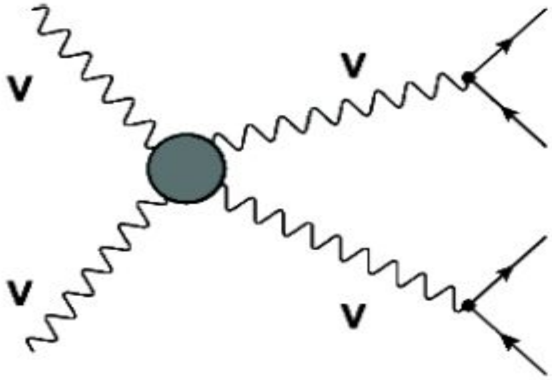
The Search for Evidence of Vector Boson Scattering between a W boson and a Photon in proton-proton collisions

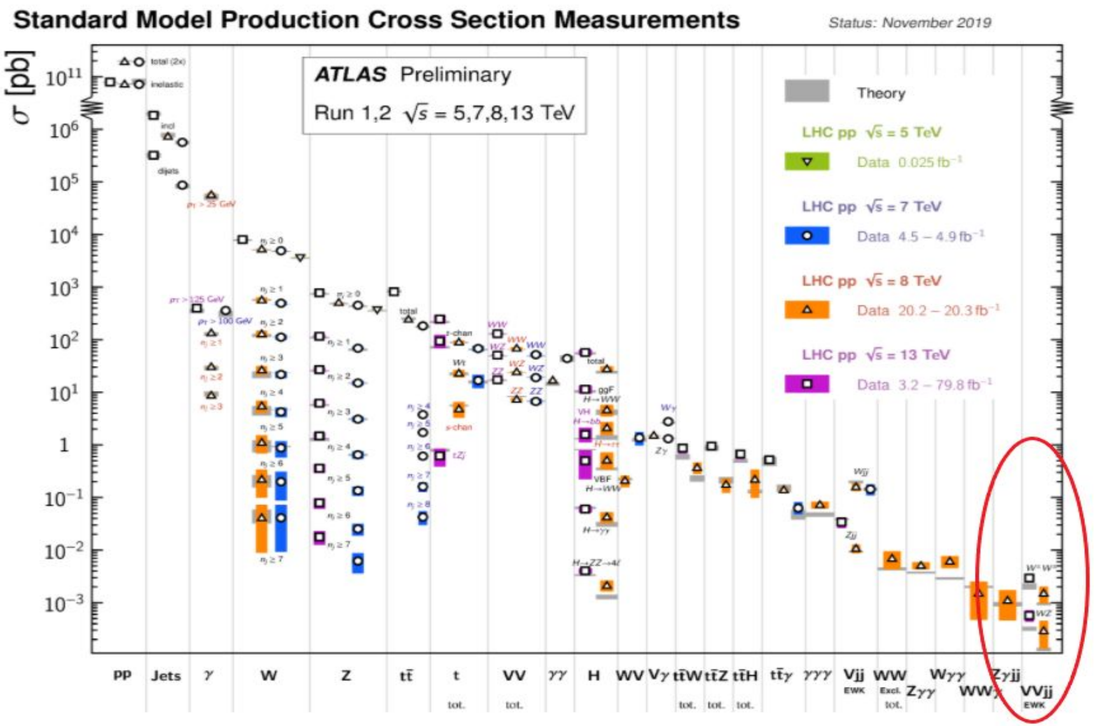
John McGowan, McGill University
Supervisor: Prof. Brigitte Vachon

Why Study Vector Boson Scattering (VBS)?

- In the Standard Model, the interactions between gauge bosons are completely specified by the SU(2) x U(1) structure of the theory.
- This makes the study of the interactions between gauge bosons a powerful approach to search for new physics.
- Any deviation from SM predictions in the self interactions of gauge bosons would indicate the presence of new physics phenomena.



VBS at the LHC



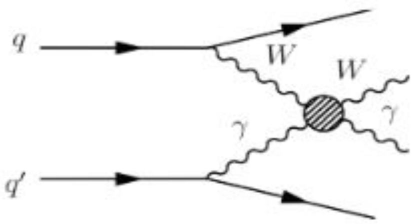
[ATL_PHYS_PUB_2019_24]

- The LHC provides a unique environment in which to study rare Standard Model processes.
- The search for evidence of scattering between a W boson and a photon is carried out using a total of 139 fb^{-1} data collected by the ATLAS detector at $\sqrt{s}$ = 13 TeV.

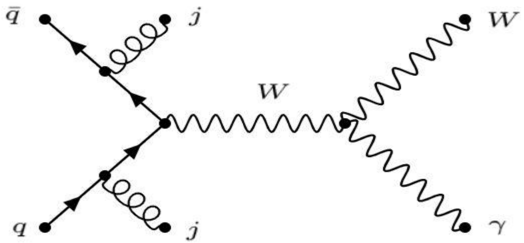
Challenges of VBS Measurements

- The search for the scattering of a W boson and a photon comes with formidable challenges.
  - There is a large and irreducible background from poorly modelled processes involving strong interactions between particles (QCD):

Signal



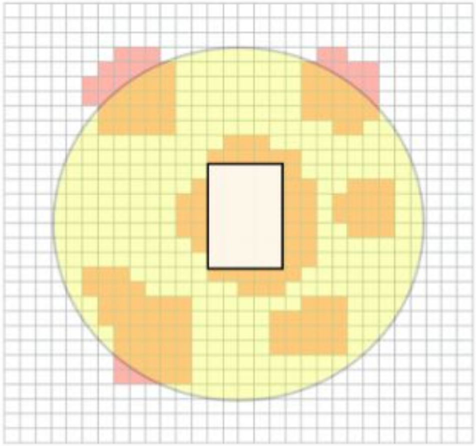
Background



  - The imperfect modelling of the detector response results in a non-negligible number of jets being misidentified as photons.
- In this talk I will discuss a data driven approach to estimating the number of jets misidentified as photons and a machine learning approach to estimating the number of signal events from the large irreducible background.

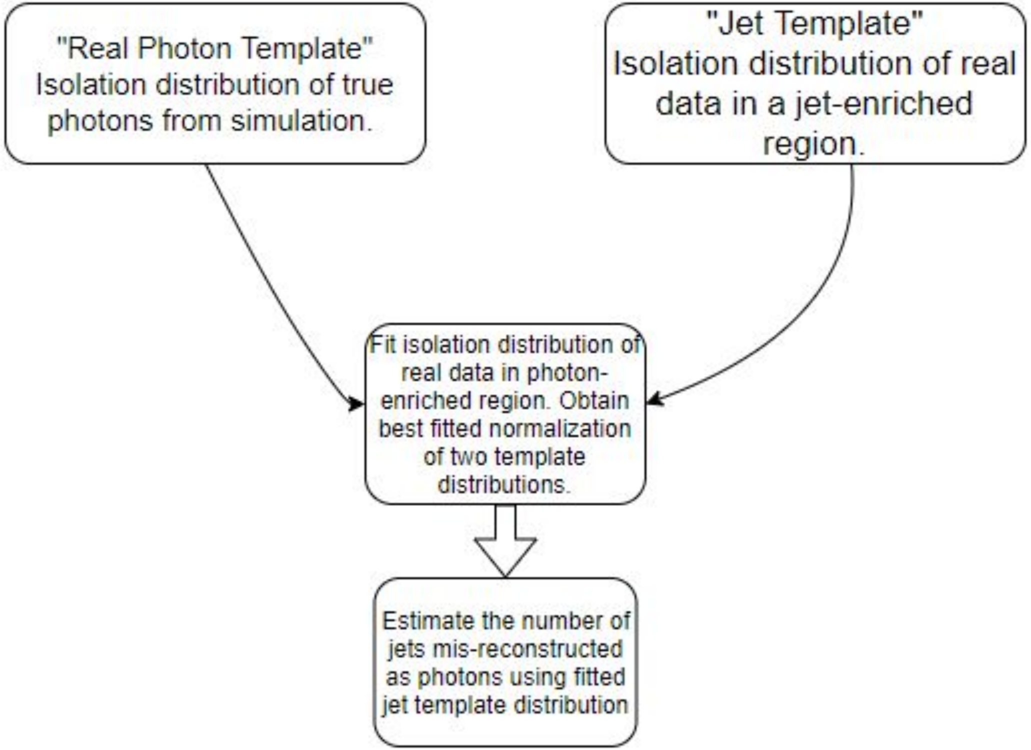
Jets Faking Photons - A Data Driven Approach

- Simulated events cannot be used to estimate the size of this background due to the imperfect modelling of the detector response to jets.
- Photon Identification relies on measurements in the calorimeter.
    - A photon candidate satisfying various cuts on calorimeter variables is considered tight.
    - A photon candidate that fails a subset of these cuts is considered non-tight.
    - A tight and isolated photon candidate is considered a signal photon.



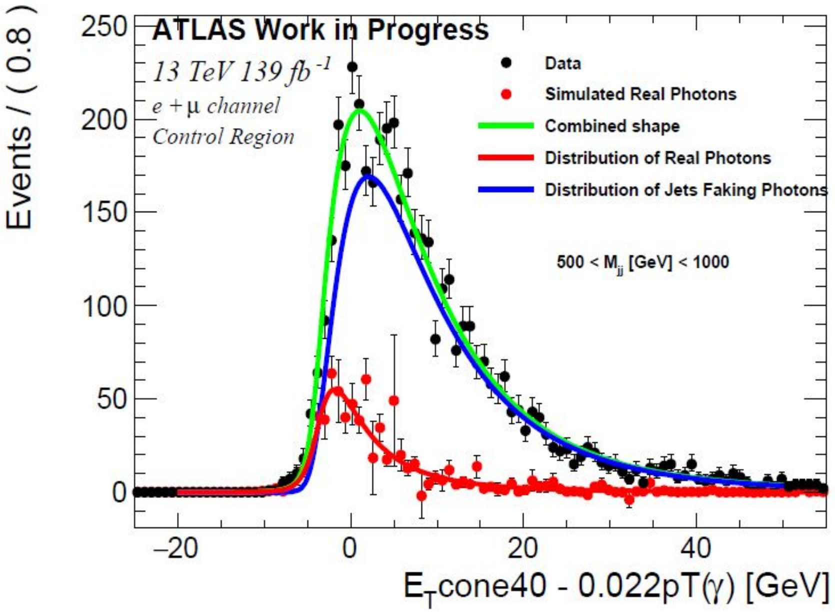
- Photon candidate identification variables are uncorrelated with isolation energy
    - The shape of the isolation distribution for tight and non-tight photon candidates, originating from jets, is the same.
    - Use this property to estimate the number of jets misidentified as photons.

The Template Fake Method



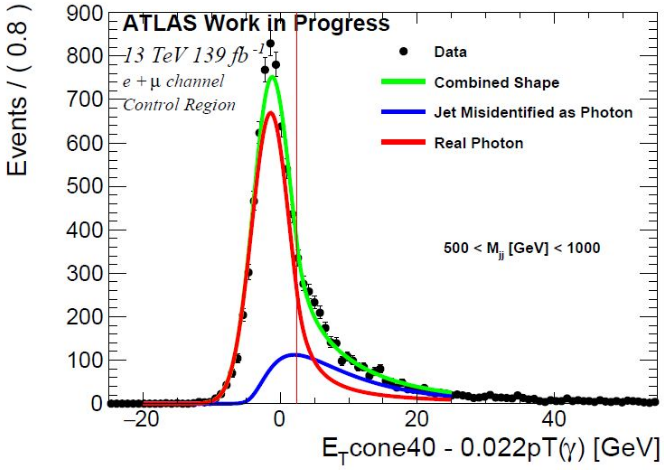
Determining the isolation shape of jets mis-identified as photons

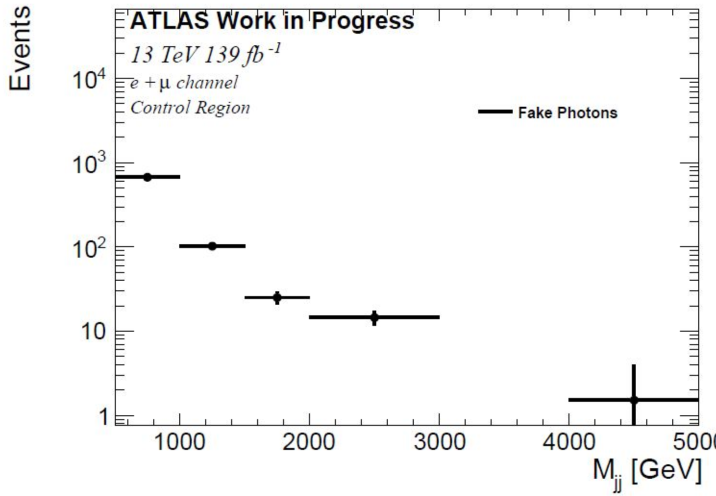
- Jet enriched region still contains a small fraction of true photons.
  - Determine the shape of the isolation of true photons from simulated data.
  - Assume a functional form for jets contribution.
  - Fit combined shapes to real data.
- Shape of isolation of jets mis-identified as photons given by best-fit parameters of jet function (blue curve).



Determining the number of reconstructed photons originating from mis-identified jets

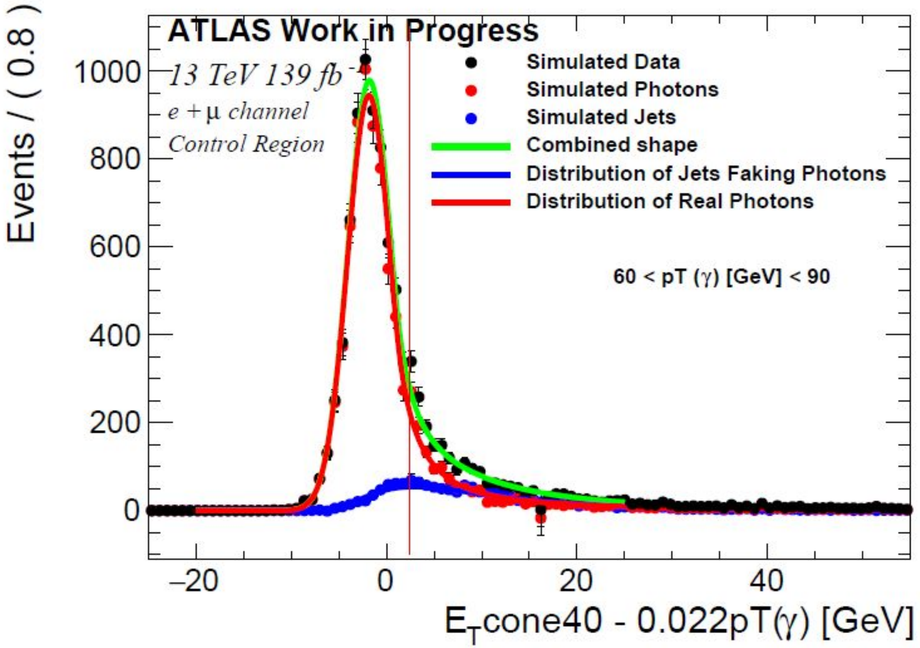
- Photon-enriched region contains a small fraction of jets.
  - Use isolation shape of true photons from simulated data.
  - Use isolation shape of jet distribution determined in previous step.
  - Fit combined shapes to real data, taking normalizations as free parameters..
- Number of jets mis-reconstructed as photons obtained by integrating jet distribution in the signal region (-25 to 2.45).





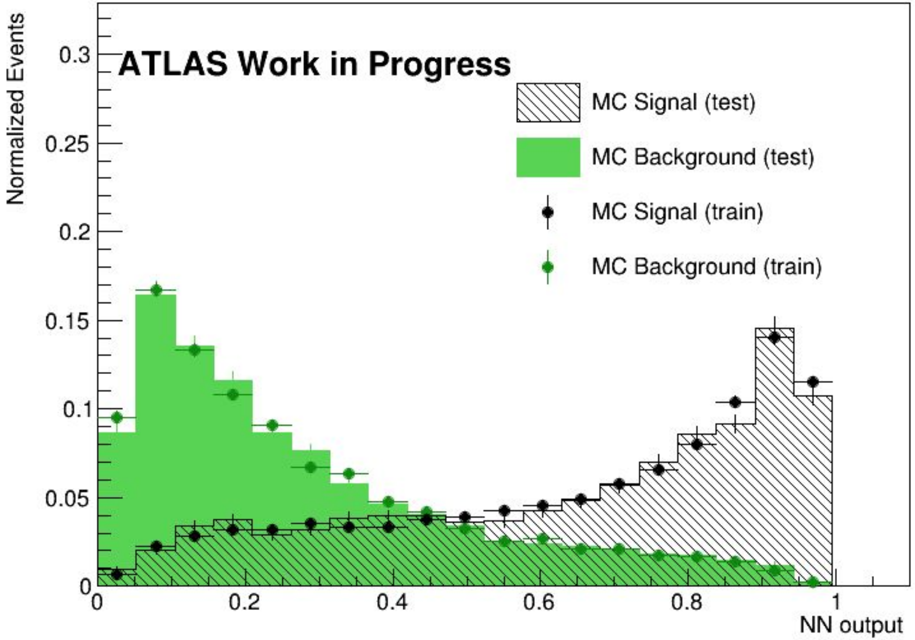
Checking the Method

- Verify method by creating a testing dataset from simulated data.
- Estimated distribution of jets mis-reconstructed as photons (blue line) matches simulated sample (blue points).



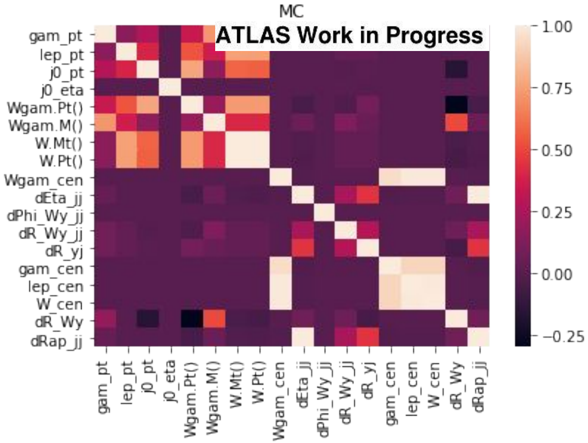
Distinguishing Electroweak Signal from Irreducible Background

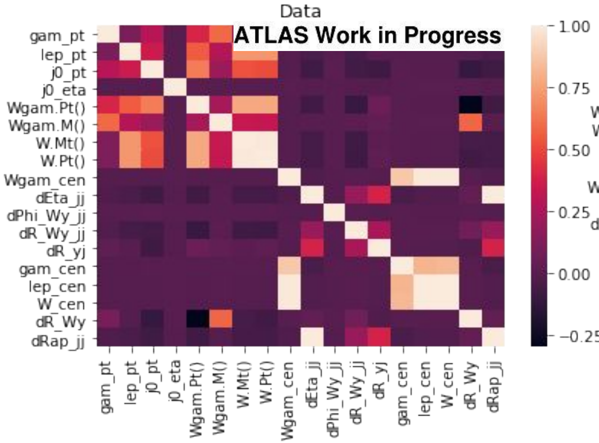
- Large irreducible background from QCD processes.
- To leverage the discriminating power of multiple variables, use them to train a neural network.
    - 5 layer fully connected neural network.
    - Trained on 18 variables.
    - 60% of simulated sample used as training set, 20% used as testing set.
- NN distinguishes signal from QCD background with 69% accuracy.
    - Signal defined as event with NN score > 0.5.

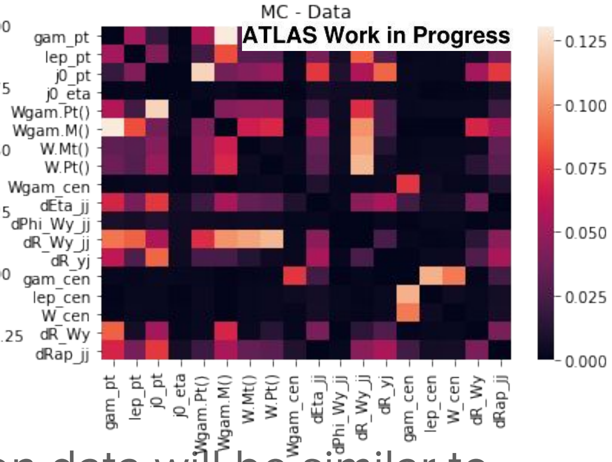


Generalizing performance from MC to data

- The neural network will learn correlations between training variables. Similar performance can be expected when evaluating the model on data if the correlation matrices are similar.



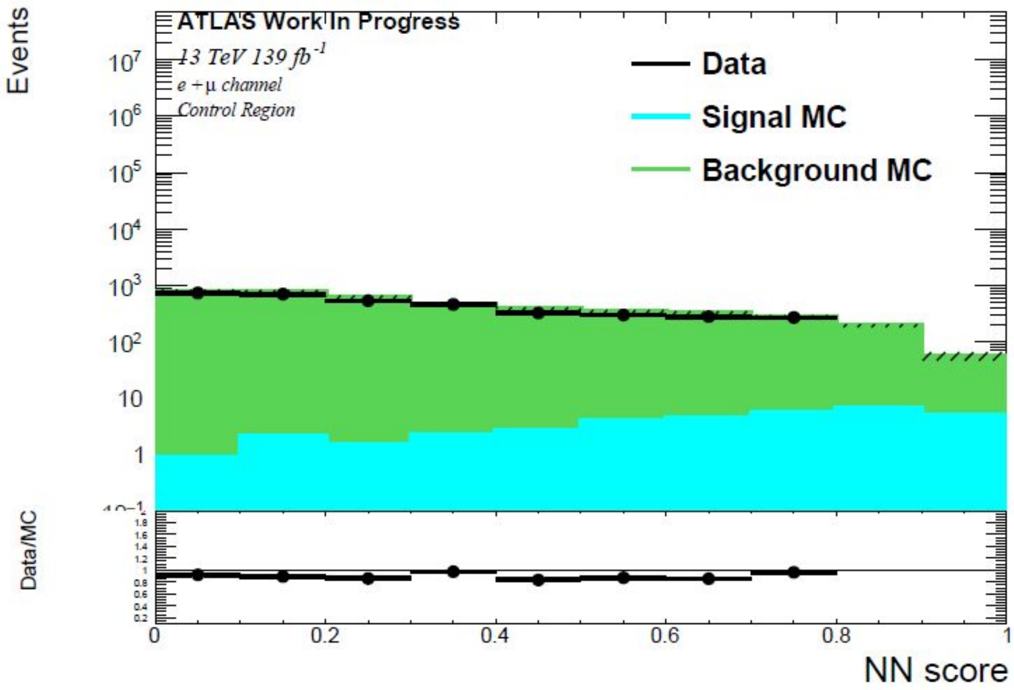




- Maximum difference of .12 suggests performance on data will be similar to performance on simulated data.

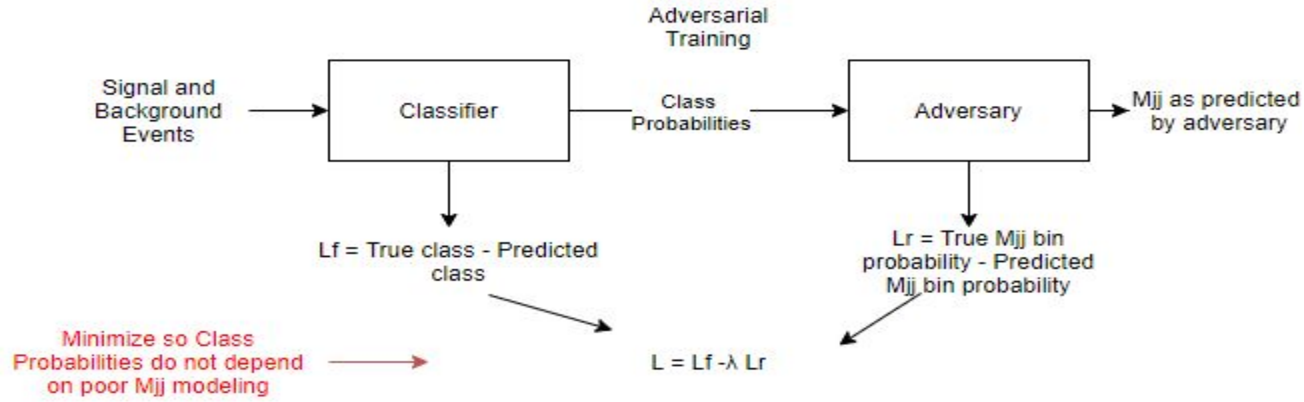
Comparing Simulated and Real Data NN output

- Apply NN to real data in background-enriched control region to verify agreement between data and simulation.
- NN model generalizes well onto real data.



Next Steps: Adversarial Training

- Irreducible QCD background is mismodelled at high values of di-jet invariant mass (M_{jj})
- Train a second neural network to learn the Mjj distribution (predict M_{jj} bins) from the output of the classifier.



Summary

- VBS measurements provide a powerful probe of new physics.
- The search for the vector boson scattering of a photon and a W boson comes with two significant challenges: a large background from jets misidentified as photons and a large irreducible QCD background.
- A data driven approach has been shown to effectively estimate the background from jets faking photons.
- A neural network has been trained to strongly discriminate the electroweak signal from the dominant QCD background.
- New techniques being explored for regularizing the neural network to not learn MC mismodelling.