency standard cut based analysis: 0.021*

*Courtesy of our IFCA/MIT comrades.



We were able to drastically reduce background while probing into higher regions of jet multiplicity.

CONCLUSIONS

WW cross-section measurements have been in disagreement with Standard Model predictions and it is unclear where the discrepancy stems. By utilizing the random forest technique we are able to explore the jet multiplicity distribution without sacrificing sample purity.

LOOKING TO THE FUTURE...

We are currently in the CMS publication pipeline working to publish an inclusive cross-section and jet multiplicity measurement.

QUESTIONS?

BACKUPS

REFERENCES

1. Wujek(2015)https://communities.sas.com/legacyfs/online/10908_rf-concept.gif
2. Mennitt "Conceptual-diagram-of-the-RANDOM-FOREST-algorithm-On-the-left-trees-are-trained.png" https://www.researchgate.net/profile/Daniel_Mennitt/publication/278677730/figure/fig2/AS:282631661998100@144 diagram-of-the-RANDOM-FOREST-algorithm-On-the-left-trees-are-trained.png
3. ATLAS Collaboration, "Measurement of W + W − production in pp collisions at √s = 7 TeV with the ATLAS detector and WWZ and WWγ couplings", Phys. Rev. D87 (2013), no. 11, 112001, doi:10.1103/PhysRevD.87.112001,10.1103/PhysRev arXiv:1210.2979. [Erratum: Phys. Rev.D88,no.7,079906(2013)].
4. CMS Collaboration, "Measurement of the W W cross section in pp collisions at √s = 8 TeV and limits on anomalous ga J. C76 (2016), no. 7, 401, doi:10.1140/epjc/s10052-016-4219-1, arXiv:1507.03268.
5. ATLAS Collaboration, "Measurement of total and differential W^+W^- production cross sections in proton-proton collisi ATLAS setector and limits on anomalous triple-gauge-boson couplings", JHEP 09 (2016) 029, doi:10.1007/JHEP09(20
6. CMS Collaboration, "Measurement of the W W Cross section in pp Collisions at √s = 7 TeV and Limits on Anomalous W Eur. Phys. J. C73 (2013), no. 10, 2610, doi:10.1140/epjc/s10052-013-2610-8, arXiv:1306.1126.