INSTITUTE FOR COLLIDER PARTICLE PHYSICS

UNIVERSITY OF THE WITWATERSRAND

University of Venda

Search for heavy scalar resonances in the LHC run 2 dataset in $Z\gamma$ final states using machine learning techniques.

HEPP WORKSHOP 2020

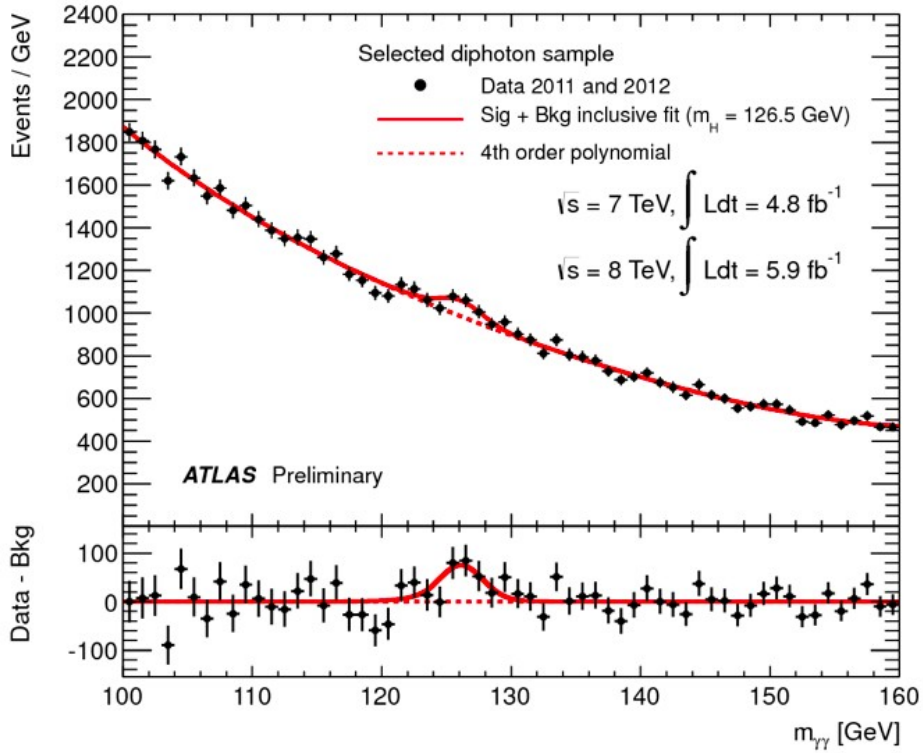
University of Venda

Gaogalalwe Mokgatitswane

Salah-Eddine Dahbi | Xifeng Ruan | Ntsoko Phuti Rapheeha | Bruce Mellado

Overview

- Motivated by the statistically significant excesses in the multi-lepton final states compatible with physics at the electroweak scale. Refs.[**arXiv:1711.07874, arXiv:1901.05300**]
- We possibly have some signal in the 240-280 GeV region, but it's too small and mixed with a huge background.
- We are setting up a search based on machine learning tools – This can split better the signal and background
- Conduct a weakly-supervised and full-supervised learning in conjuction with a Deep Neural Network Algorithm.

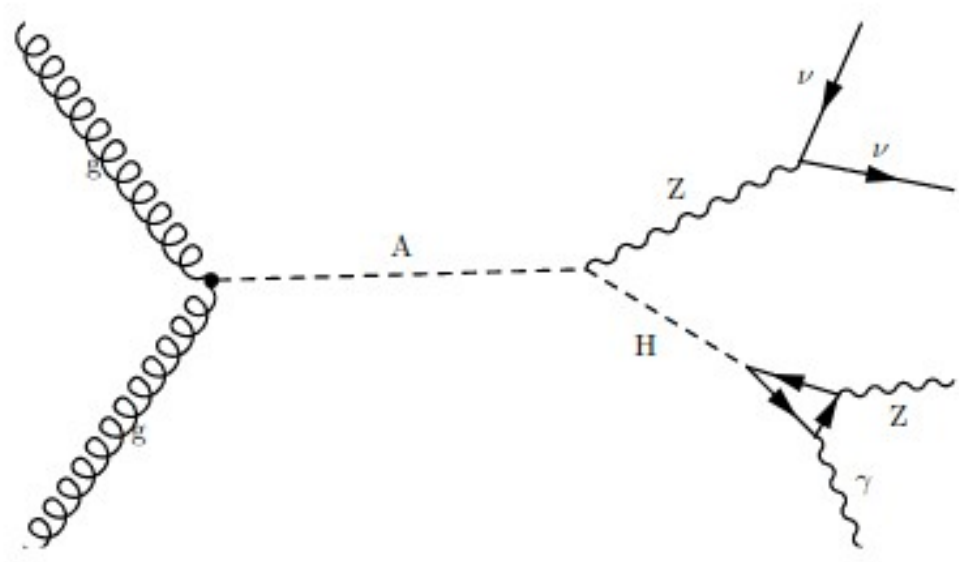


- Here is an illustration of the discovery of Higgs Boson, showing how in general HEP people are doing.
- We are trying to find a signal peak from such kind of a continuos background.

- A simillar problem we are dealing with!

Z + gamma

- We want to search for a heavy Higgs (*H*) decaying to Zγ in association with E_T^{miss} .
- An illustration of one of the possible processes which is a heavy mass of particle A decaying to Z and H



The production and decay of the scalars in a simplified AZH model

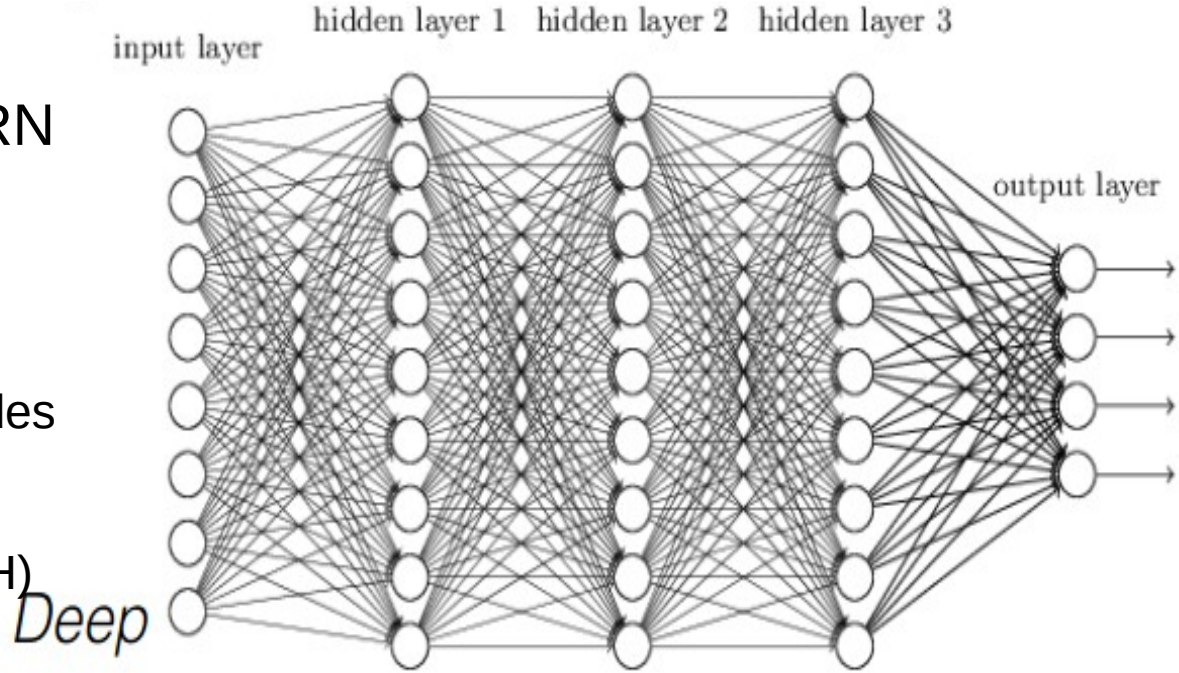
Machine Learning (methodology)

- Full-supervised learning: requires distinct Signal and a Background samples (labelled data)
- Weakly-supervised learning: requires a Background sample and a sample from data that contains Background and an unknown mixture of Signal (unlabelled data), hidden in the real data.

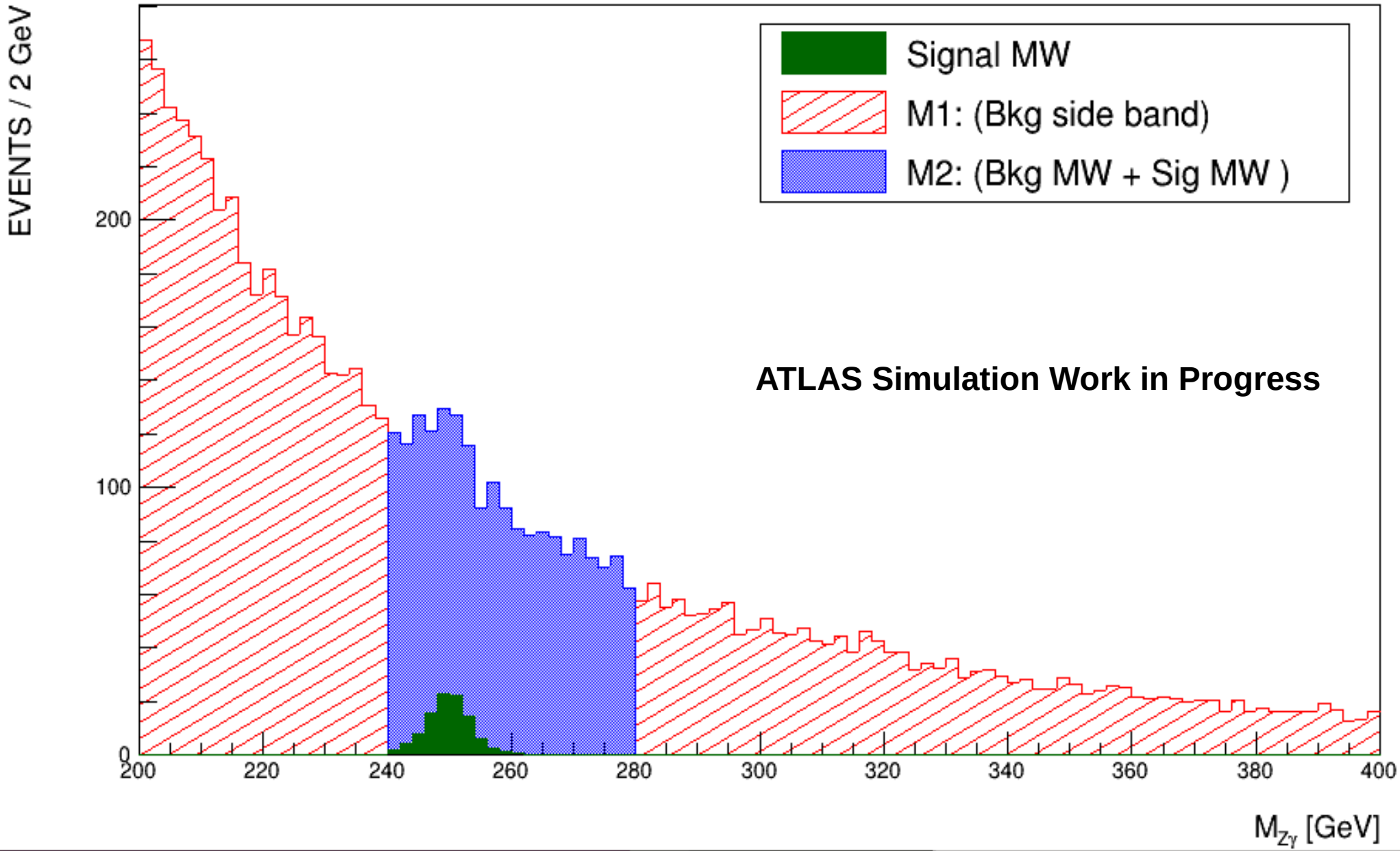
Toolkit for Multivariate Data Analysis (TMVA) framework developed by CERN for high energy physics

MODEL CONFIGURATION: DNN_CPU

- One input layer with variation number of nodes
- 3 hidden layers with each 128 nodes
- One output layer
- Hyperbolic tangent activation function (TANH)
- BatchSize = 256
- Number of epochs = 80

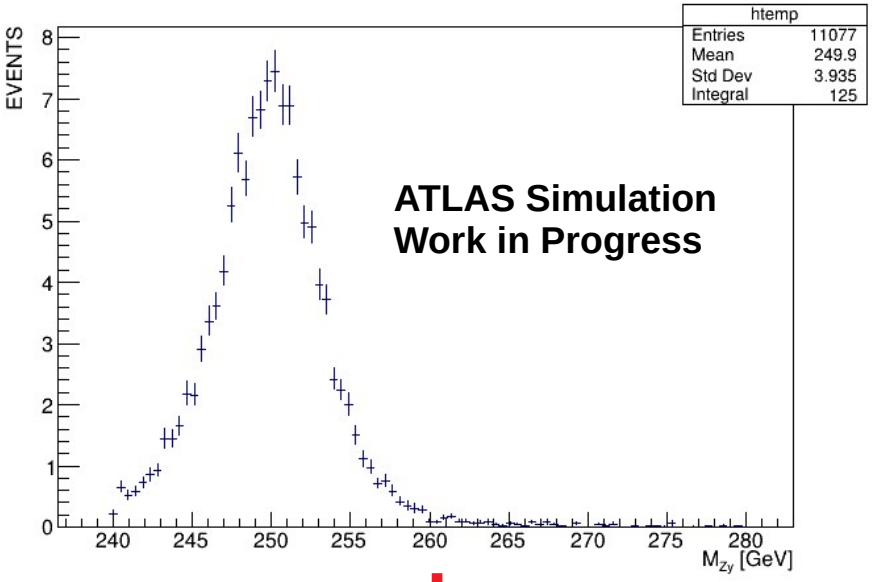


$Z\gamma$ mass spectrum

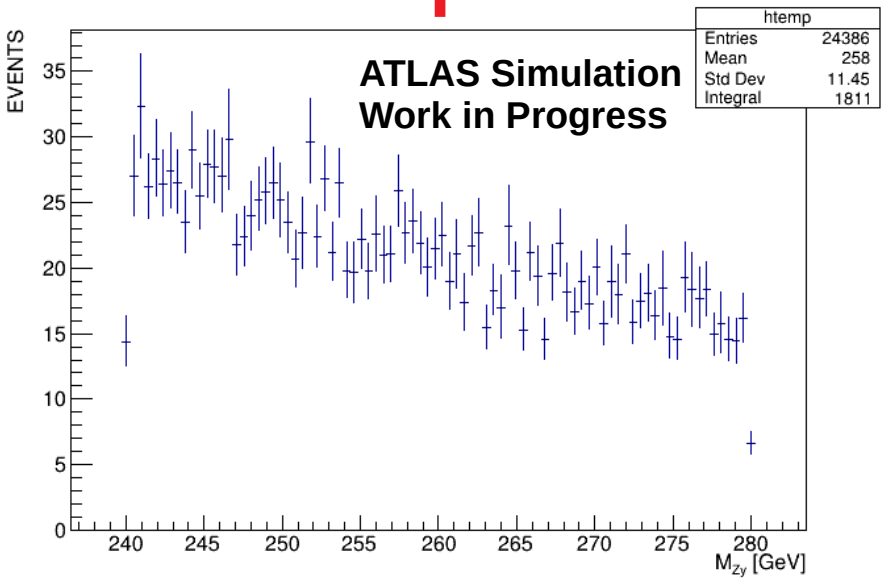


Cuts requirement

- Signal + Background in Mass Window:
  - m_{Zy}<=280 && m_{Zy}>=240 GeV

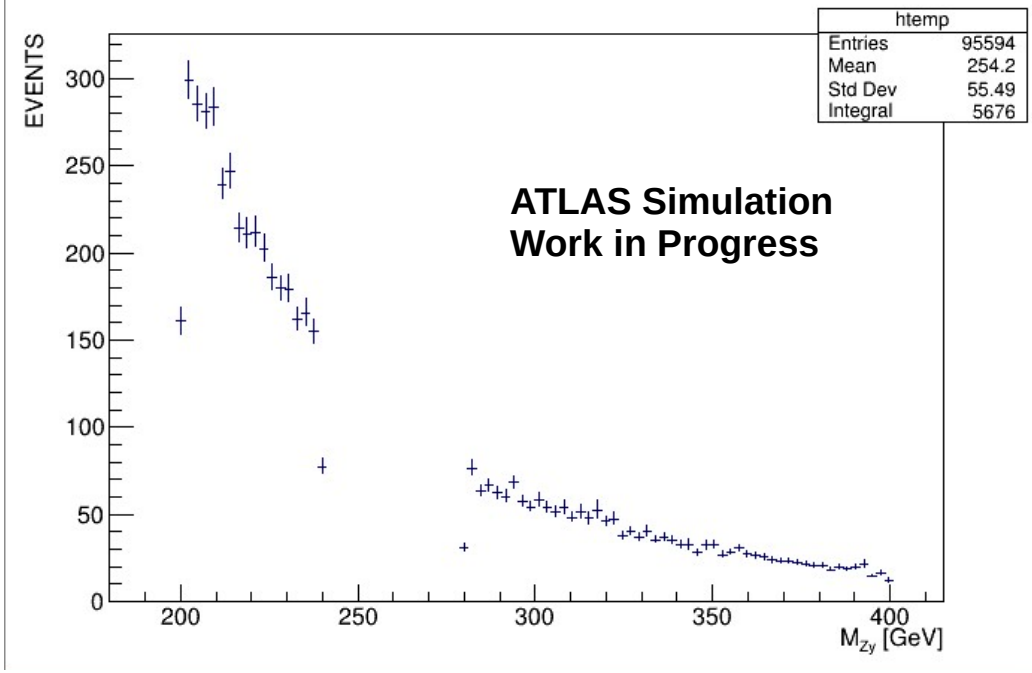


+



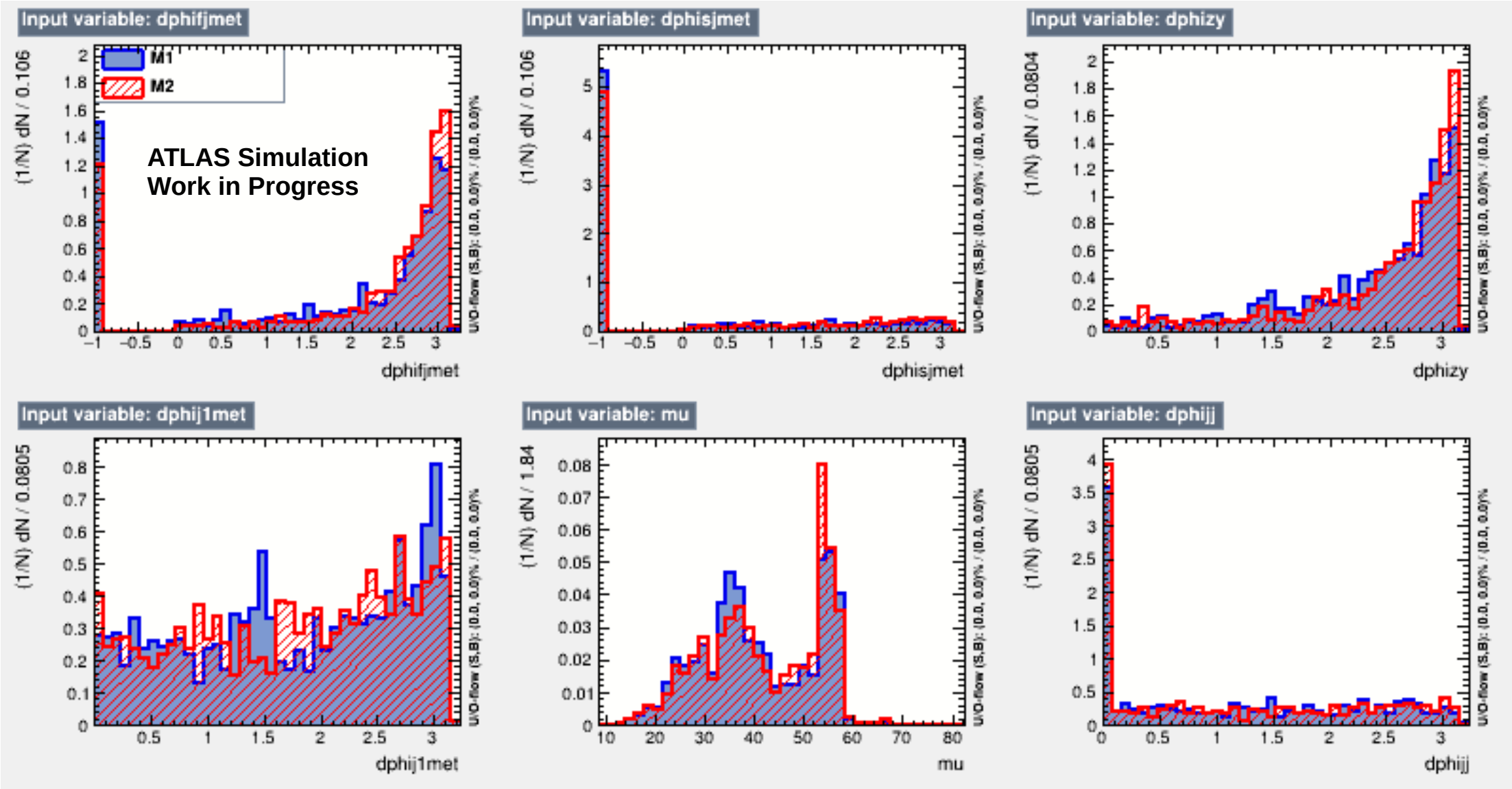
VS

- Background side band:
  - m_{Zy}<=240 && m_{Zy}>=200 || m_{Zy}<=400 && m_{Zy}>=280 GeV



$$E_{T,sig}^{miss} \geq 2.5\,GeV$$

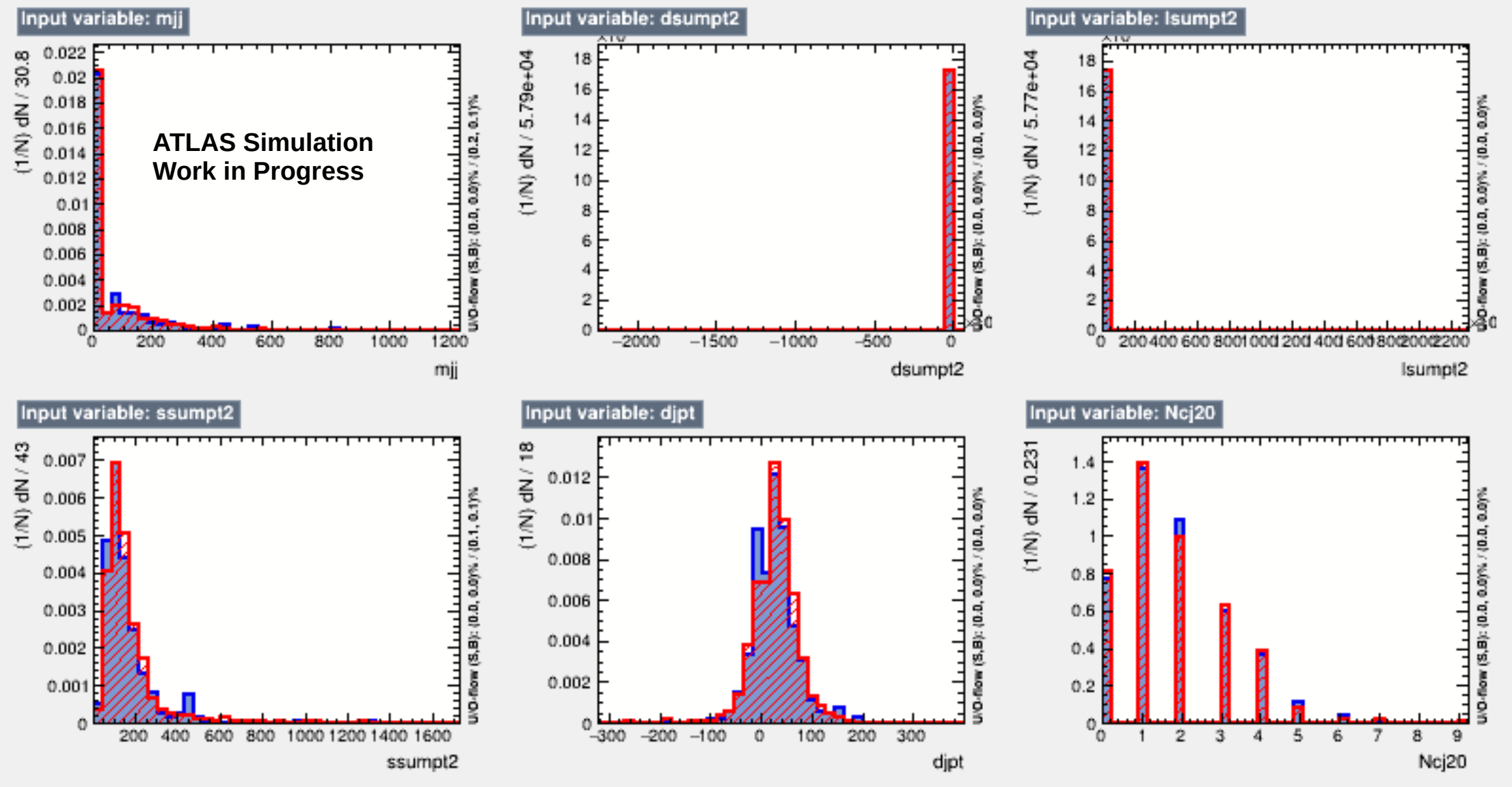
1D distribution of input variables



M1: Signal MW + Background MW

M2: Background side band

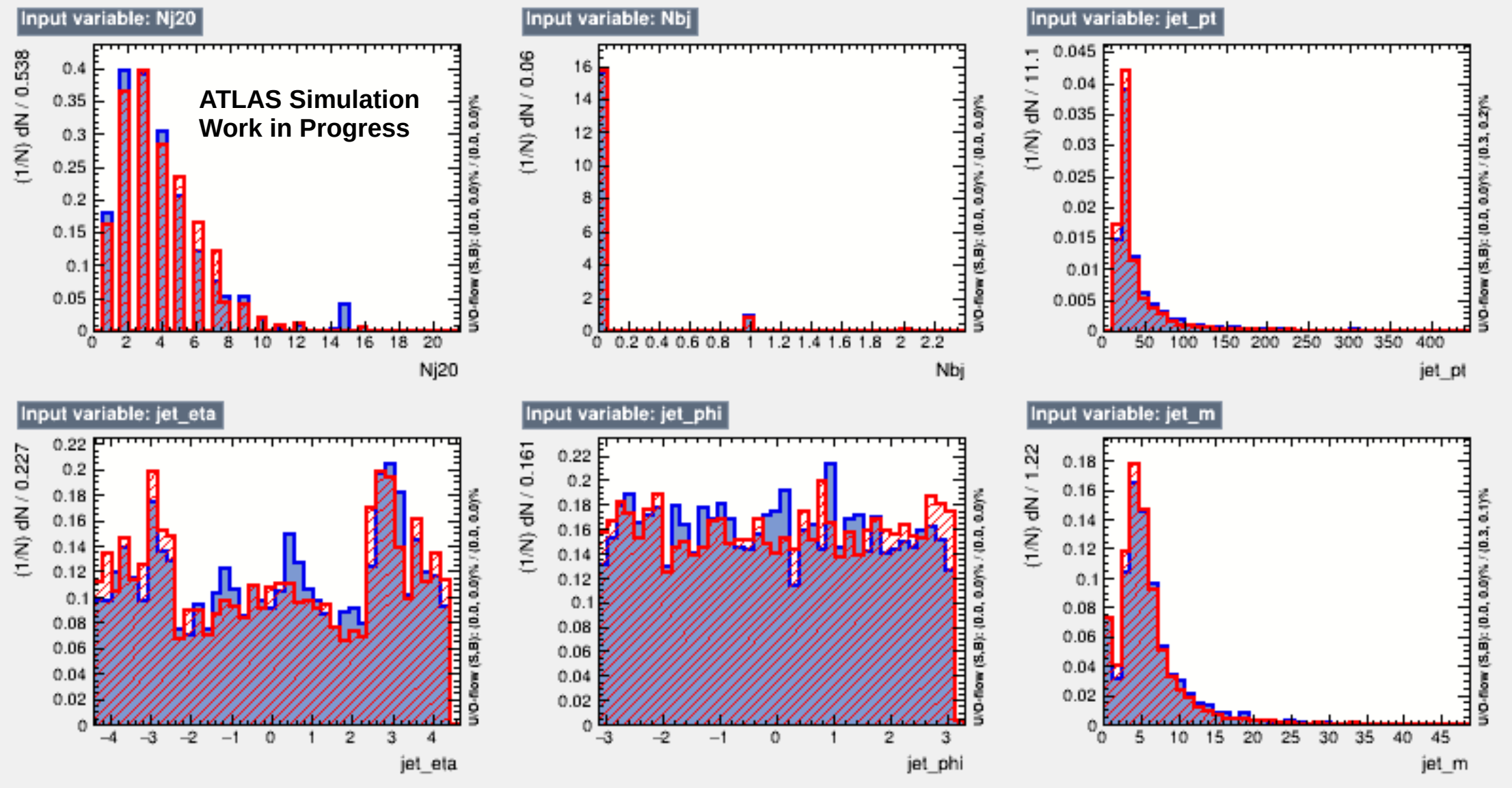
1D distribution cont...



M1: Signal MW + Background MW

M2: Background side band

1D distribution cont...

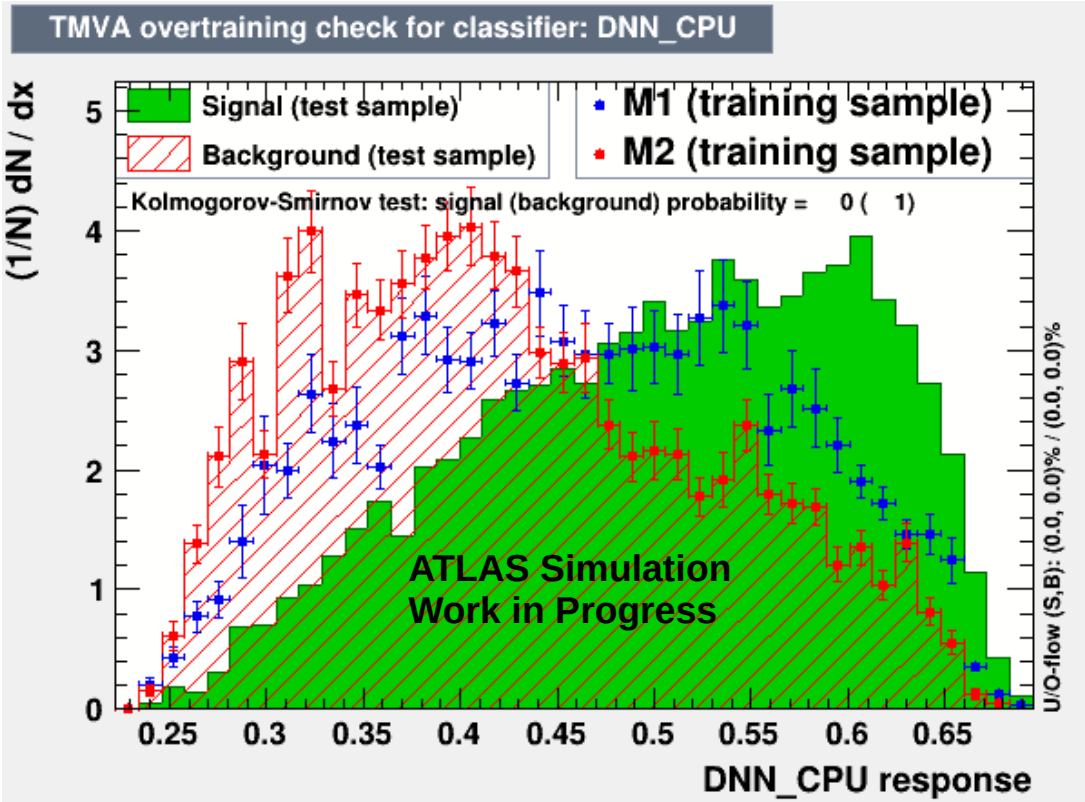


M1: Signal MW + Background MW

M2: Background side band

Overtraining check for DNN_CPU classifier

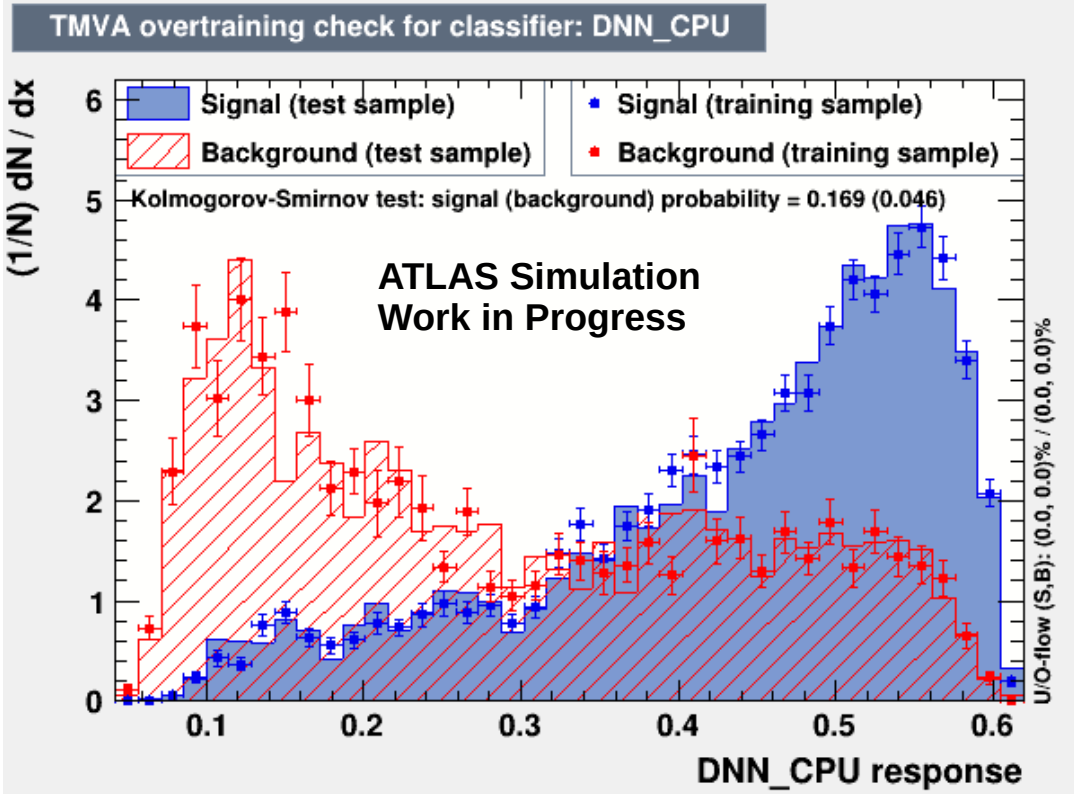
Weak-supervision



M1: Signal MW + Background MW
M2: Background side band

Training: 60502 events
Testing: 51535 events

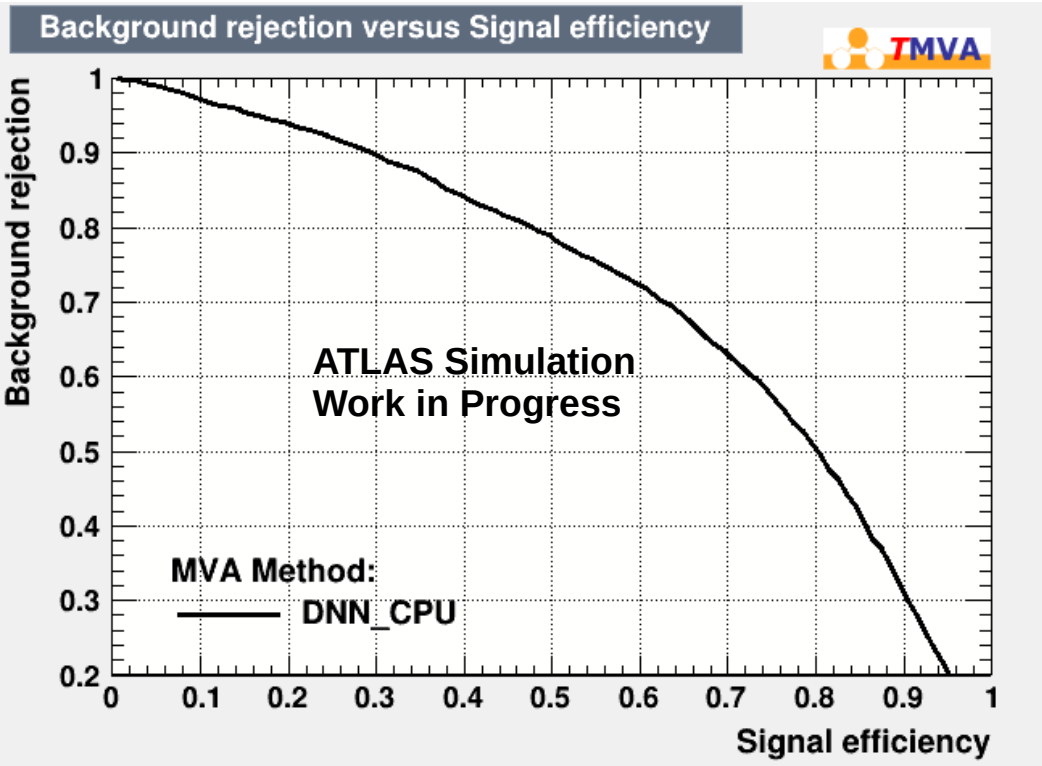
Full-supervision



Training: 25767 events
Testing: 25767 events

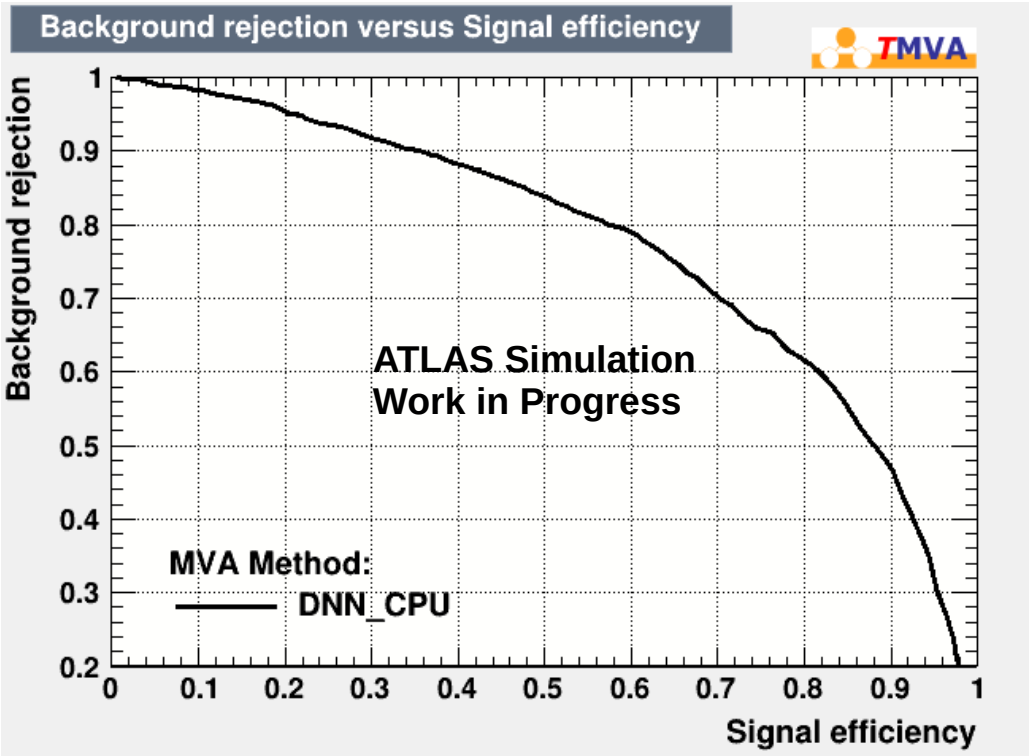
ROC curve plot for DNN classifier

Weak-supervision



ROC-integ: 0.715

Full-supervision



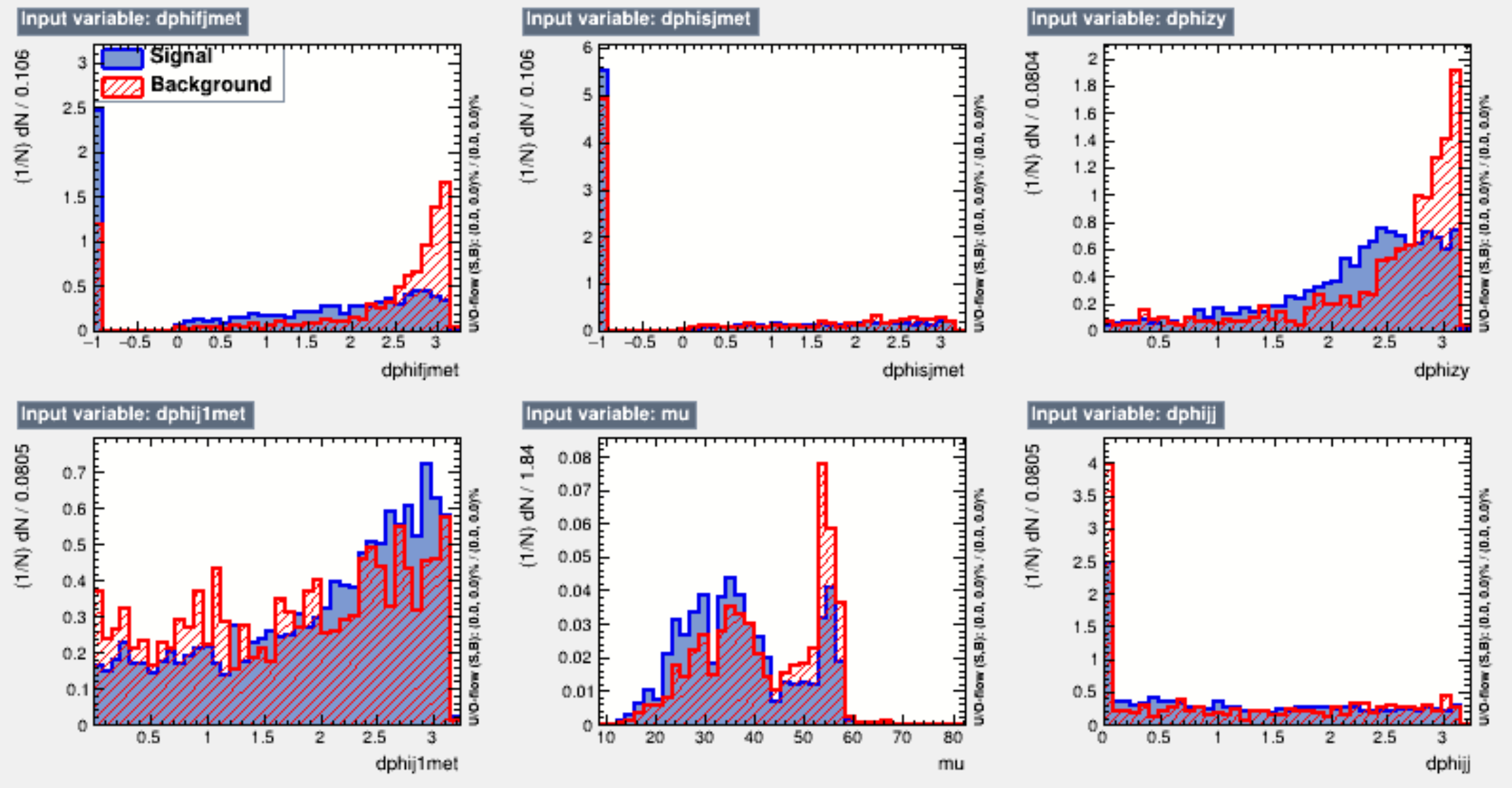
ROC-integ: 0.773

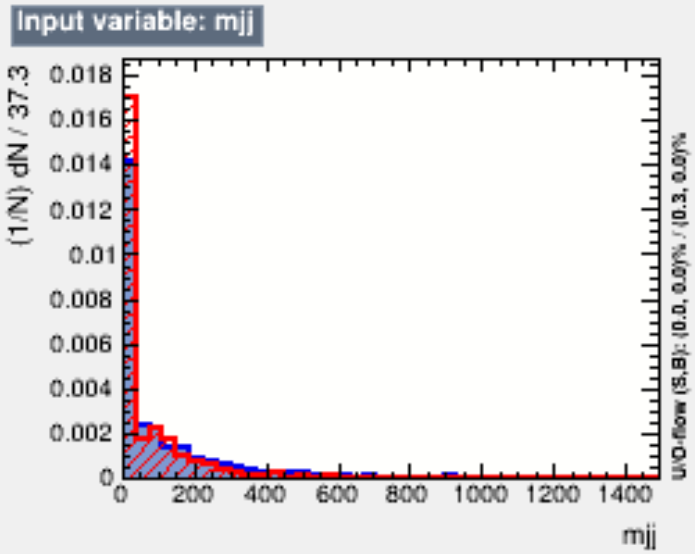
Conclusion

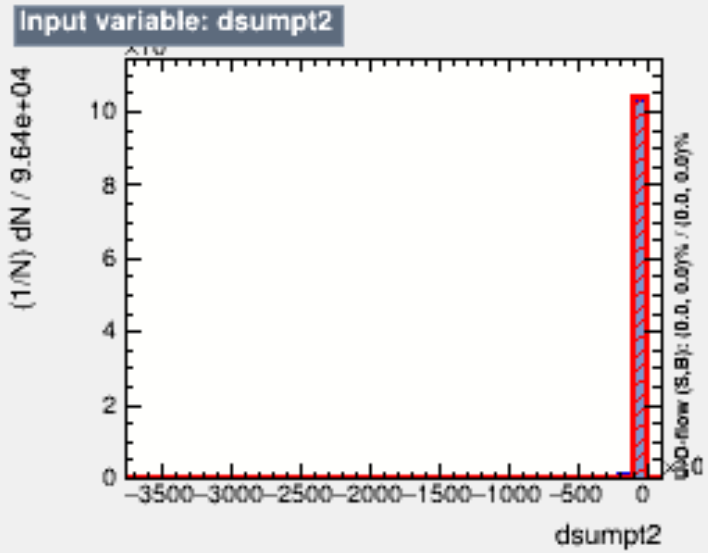
- A search for heavy Higgs in the Zy decay channel has been established. This was motivated by the statistically significant excesses in thte multi-lepton final states compatible with physics at the EW scale.
- The search is based on machine learning techniques for separation of the signal form background - Toolkit for Multivariate Data Analysis with ROOT (TMVA).
- The combination of DNNs, advanced machine learning techniques and weak supervision methodology can help for finding new particles in High Energy Physics
- What we have tested in the Zy decay channel showed some improvements and a proof that indeed the search strategy is working.
- This methodology can be implemented in the LHC searches and analyses towards discovery of new physics.

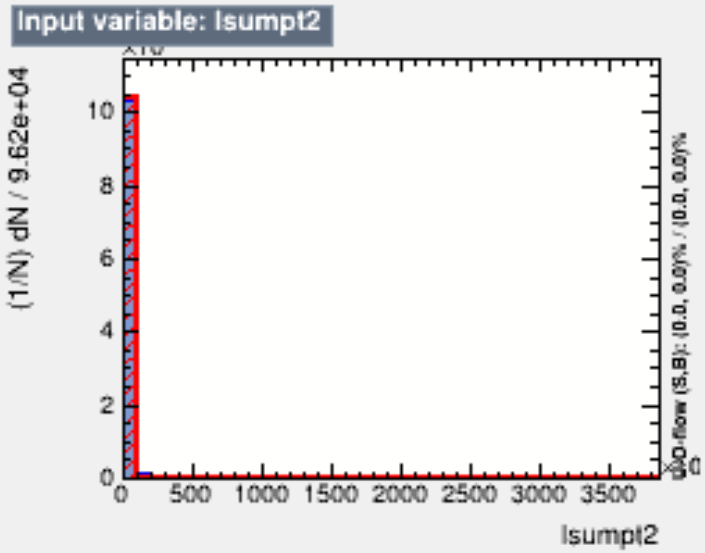
THANK YOU

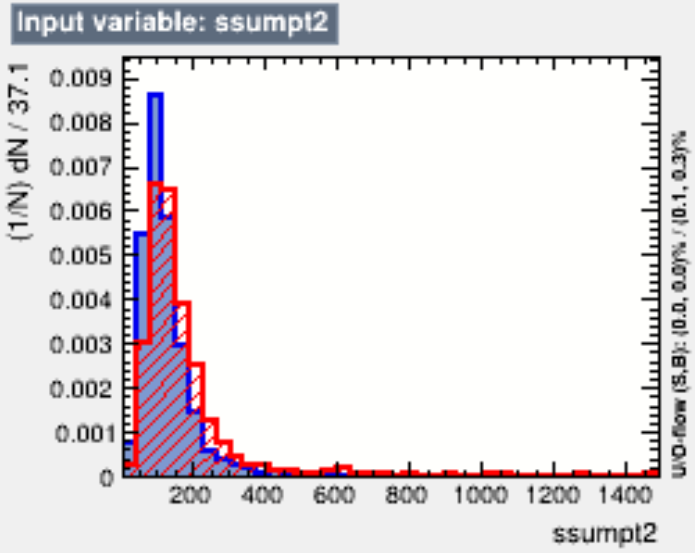
Back-up slides (1D distribution of input variables- Full supervision)

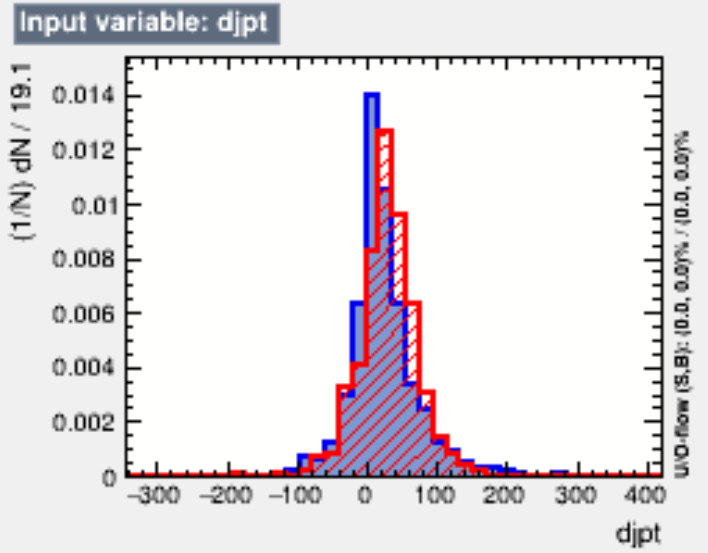


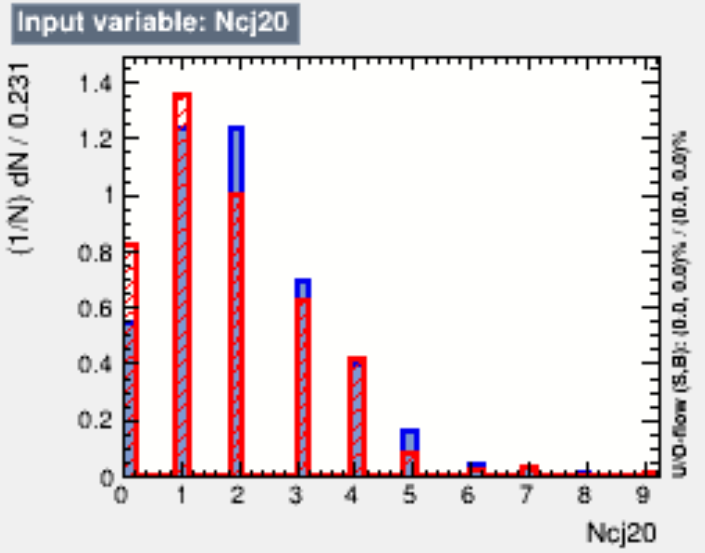

Input variable: mjj
(1/N) dN / 37.3
mjj
U/O-flow (S,B): (0.0, 0.0)% / (0.3, 0.0)%

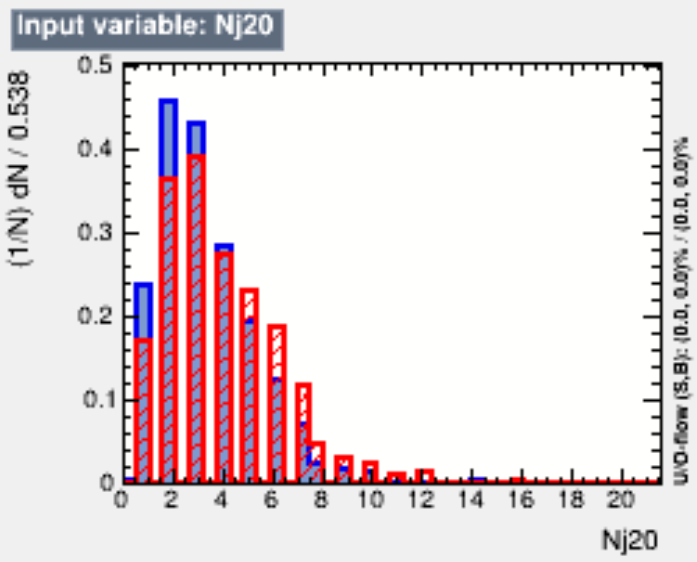

Input variable: dsumpt2
(1/N) dN / 9.64e+04
dsumpt2
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

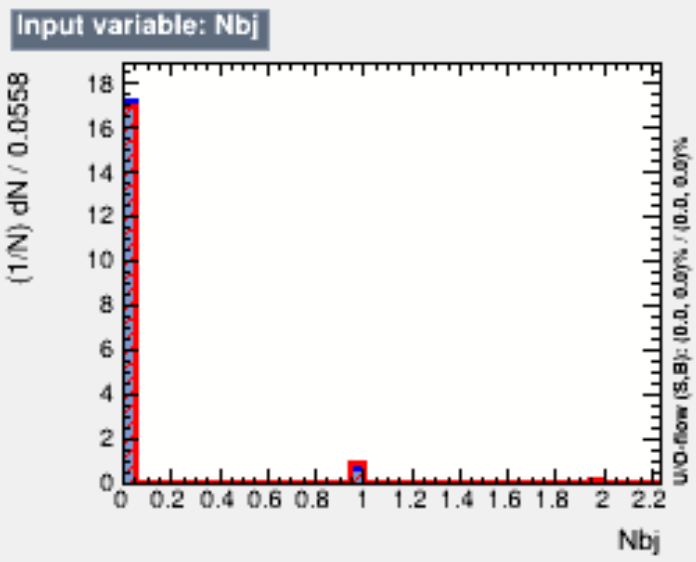

Input variable: lsumpt2
(1/N) dN / 9.62e+04
lsumpt2
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

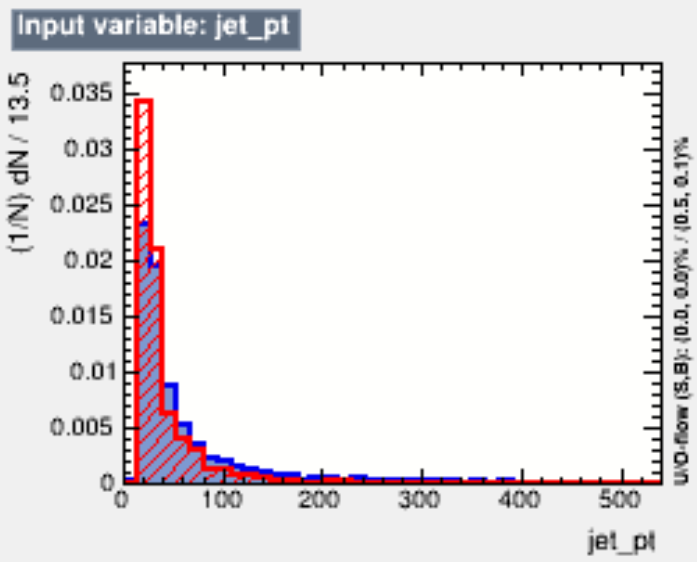

Input variable: ssumpt2
(1/N) dN / 37.1
ssumpt2
U/O-flow (S,B): (0.0, 0.0)% / (0.1, 0.3)%

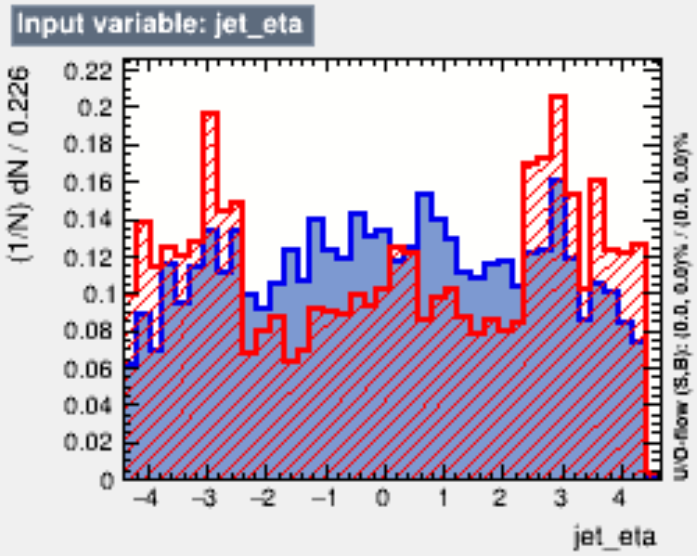

Input variable: djpt
(1/N) dN / 19.1
djpt
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

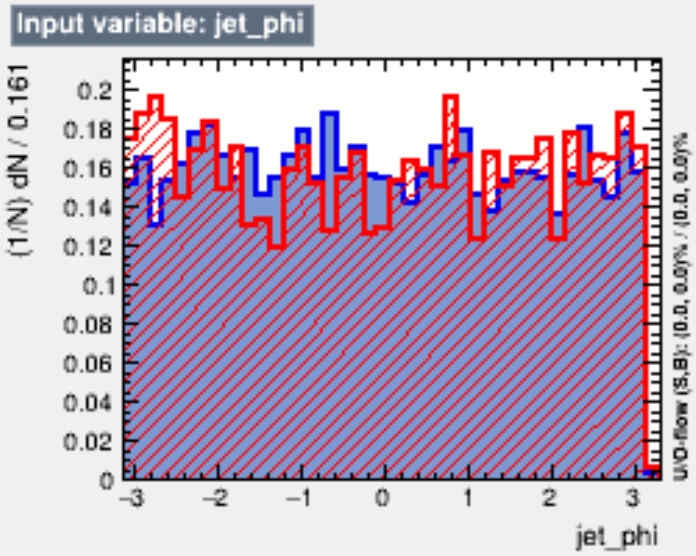

Input variable: Ncj20
(1/N) dN / 0.231
Ncj20
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

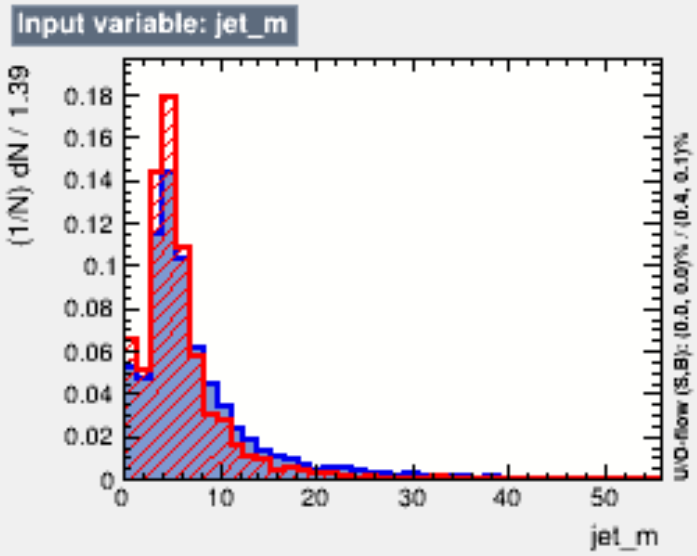

Input variable: Nj20
(1/N) dN / 0.538
Nj20
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


Input variable: Nbj
(1/N) dN / 0.0558
Nbj
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


Input variable: jet_pt
(1/N) dN / 13.5
jet_pt
U/O-flow (S,B): (0.0, 0.0)% / (0.5, 0.1)%


Input variable: jet_eta
(1/N) dN / 0.226
jet_eta
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


Input variable: jet_phi
(1/N) dN / 0.161
jet_phi
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


Input variable: jet_m
(1/N) dN / 1.39
jet_m
U/O-flow (S,B): (0.0, 0.0)% / (0.4, 0.1)%