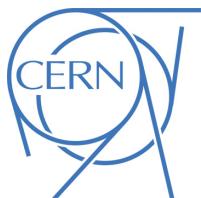




ATLAS Note

ATL-COM-PHYS-2020-XXX

5th January 2020



Draft version 0.1

1

2 Search for flavour-changing neutral currents in top 3 pair events with an associated photon using $\sqrt{s} = 13$ 4 TeV with the ATLAS detector

5

Jason Barkeloo^a and Jim Brau^a

6

^a*University of Oregon*

7

This note presents the search for flavour-changing neutral currents in processes with top pairs where one top decays through the flavour-changing neutral current decay mode (to an up type quark and a photon) and the other decays through the most common standard model mode to a b-quark and a W boson. The W boson then decays leptonically defining the channels searched (electron+jets and muon+jets). Using the entire Run-2 data set of $\sqrt{s} = 13$ TeV data collected using the ATLAS experiment between 2015 and 2018 corresponding to a total integrated luminosity of 139 fb^{-1} . A neural network was developed to separate signal and background events using both low level kinematic variables and high level variable combinations as inputs. The signal events contain an isolated very high p_T photon, a lepton, a b-tagged jet, another light jet, and missing transverse energy. Various data driven techniques were used to estimate contributions to the background from events with a hard scatter photon or a photon faked by either a jet or an electron. As there is no sensitivity for observation an upper limit on the branching ratio of this process is set for each channel and a combination of the channels.

20 © 2020 CERN for the benefit of the ATLAS Collaboration.

21 Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

22 **Contents**

23	1 Change Log	4
24	2 Introduction	5
25	3 Object Reconstruction	7
26	3.1 Electrons	8
27	3.2 Muons	8
28	3.3 Photons	9
29	3.4 Jets	9
30	3.4.1 B-Jets	12
31	3.5 Missing Transverse Energy	13
32	3.5.1 Neutrino Reconstruction	14
33	3.6 Overlap Removal	14
34	3.7 Duplicate Event Removal	14
35	4 Initial Event Selection	15
36	4.1 Triggers	15
37	4.2 Data and MC Pre-Selection Cuts	15
38	4.3 Background Evaluation: Control and Validation Regions	17
39	4.3.1 Backgrounds Without Photons	17
40	4.3.2 Background With Photons	22
41	4.3.3 W+ γ Control Region	22
42	4.3.4 $t\bar{t} + \gamma$ Control Region	24
43	5 Background Estimations	30
44	6 Neural Network for Signal Background Separation	31
45	6.1 Event Classification: Neural Network Optimization	31
46	6.1.1 Input Variables	31
47	6.1.2 Architecture	32
48	6.1.3 Training and Validation of Neural Networks	34
49	6.1.4 Hidden Layer Studies	34
50	6.1.5 B-Tagging Working Point Studies	35
51	7 Analysis	45
52	8 Conclusion	46
53	Appendices	49
54	A Derivation Information (TOPQ1)	49
55	B Complete List of Monte Carlo Samples Used	50
56	C Complete List of Monte Carlo Samples Used	55

57	D Additional Shape Comparison Plots: μ+jets channel	55
58	E Shape Comparison Plots: e+jets channel	55

59 **1 Change Log**

60 **Version 0.1**

- 61 • First Draft

62 2 Introduction

63 The Standard Model of particle physics has proven itself an exceptional and resilient theory since the
 64 combination of the electromagnetic and weak interactions in 1961[1]. Further theoretical work combined
 65 the Higgs Mechanism[2, 3] with the electroweak theory[4, 5]. The resiliency of this theoretical model has
 66 been tested to further degrees of accuracy over the decades with one of my most recent portions being
 67 the experimental confirmation of the Higgs Boson in 2012[6, 7] using the Large Hadron Collider (LHC).
 68 Further precision measurements are ongoing at various experiments at the LHC, including the ATLAS
 69 experiment.

70 However, the Standard Model is known to have flaws and disagreements with nature. For example, the
 71 Standard Model predicts massless neutrinos which is in conflict with experimental observation of neutrino
 72 flavor oscillation and doesn't provide an explanation for dark matter particles or their interactions with
 73 currently known particles. While these large gaps in the Standard Model are well known every precision
 74 measurement made has yet to give any significant new hints toward physics beyond the Standard Model.
 75 One new pathway to look for these hints at the LHC is through top quark decays.

76 The top quark was first observed at Fermilab's Tevatron in 1995[8] but the increase in energy and amount
 77 of data at the LHC has produced orders of magnitude more top quarks than previously seen opening up
 78 a pathway to precision measurements of the properties of the top quark. The top quark is the heaviest
 79 fundamental particle with a mass of 172.51 ± 0.27 (stat) ± 0.42 (syst)[9]. This large mass also means that
 80 top top quark lifetime is very short ($5 * 10^{-25}$ s) and decays before it can hadronize. This allows us to probe
 81 its' branching ratios and decay modes directly. The Standard Model predicts that the top quark decays
 82 through the charged current mode nearly 100% of the time; $t \rightarrow qW$ ($q = b, s, d$)[10]. The Standard Model
 83 also predicts a rare branching ratio of the top quark through a flavor changing neutral current (FCNC)
 84 process, to a neutral boson (photon, Z boson, Higgs Boson, or gluon) and up type quark with a branching
 85 ratio on the order of 10^{-14} .

86 Precision measurements are an important litmus test for the Standard Model. Predicted branching ratios for
 87 FCNC processes in top quark decays are far beyond the experimental reach of the LHC and any observation
 88 of these decay modes would be a sure sign of new physics. Branching ratios are an important measurement
 89 due to a litany of theories for new physics beyond the Standard Model (BSM). These BSM theories predict
 90 enhancements in the top sector by many orders of magnitude such as Minimal Supersymmetric models[11],
 91 R-parity-violating Supersymmetric models[12] and two Higgs doublet models[13] can all increase this
 92 branching ratio many orders of magnitude. Even a null result to a search will set an upper limit on the
 93 branching ratio that can assist in ruling out future physical models based on their amount of large top sector
 94 enrichment.

95 This note presents the ATLAS search for top FCNCs with an associated photon in the decay mode of $t\bar{t}$
 96 events using the entire Run 2 dataset at the LHC, containing combined 2015-2018 datasets taken by the
 97 ATLAS experiment totaling 139 fb^{-1} of integrated luminosity taken at $\sqrt{s} = 13 \text{ TeV}$. The analysis presented
 98 looks for an excess of events coming from top quark pair produced events where one top quark decays to
 99 the most likely decay mode (a bottom quark and W boson) and the other to an up type quark (up or charm)
 100 and a photon. The W boson decay determines the final state channels looked for, two separate channels are
 101 considered: the electron+jets and muon+jets channels.

102 Section ?? a theoretical background will be presented for both the Standard Model with a closer view on
 103 the usual extensions to include the FCNC vertices. Following this Section ?? will discuss the LHC and
 104 the ATLAS experiment used in the creation of the dataset used in the analysis. In Section ?? the special

105 signal simulation requirements will be presented as well as the common background event simulation
106 methodology. The search strategy including the creation of signal, control and validation regions and the
107 training of a neural network will be examined in Section ?? . Section ?? will discuss the results and the
108 conclusions drawn from these results will be presented in Chapter ??.

109 3 Object Reconstruction

110 After the events are simulated, or collected in case of real data, the collections of energy deposits
 111 within the detector systems must be transformed into meaningful physics objects through reconstruction.
 112 Reconstruction is typically done in two major parts using the specialized detectors covered in Chapter ??.
 113 The Inner Detector and Muon System turn patterns of hits within the tracking detectors into tracks that
 114 have direction and momentum information. The calorimeter system transforms the energy deposits within
 115 the calorimeters into calibrated energy deposits with a particular position. These tracks and calorimeter
 116 deposits are used to create physics objects (electrons, muons, etc.) by using particle identification techniques
 117 to reconstruct the underlying physics event. For the analysis presented in this dissertation, the final state
 118 signal particles that need to be reconstructed are one lepton (an electron or a muon), one photon, two
 119 quarks (one light flavor and one b quark), and one neutrino (missing transverse energy as it is the only
 120 particle that does not interact with the detector). Each of these particles has a particular signature in the
 121 subdetectors of the ATLAS detector, shown in Figure 1.

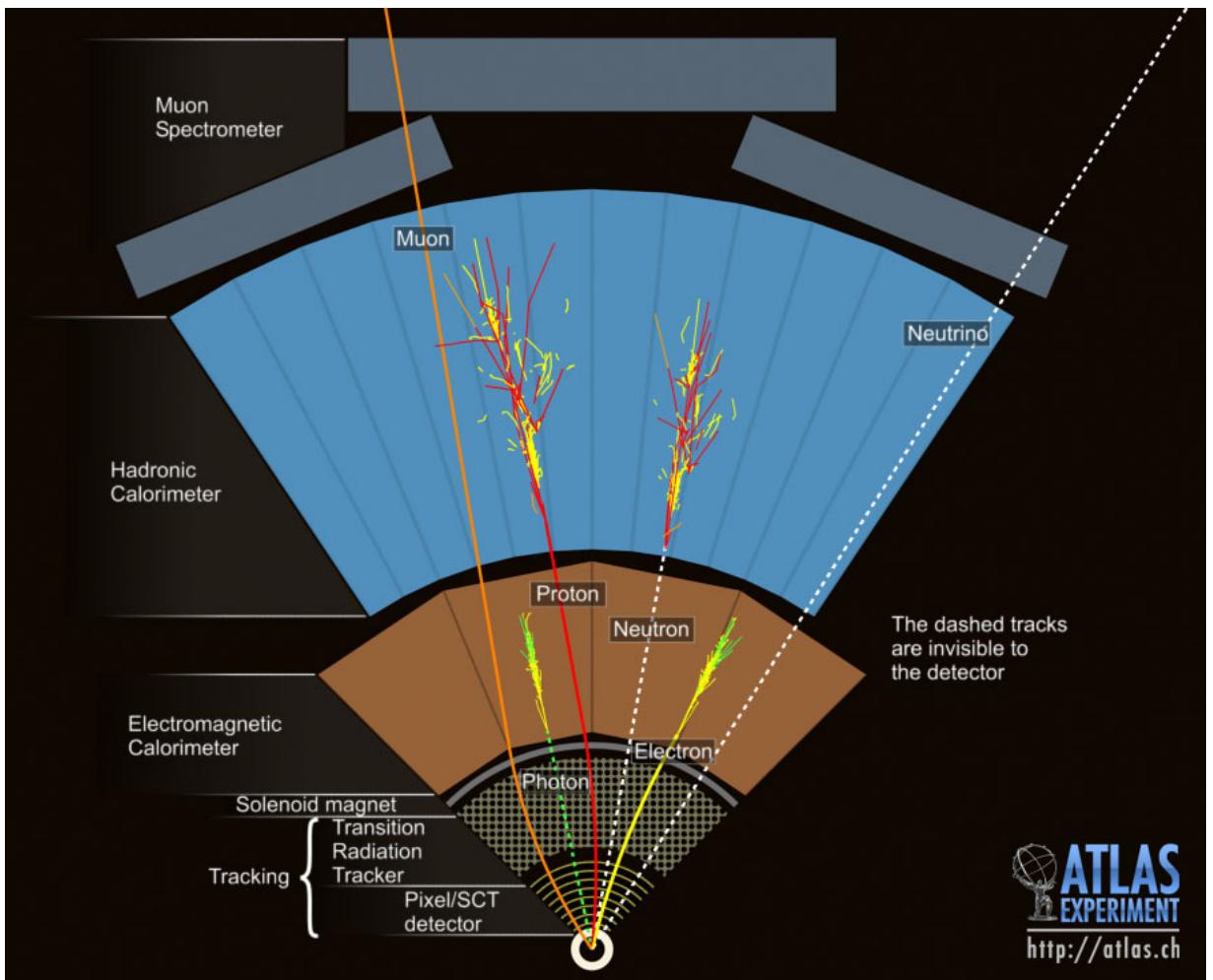


Figure 1: Cross section of a simulated ATLAS detector showing how various particles interact with ATLAS subsystems. Solid lines indicate interactions while dashed lines indicate that no interactions typically occur in that section of the detector. [14]

122 3.1 Electrons

123 Electrons interacting with within the ATLAS detector leave a track in the Inner Detector as well as a
 124 cluster of energy in the electromagnetic calorimeter. The track and cluster are required to be matched
 125 together to be identified as an electron candidate[15]. As electrons move through the detector they create
 126 electromagnetic showers through bremsstrahlung which can produce electron-positron pairs. The process
 127 continues as the particles continue to give energy to the detector. This collection of electrons, positrons,
 128 and photons creates a signature energy cluster in the calorimeter.

129 Electron identification algorithms are applied to the electron candidates to separate prompt and isolated
 130 electron candidates from electrons that come from backgrounds such as converted photons and misidentified
 131 jets. The electron idenification algorithms use a sliding window (3×5 in $\eta \times \phi$) in the high granularity
 132 section of the LAr electromagnetic calorimeter to search for electron cluster “seeds” greater than 2.5 GeV.
 133 Clusters are created around these seeds to form the electromagnetic shower and remove possible duplicate
 134 electron signals by containing them within the cluster. Further pattern recognition for the track fitting
 135 allows even larger amounts of energy into the shower to account for bremsstrahlung in the shower shape.
 136 Tracks and clusters are then matched to give electron candidates.

137 Electrons coming from background jets or photon conversion are called non-prompt as they do not originate
 138 from a signal object/the primary vertex. In order to reject these electrons, other discriminating variables
 139 are used in addition to the track-cluster matching. These variables inclue the amount of energy leakage into
 140 the hadronic calorimeter, thee shower development throughout the electromagnetic calorimeter, and the
 141 amount of radiation measured in the TRT. Three electron identification working points are used: Loose,
 142 Medium, and Tight. Each of these operating points have their own level of background rejection and signal
 143 efficiency. Working points with higher background rejection are a subset of those with lower background
 144 rejection.

145 Isolation variables are another useful tool in the identification of signal electrons from converted photons
 146 produced in hadron decays and light hadron misidentification. These variables are defined by a cone size
 147 around the electron candidate and are the sum of the transverse variable (momentum or energy) of all of
 148 the tracks within the cone, $p_T^{\text{cone}0.2}$ with a cone of $\Delta R = 0.2$ (or $10 \text{ GeV}/E_T$, for high energy electrons) and
 149 $E_{T,\text{Topo}}^{\text{varcone}0.4}$ with a cone defined in a similar manner.

150 Because the LAr calorimeter is a sampling calorimeter, the energy deposits must be calibrated and scaled
 151 such that the true electron energy is read out and not just the small amount of energy deposited into the
 152 active layers as discussed in Section ???. The energy scale is calibrated to be uniform throughout the
 153 detector. Any residual differences between data and simulation are corrected. The calibration strategy was
 154 developed for optimal preformance in LHC Run 1[16] and updated for the conditions of LHC Run 2[17].

155 3.2 Muons

156 Muons behave differently from other particles as they traverse the detector. They act as minimum-ionizing-
 157 particles (MIPs) throughout the calorimeter. The Muon Spectrometer (MS), Section ??, specializes in
 158 precision measurements of muons. The Inner Detector (ID) plays a pivotal role in the identification of
 159 muons as it offers an independent measure of the muon characteristics. The muron reconstruction process
 160 uses a specific set of variables as well[18]. These variables include: *q/p significance*: the difference in
 161 the ratio of track charge and momentum measured with the ID and MS, ρ' : the difference between the

162 transverse momenta measured with the ID and MS, and χ^2 of the combined track fit using tracks from both
 163 the ID and MS.

164 Muons are separated out into four separate types depending on their interactions with the various
 165 subdetectors. The best muon candidates are combined muons that use hits in the MS to trace back to a
 166 track in the ID in order to reconstruct the entire muon track. Segment-tagged muons are muon candidates
 167 that leave a track in the ID but only a segment in the MS instead of a full track. Segment-tagged muons can
 168 occur because of the muon having low p_T or crossing through a region of the MS with reduced acceptance.
 169 Extrapolated muons require only tracks in the MS and are used in regions of η, ϕ phase space that the ID
 170 does not cover. Calorimeter-tagged muons are muons identified by MIPs in the calorimeters and are used
 171 to find muons that cross the ID and MS in regions where cabling might prevent particle detection.

172 Muons also have their own set of isolation criteria which is track-based $p_T^{\text{varcone}0.3}$, with a cone of
 173 $\Delta R = \min(0.3, 10 \text{ GeV}/p_T)$. Similar to electrons various working points are available at the analysis level
 174 for muons. These working points are named similarly: Loose, Medium, Tight, and High- p_T in order of
 175 background rejection.

176 High p_T jets that punch through the hadronic calorimeter can leave tracks in the MS which could be
 177 identified as muons. These would be identified as a bad or a fake muon because of the high-hit multiplicities
 178 they leave in the MS as opposed to a single track left by a muon as it is a MIP. Another source of bad
 179 muons is a mismeasured ID track that gets incorrectly matched to segments in the MS. Fake muons are a
 180 source of fake missing transverse energy, \cancel{E}_T

181 3.3 Photons

182 Photons behave very similarly to electrons in the calorimeter in that they also produce an electromagnetic
 183 shower in the calorimeter. However, they are neutrally charged particles meaning that they should not
 184 leave a track in the ID as they do not bend and produce bremsstrahlung photons traveling through the
 185 magnetic field. Prompt photons pair-produce electrons in the tracker, but this process can be identified as
 186 the associated cluster in the electromagnetic calorimeter is matched to two tracks with opposite charge.
 187 This process produces what is called a converted photon. Unconverted photons have no matching tracks
 188 associated with an electromagnetic cluster.

189

190 Prompt photons produce narrower energy deposits in the electromagnetic calorimeter and have smaller
 191 leakage into the hadronic calorimeter compared to background photons. The energy contained within
 192 narrow structure in $\eta \times \phi$ strips compared to the energy containid in a larger section can help identify prompt
 193 from non-prompt photons [19]. Cuts on this and the other variables listed in Table 1 are tuned to reduce
 194 dependency of identification efficiency on the pileup conditions of Run 2.

195 3.4 Jets

196 Contrasting with electromagnetic showers produced by electrons and photons, hadronic showers form
 197 through QCD processss. Quarks very quickly undergo showering by emitting gluons which further
 198 produce quark-antiquark pairs, analogous to the photons and pair-produced electron-positron pairs of

Category	Description	Name	<i>loose</i>	<i>tight</i>
Acceptance	$ \eta < 2.37$, with $1.37 \leq \eta < 1.52$ excluded	-	✓	✓
Hadronic Leakage	Ratio of E_T in the first sampling layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \eta $ or $ \eta > 1.52$)	R_{had_1}	✓	✓
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \eta < 1.37$)	R_{had}	✓	✓
EM Middle Layer	Ratio of the energy in $3 \times 7 \eta \times \phi$ cells over the energy in 7×7 cells centered around the photon cluster position	R_η	✓	✓
	Lateral shower width, $\sqrt{(\sum E_i \eta_i^2)/(\sum E_i) - ((\sum E_i \eta_i)/(\sum E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	ω_{η_2}	✓	✓
	Ratio of the energy in $3 \times 2 \eta \times \phi$ strips, over the energy of 3×6 cells centered around the photon cluster position	R_ϕ	✓	✓
EM Strip Layer	Lateral shower width, $\sqrt{(\sum E_i(i - i_{\max})^2)/(\sum E_i)}$, where i runs over all strips in a window of $3 \times 2 \eta \times \phi$ strips, and i_{\max} is the index of the highest-energy strip calculated from three strips around the strip with maximum energy deposit	$\omega_{s,3}$		✓
	Total lateral shower width $\sqrt{(\sum E_i(i - i_{\max})^2)/(\sum E_i)}$, where i runs over all strips in a window of $20 \times 2 \eta \times \phi$ strips, and i_{\max} is the index of the highest-energy strip measured in the strip layer	$\omega_{s,\text{tot}}$		✓
	Energy outside the core of the three central strips but within seven strips divided by energy within the three central strips	f_{side}		✓
	Difference between the energy associated with the second maximum in the strip layer and the energy reconstructed in the strip with the minimum value found between the first and second maxima	ΔE_s		✓
	Ratio of the energy difference between the maximum energy deposit and the energy deposit in the secondary maximum in the cluster to the sum of these energies	E_{ratio}		✓
	Ratio of the energy in the first layer to the total energy of the EM cluster	f_1		✓

Table 1: Photon identification variables used for *loose* and *tight* photon identification, taken from [19]

199 electromagnetic showers. When quarks have enough energy they hadronize by producing bound states of
 200 particles. These particles are typically pions or mesons that are measured by the ATLAS detector. The
 201 top quark is the only quark that decays before hadronization because it decays so fast (5×10^{-25} s). The
 202 spray of hadrons coming from a quark from the initial interaction is called a jet and is a collection of
 203 detector objects that are traced back and assigned to the quark(s) in the final state of the interaction. These
 204 algorithms are called jet-finding algorithms. Pictoral representations of the same event reconstructed with
 205 four various algorithms is shown in Figure 2.

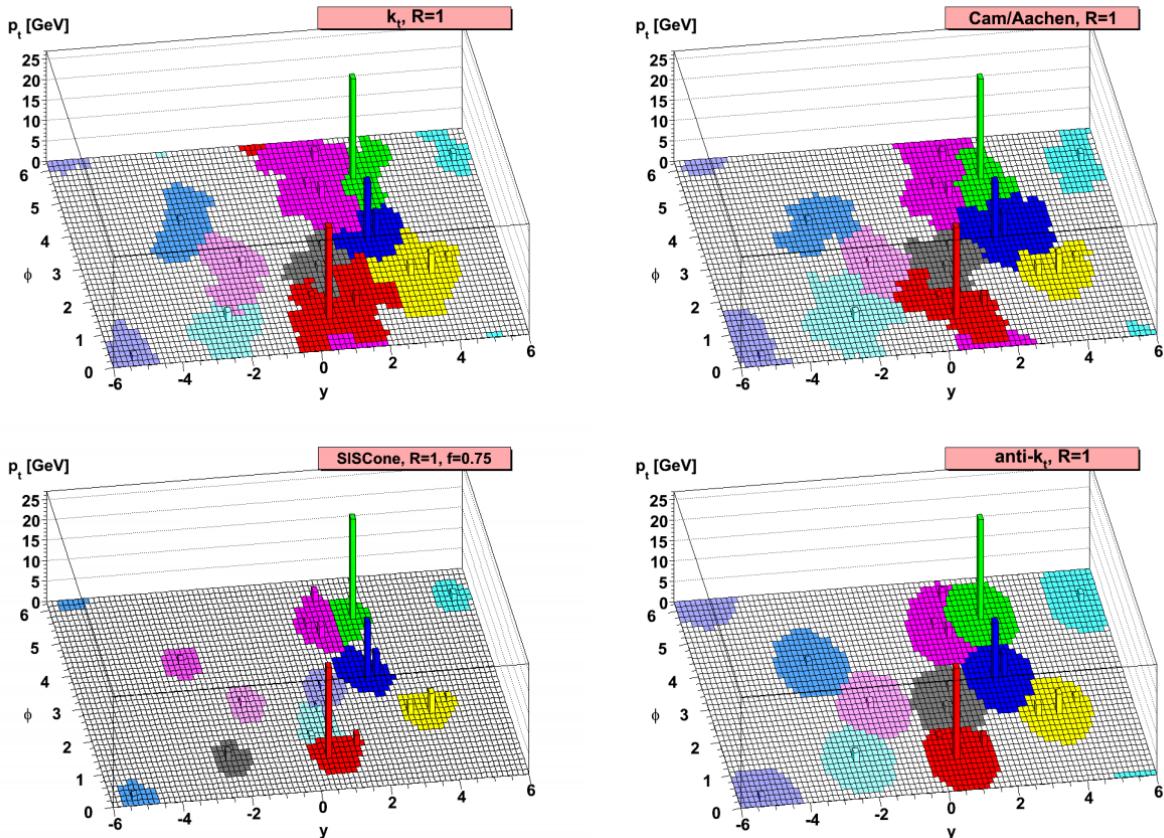


Figure 2: A sample parton-level event with many random soft jet objects, clustered with four different jets algorithms, illustrating the areas of the resulting hard jets. For k_T and Cambridge/Aachen the detailed shapes are in part determined by the specific set of ghosts used, and change when the ghosts are modified [20]

206 The jets in this analysis use the anti- k_T algorithm [20] with a radius parameter $R = 0.4$. Jets are not
 207 physical objects but collections of clustered particles so how they are defined can change the physics objects
 208 that are eventually analyzed. The anti- k_T algorithm is preferred because it is infrared and collinear safe.
 209 Infrared safe jet algorithms do not merge two jets with a soft emission between them. Adding or removing
 210 a soft term between two jets should not change which objects are called jets. Collinear safe jet algorithms
 211 do not change the jet collection if the high transverse momentum particles is split or merged . Another
 212 added benefit of the anti- k_T jet finding algorithm is that it produces roughly circular jet objects, thereby
 213 simplifying the calculation of the energy density and simplifying the calibration of the jet object.

214 The anti- $=k_T$ algorithm calculates the distance between an object i and all possible jet objects j (d_{ij}) and

215 the beam (d_{iB})

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R}, \quad d_{iB} = k_{ti}^{2p}$$

216 where k_{ti} is the transverse momentum, Δ is the distance between the objects, and $p = -1$. This is a general
 217 form for the type of algorithm where the inclusive k_T algorithm has a p value of 1 and the inclusive
 218 Cambridge/Aachen algorithm has a p value of 0 [21]. The algorithm then follows that if d_{ij} is smaller than
 219 d_{iB} then objects i and j are merged, otherwise i is labeled as a jet and removed from the list of entries of
 220 possible jet objects. This is repeated for all entries in the list of possible jet objects.

221 Jet cleaning is also applied to remove events with jets built from known noisy parts of the calorimeter due
 222 to particular calorimeter cells or non-collision background in those areas [22]. To reduce selecting jets that
 223 originate from pileup interactions, another requirement on the jet object is made on the jet vertex tagger
 224 [23, 24] as follows:

- 225 1. For jets with $20\text{GeV} < p_T < 60\text{GeV}$ and $|\eta| < 2.4$: if any jet is bad AND that jet is not marked as
 226 pileup by JVT, then reject the event
- 227 2. For jets with $20\text{GeV} < p_T < 60\text{GeV}$ and $|\eta| \geq 2.4$: if any jet is bad, then reject the event
- 228 3. For jets with $p_T \geq 60\text{GeV}$: if any jet is bad, then reject the event

229 3.4.1 B-Jets

230 While jets originate from any quark, jets coming from b quarks can be identified due to their decay products.
 231 B quarks hadronize into b-hadrons which have a relatively long lifetime compared to many other hadrons
 232 produced from light quarks. The longer lifetime and the relativistic speeds at which the hadrons travel
 233 mean the particle travels a measureable distance before it decays ($400 - 500\mu\text{m}$)[25]. Thus, the vertex
 234 reconstructed from the energy coming from a b hadron decay can be traced back to a point that does not
 235 correspond to the primary vertex of the event. A pictoral representation of a b quark decay is shown in
 236 Figure 3. The b-jet vertex is called the secondary vertex.

237 In addition to the secondary vertex, other variables are helpful in identifying jets coming from b quarks.
 238 By back tracing the tracks within the displaced vertex the minimum distance between the track and the
 239 interaction point can be measured, known as the impact parameter. Reconstructing the decay chain of the
 240 jet is also used in determining the providence of the jet. This information is used in a multivariate analysis
 241 (MVA) to identify jets coming from b quarks and reject jets coming from light quarks.

242 The MVA used in this analysis is the MV2c10, the discriminant used for b-jet identification [27]. The
 243 output distributions for various flavors of jets as well as background rejection and signal efficiency plots are
 244 shown in Figure 4. The c10 in the algorithm name refers to the background training sample of the MVA
 245 consisting of a 10% fraction of c-jets. The 77% efficiency fixed-cut working point for b-jet identification
 246 was chosen for this analysis, discussed in Section 6.1. Differences in efficiency of b-tagging between data
 247 and simulation is taken into account with working point specific scale factors provided by the ATLAS
 248 Flavour Tagging Combined Performance group.

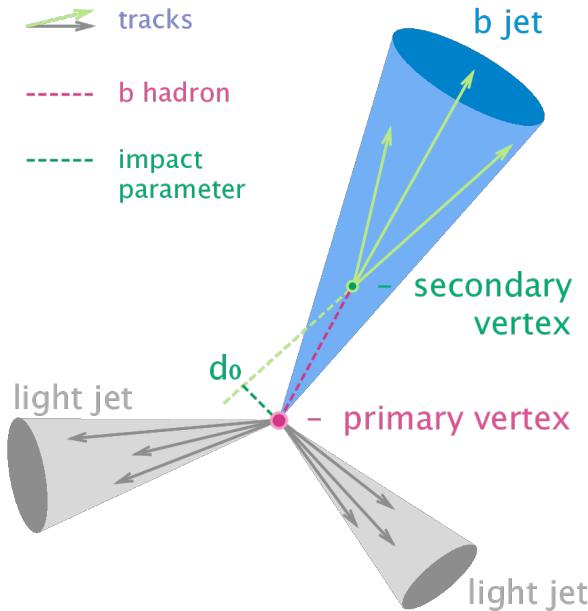


Figure 3: Pictoral representation of an event with a b-jet showing the secondary vertex and impact parameter [26]

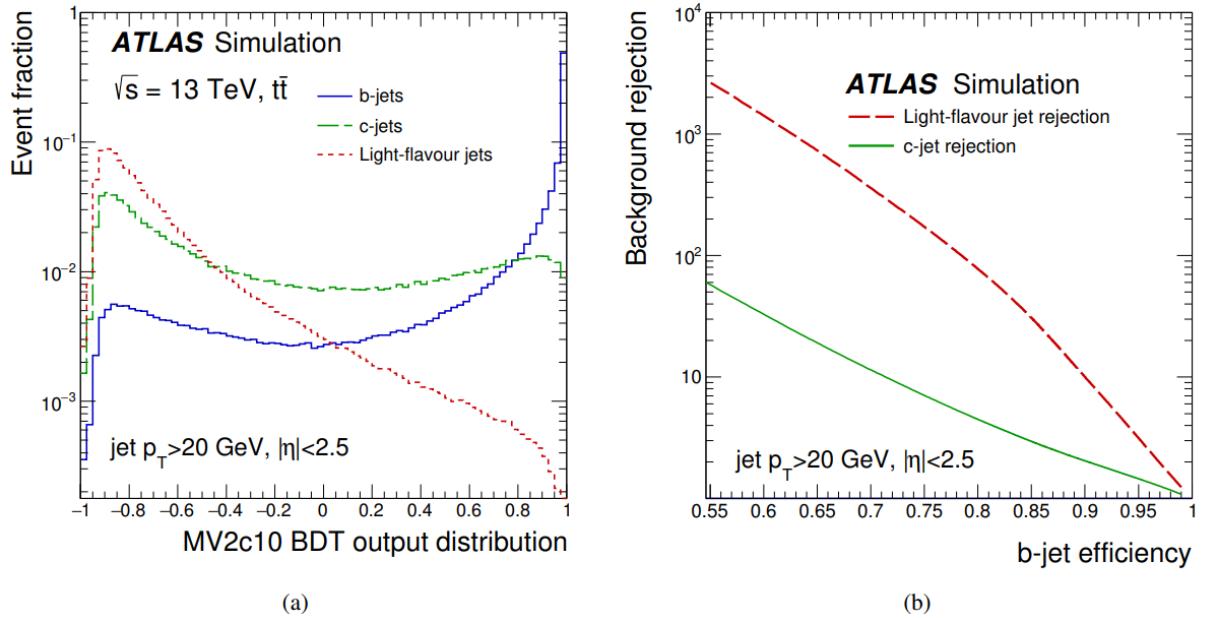


Figure 4: The MV2c10 output for b, c, and light flavored jets in simulated $t\bar{t}$ and the background rejection as a function of the b-jet efficiency [28]

249 3.5 Missing Transverse Energy

250 The remaining signal object that has yet to be discussed is the neutrino coming from the W boson decay.
 251 Neutrinos do not interact with the detectors as they pass through the ATLAS detector. The only way
 252 to measure any properties of the neutrino in ATLAS events is to use conservation of momentum. As
 253 previously mentioned the collision energy is unknown as partons do not carry a consistent fraction of the

beam proton energy. However, in the transverse plane to the beamline the total momentum is known to be very small. Before the collision there is on the order of 1 GeV of momentum in the transverse plane. Therefore, the total transverse momentum of the collision products should be approximately zero.

Any imbalance in the momentum is referred to as Missing Transverse Momentum (\cancel{E}_T). The negative vector sum of all reconstructed objects plus an additional soft term are used to calculate the missing energy in the x-plane and the y-plane[29]. A magnitude and an azimuthal angle are calculated to give the \cancel{E}_T vector in the transverse plane but this does not directly correspond to a neutrino which also has a momentum in the z direction.

3.5.1 Neutrino Reconstruction

In this analysis the signal contains only one source of missing energy, therefore all of the missing energy can be used to reconstruct a neutrino object. There is an ambiguity in the choice of the neutrino z-momentum. To find the z-momentum a χ^2 minimization is done:

$$\chi^2 = \chi_{\text{SMTop}}^2 + \chi_W^2$$

$$\chi^2 = \frac{(m_{\text{bjet},l,\nu} - m_t)^2}{\sigma_{\text{SMTop}}^2} + \frac{(m_{l,\nu} - m_W)^2}{\sigma_W^2}$$

The widths σ_{SMTop} and σ_W^2 are determined from signal Monte Carlo. The event objects are combined to calculate the invariant mass of the top quark (the combination of the b-jet, lepton, and neutrino) and the W boson (combination of the lepton and neutrino). The χ^2 minimization is done while varying the z-momentum of the neutrino. The neutrino momentum that corresponds to the smallest χ^2 value is given assigned to the neutrino object for further use in the analysis. The χ^2 values are also used as a discriminating variable and fed into a neural network (Section 6.1).

3.6 Overlap Removal

Tracks and energy deposits within the detector can, in some cases, be used to reconstruct multiple objects. To prevent using these tracks and deposits multiple times a standard overlap removal procedure is applied to objects. First, electrons that share tracks with any other electrons are removed. Any electron sharing a track with a muon is then also removed. Any jet that is found within $\Delta R < 0.2$ of an electron is removed. Then any jet with less than 3 tracks associated with it within $\Delta R < 0.2$ of a muon object is removed. After that any muon found within $\Delta R < 0.4$ of a jet is removed and any photon within $\Delta R < 0.4$ of an electron or muon object is removed.

3.7 Duplicate Event Removal

As specialized higher statistic samples are used for processes with prompt photons a double counting of events could occur with the nominal MC samples. For example, in addition to the $t\bar{t}$ sample a sample of $t\bar{t} + \gamma$ events are used. This is true for the W+jets/Z+jets and special samples of W+jets+ γ and Z+jets+ γ . Therefore a truth based matching scheme is used to remove events in the nominal samples that match with the photon types produced in the specialized $+ \gamma$ samples i.e., they contain a truth photon that does not originate from a hadron or lepton.

288 4 Initial Event Selection

289 Initial event selection is done to ensure that events that are accepted into the analysis are not contaminated
 290 by extremely noisy detector environments and happened during times when the ATLAS detector was
 291 accepting events properly. All of the events have the same initial set of criteria for determining whether or
 292 not the event is looked at any further for this analysis, applying to both MC and Data. These initial checks
 293 are as follows:

- 294 • Only events occurring during runs good for physics
- 295 • Good Calorimeter status: Ensures that the LAr and Tile calorimeters are not experiencing a noise
 296 burst at the time of the event
- 297 • Requires a primary vertex to be reconstructed for the event which ensures timing of further reconstructed
 298 objects are placed with the correct vertex
- 299 • Global Trigger Decision: Selects events based on whether they passed one of the triggers including
 300 the trigger thresholds, further discussion in Section 4.1
- 301 • Trigger Match: Select events where an electron or muon matches the trigger
- 302 • Overlap Removal as discussed in Section 3.6
- 303 • Ignore events that have a bad muon, occurs mostly in the transition region and the cathode strip
 304 chamber regions.
- 305 • Jet Cleaning: Removes events with jets formed from calorimeter information from sources that have
 306 nothing to do with the energy flow from the initial hard scatter interaction

307 These basic event selection values are applied to every event, in both MC and Data. On top of these
 308 various kinematic cuts are added to form the additional analysis level objects and regions used in the
 309 analysis. These additional kinematic cuts are examined more closely in Section 4.2 and in the discussion of
 310 kinematic region creation throughout the rest of the analysis e.g., Section 4.3.

311 4.1 Triggers

312 Different HLT triggers are used for data taking periods for each year of Run 2. This analysis takes advantage
 313 of single lepton triggers for electrons and muons to dramatically reduce backgrounds due to QCD events
 314 without leptons.

315 4.2 Data and MC Pre-Selection Cuts

316 The Signal Region pre-selection is defined to select events that have an opportunity to enter the final search
 317 selection. This pre-selection selects events with exactly one massive lepton, at least two jets (at least one
 318 of which is b-tagged at the 77% working point), transverse momentum and exactly one photon such that
 319 it resembles the expected final-state topology for the signal. All of the events have the same initial set of
 320 criteria for determining whether or not the event is looked at any further for this analysis, applying to both
 321 MC and Data. These initial checks are as follows:

Year	p_T threshold [GeV]	Identification Menu	Isolation menu	L1 Seed
2015	≥ 24	Medium	None	L1EM20VH
	≥ 60	Medium	None	-
	≥ 120	Loose	None	-
2016-2018	≥ 26	Tight	Gradient (Loose)	-
	≥ 60	Medium	None	-
	≥ 140	Loose	None	-

Table 2: The electron trigger requirements in the event selections

Year	p_T threshold [GeV]	Identification Menu	Isolation menu	L1 Seed
2015	≥ 20	None	Gradient (Loose)	L1MU15
	≥ 50	None	None	-
2016-2018	≥ 26	None	Gradient (Medium)	-
	≥ 50	None	None	-

Table 3: The muon trigger requirements in the event selections

- Exactly 1 lepton (electron or muon) $p_T > 25$ GeV
- At least two good jets ($p_T > 25$ GeV)
- At least one b-tag (MV2c10, 77% Working point)
- $\cancel{E}_T > 30$ GeV and $m_T^W > 30$ GeV (for events with electrons)
- $\cancel{E}_T > 20$ GeV and $\cancel{E}_T + m_T^W > 60$ GeV (for events with muons)
- Exactly 1 photon, $p_T > 50$ GeV

These plots are also produced before additional scale factors to account for mismodelling of various processes are taken into account. These include the fake rate scale factors for processes where a truth electron or hadron are reconstructed to a photon and scaling to account for further mismodeling based on the order of the MC events produced (leading order, next-to-leading order, etc.). Only statistical uncertainties are shown

Signal photons, which originate from a top quark decay, are very high p_T whereas background photons typically result from soft processes. A cut on the photon candidate p_T removes much of the backgrounds while keeping a majority of the signal. The photon p_T in the preselection region is shown in Figure 5.

Other variables of interest, i.e. those being used as inputs into the neural network are also showed in this section. Figure 6 shows the S_T and m_T^W distributions. Figure 7 shows the invariant mass distributions for both top quark candidates, m_{Wb} and $m_{q\gamma}$. The kinematic variables for the electron channel are shown in Figure 8 and for the muon channel in Figure 9. The neural network output of these events are shown in Figure 10

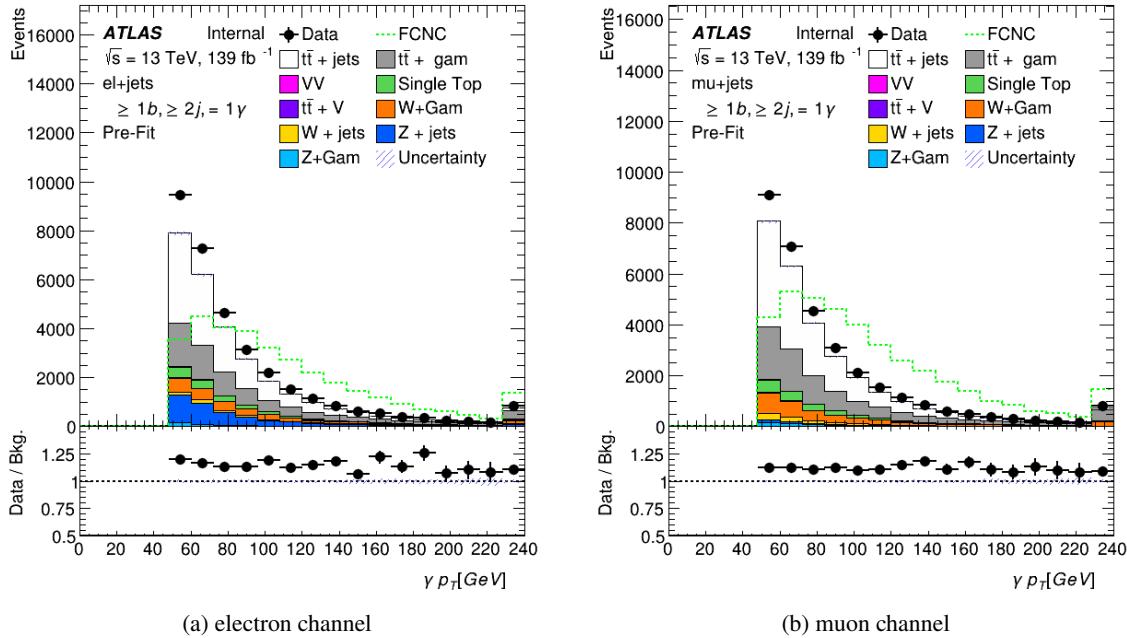


Figure 5: Photon p_T in the signal region pre-preselection region. FCNC signal branching ratio is scaled to 1%.

341 4.3 Background Evaluation: Control and Validation Regions

342 Orthogonal regions to the signal region have been created to test the performance of Monte Carlo samples.
 343 Control and validation regions are designed to isolate specific physics processes to determine and test the
 344 efficacy of scale factors that will be applied to the final signal region Monte Carlo events. These control
 345 and validation regions need to be kinematically similar to the signal region such that derived scale factors
 346 can be translated directly into the signal region and orthogonal to make sure that there is little signal
 347 contamination in the regions. Regions have been made to test the major backgrounds expected in the
 348 signal region: $t\bar{t}$, W+jets, as well as events similar events produced with an associated photon: $t\bar{t} + \gamma$ and
 349 W+Jets+ γ . Events without real photons are described in Section 4.3.1 and regions with a real photon are
 350 described in Section 4.3.2.

351 4.3.1 Backgrounds Without Photons

352 Various background processes that do not have a real photon produced in the events can still enter the signal
 353 region if an electron or jet is mis-reconstructed as a photon. Of these processes the largest contributors in
 354 the signal region are Standard Model $t\bar{t}$ and W+jets. As the LHC attains higher and higher energies the
 355 QCD multijet backgrounds become increasingly hard to model due to the non-perturbative nature of the
 356 interactions. A data-driven technique for accounting for these backgrounds is developed by scaling the
 357 major backgrounds to account for the QCD backgrounds that contribute extra jets to the major backgrounds.
 358 Designing a single control region satisfactorily close to the signal region is impossible. Thus, two control
 359 regions are designed, one which is W+jets rich and the other $t\bar{t}$ rich. Scale factors for these backgrounds
 360 are derived simultaneously and tested in a third similar region for validation before being applied to other
 361 regions. These control and validation regions are defined as follows:

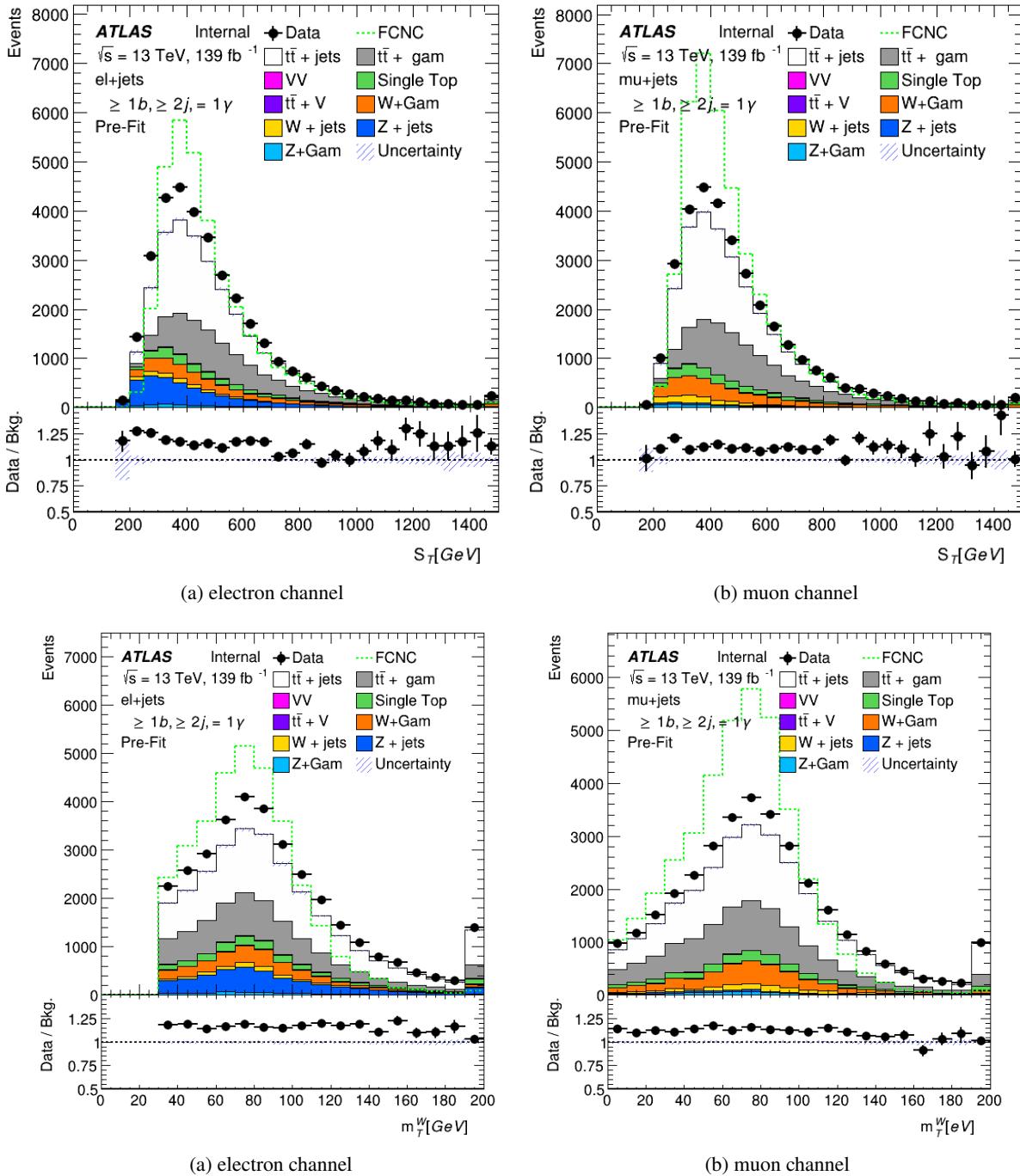


Figure 6: S_T and m_T^W in the signal region pre-selection region. FCNC signal branching ratio is scaled to 1%.

- All of the Initial Event Selection as outlined in Section 4
 - Exactly 1 lepton (electron or muon) $p_T > 25 \text{ GeV}$
 - Number of Jets ($p_T > 25 \text{ GeV}$) to define the regions

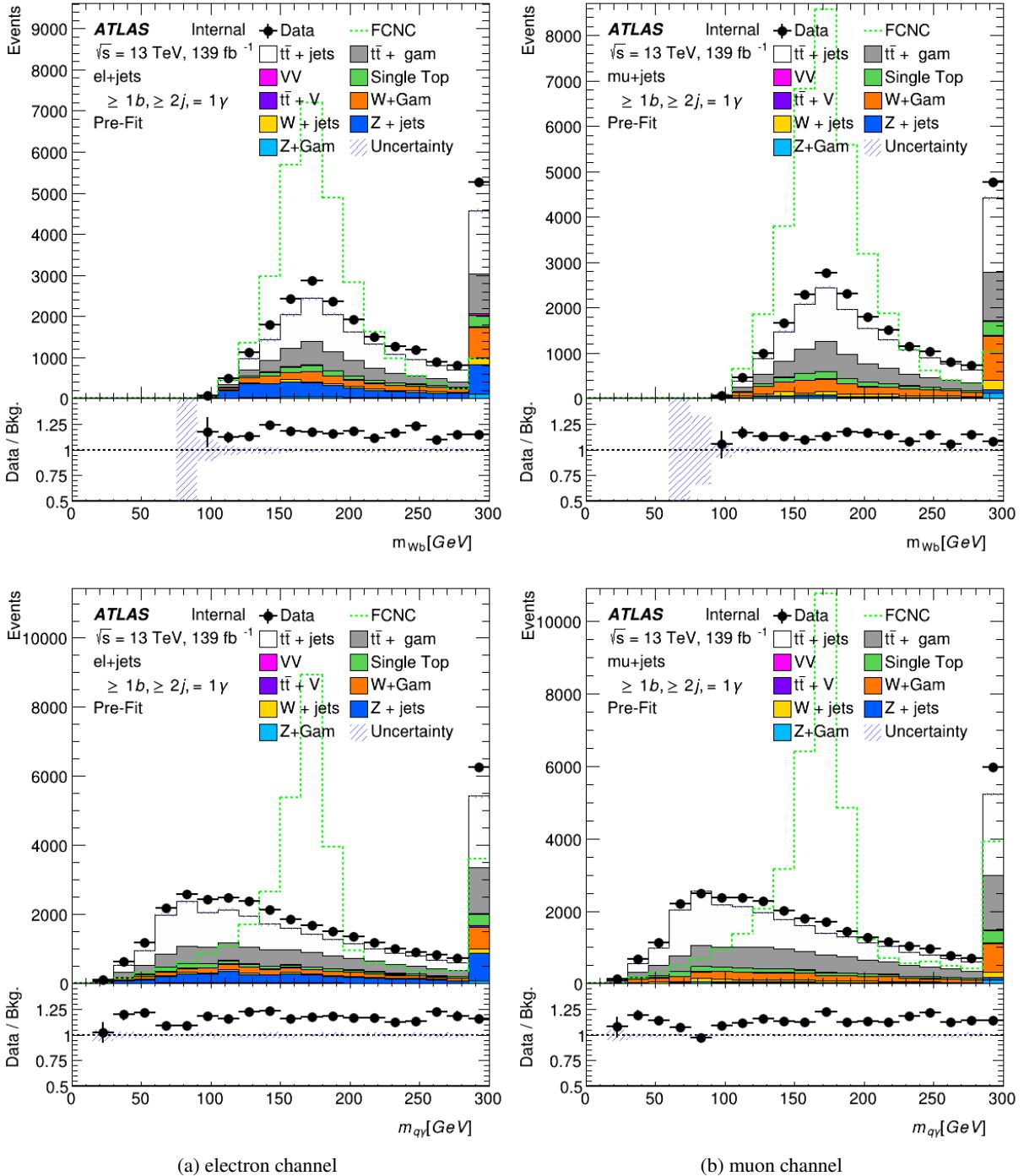


Figure 7: Top mass candidates in the signal region pre-selection: m_{Wb} and m_{qy} . FCNC signal branching ratio is scaled to 1%.

- Control Region 1 (W+Jets enriched): $n_{\text{jets}} = 3$
 - Validation Region: $n_{\text{jets}} = 4$

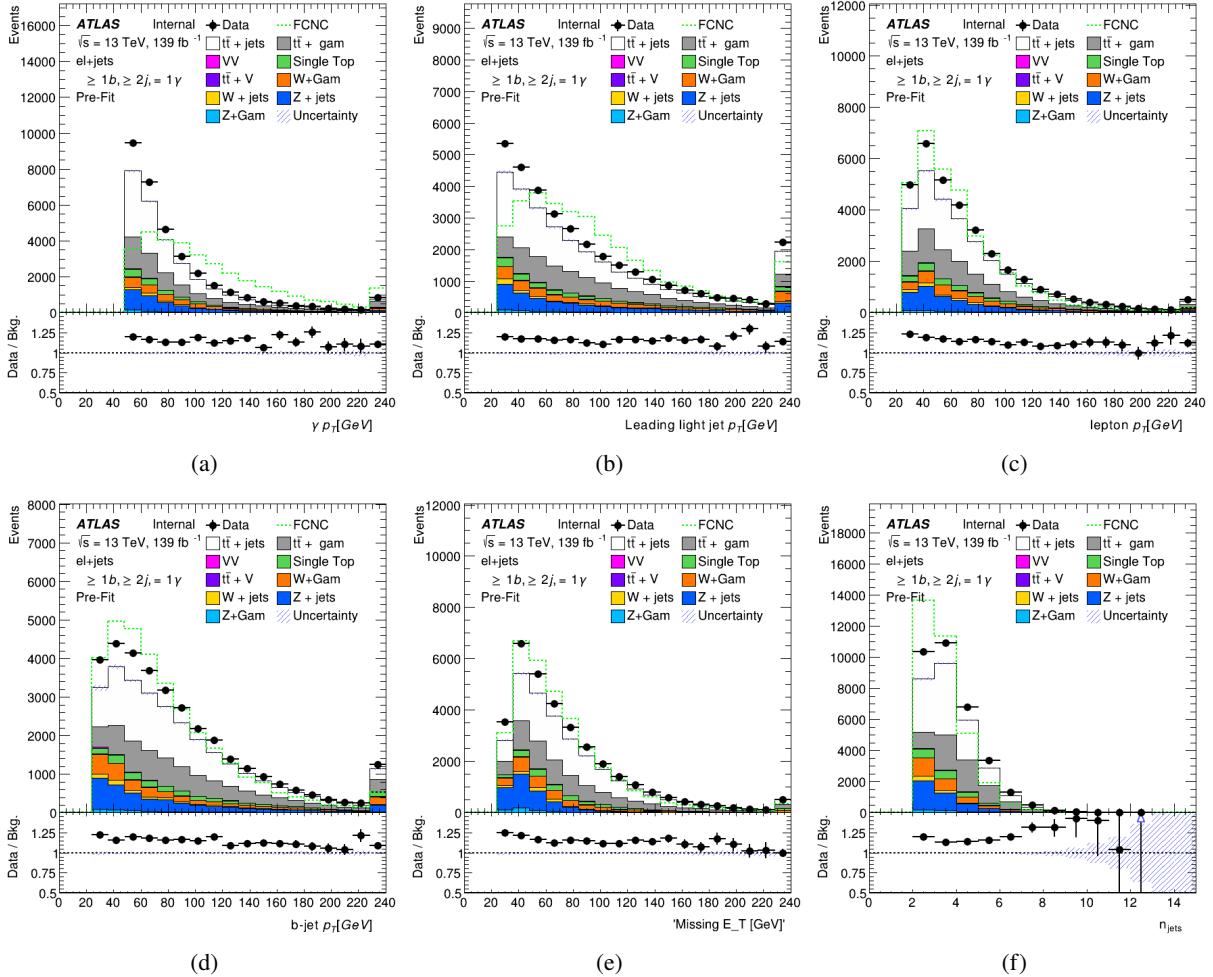


Figure 8: Photon p_T (a), leading light jet p_T (b), lepton p_T (c), b-jet p_T (d), \cancel{E}_T (e), and n_{jets} (f) plots in the signal region pre-selection for the electron+jets channel. FCNC signal branching ratio is scaled to 1%.

367 – Control Region 2 ($t\bar{t}$ enriched): $n_{\text{jets}} \geq 5$

- 368 • $\cancel{E}_T > 30 \text{ GeV}$ and $m_T^W > 30 \text{ GeV}$ (for events with electrons)
- 369 • $\cancel{E}_T > 20 \text{ GeV}$ and $\cancel{E}_T + m_T^W > 60 \text{ GeV}$ (for events with muons)
- 370 • Exactly 1 b-tagged jet (MV2c10, 77% Working point)
- 371 • 0 photons, $p_T > 15 \text{ GeV}$

372 The efficiency of scale factors derived using control regions 1 ($n_{\text{jets}}=3$) and 2 ($n_{\text{jets}} \geq 5$) are then tested in
373 the validation region ($n_{\text{jets}}=4$). The scale factors for the $t\bar{t}$ and W+jets MC are derived using:

$$\begin{bmatrix} N(W)_{3j} & N(t\bar{t})_{3j} \\ N(W)_{5+j} & N(t\bar{t})_{5+j} \end{bmatrix} \begin{bmatrix} W_{SF} \\ t\bar{t}_{SF} \end{bmatrix} = \begin{bmatrix} N(\text{data-bkg})_{3j} \\ N(\text{data-bkg})_{5j} \end{bmatrix}$$

374 Figure 11 shows the S_T distribution in both electron and muon channels before scale factors are calculated
375 for all three kinematically separate regions. The large mismodelling occurs at low S_T values as expected as

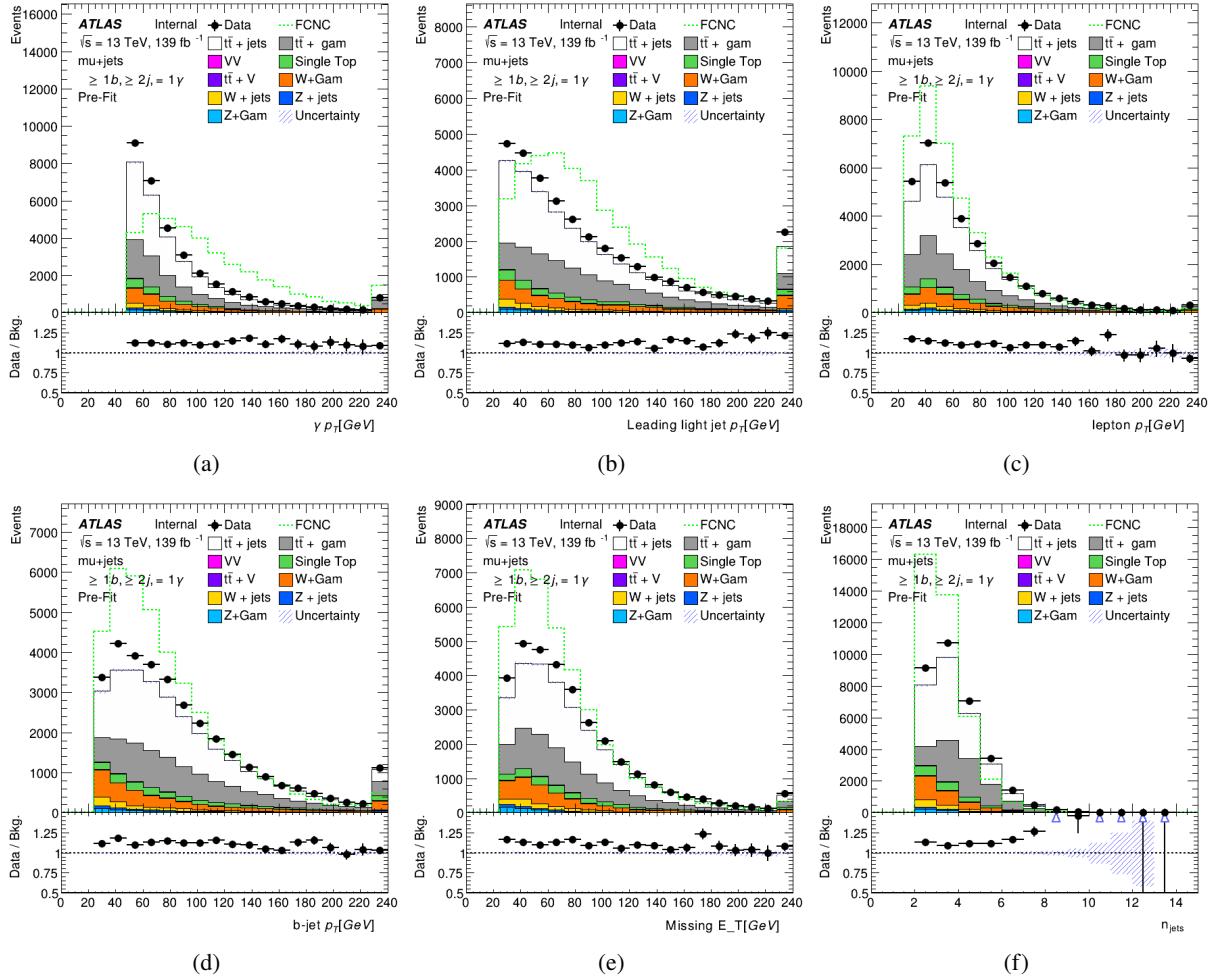


Figure 9: Photon p_T (a), leading light jet p_T (b), lepton p_T (c), b-jet p_T (d), \cancel{E}_T (e), and n_{jets} (f) plots in the signal region pre-selection for the muon+jets channel. FCNC signal branching ratio is scaled to 1%.

376 QCD processes will typically add low energy jets to the events. The Figures 12(electron channel) and
 377 13(muon channel) show various event-level variable plots for the validation region after the scale factors
 378 have been applied.

The derived scale factors using these regions are shown in Table 4

Sample	e+jets SF	μ +jets SF
W+jets	1.22	1.25
$t\bar{t}$	1.06	1.01

Table 4: Derived $t\bar{t}$ and W+jets scale factors for QCD multijet backgrounds.

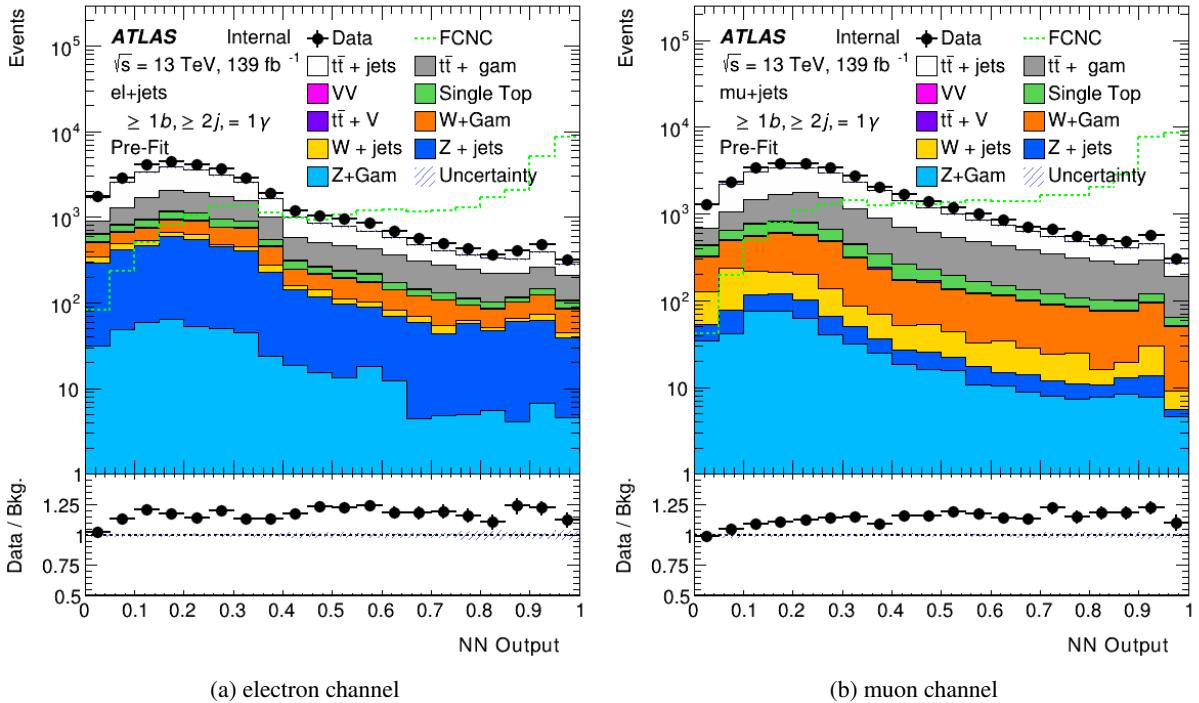


Figure 10: Output of the Neural Network in the signal region pre-selection region. FCNC signal branching ratio is scaled to 1%.

380 4.3.2 Background With Photons

381 Standard Model processes that are produced with an extra real photon are an irreducible background for
 382 this search as they can share the same final state as the signal events. The largest contributors of these
 383 irreducible backgrounds are the major background samples discussed in the previous section with an
 384 associated photon ($t\bar{t} + \gamma$ and $W+jets+\gamma$). Special Monte Carlo samples are produced for these samples
 385 (along with $Z+jets+\gamma$) that have higher statistics of these photon enriched events than the nominal samples.
 386 However, as these samples ($X+jets+\gamma$) are subsets of the nominal sample($X+jets$) duplicate events must
 387 be removed from the nominal sample. This is done using the **MCTruthClassifier** tool which is detailed
 388 further in Section ???. All events with a photon from the hard scattering are removed from the $X+jets$
 389 samples as they are contained within the $X+jets+\gamma$ samples.

390 4.3.3 $W+\gamma$ Control Region

391 A validation region for $W+jets+\gamma$ was created as it is one of the more dominant backgrounds other than $t\bar{t}$
 392 and $t\bar{t} + \gamma$ events. The normalization for the $W+jets+\gamma$ validation region enters as a free parameter into the
 393 final fit. The region selection for the $W+jets+\gamma$ is as follows:

- 394 • All of the Initial Event Selection as outlined in Section 4
- 395 • Exactly 1 lepton (electron or muon) $p_T > 25$ GeV

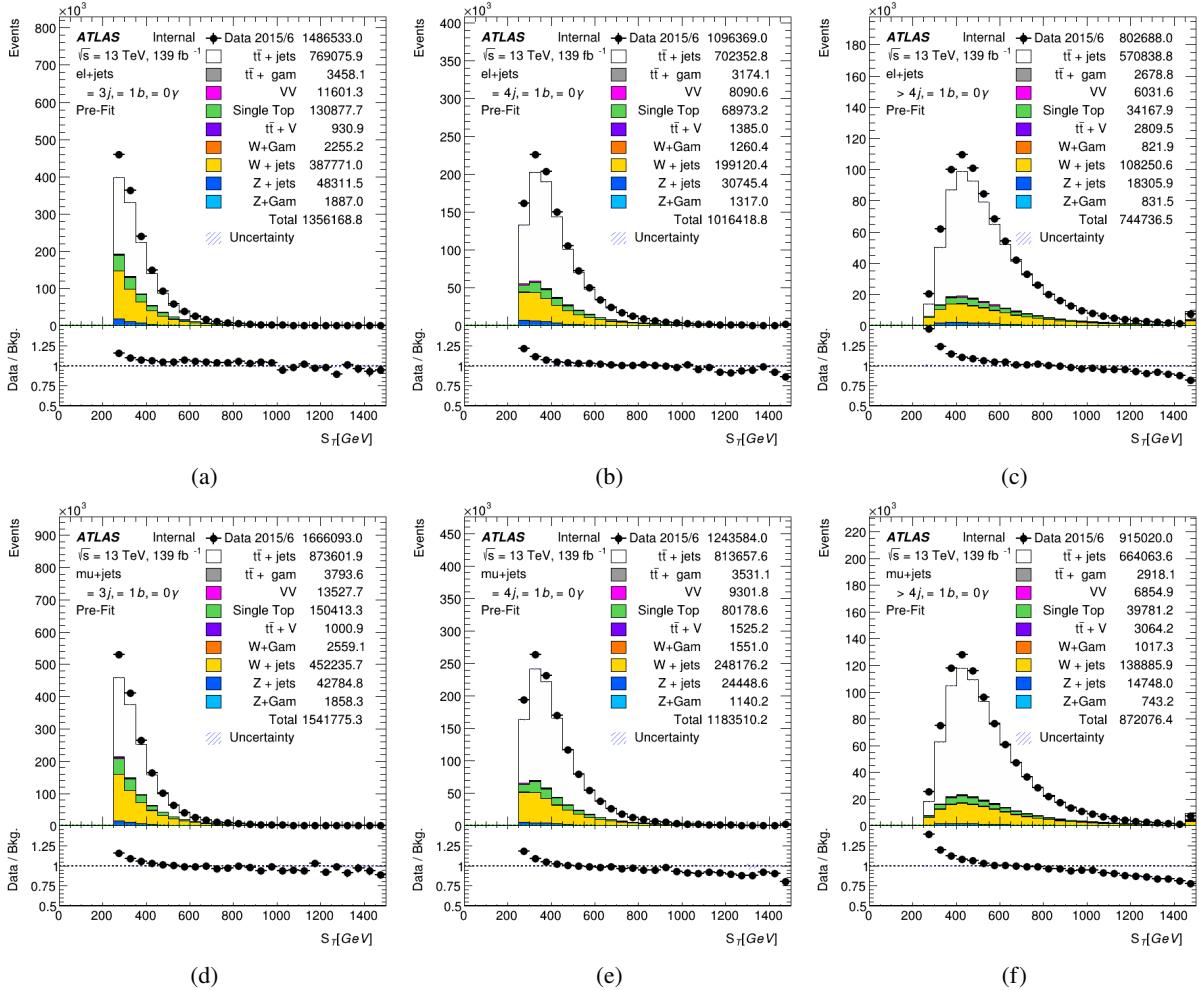


Figure 11: S_T distributions in the 3(a,d), 4(b,e), and 5+(c,f) jets control and validation regions. Electron channel is shown on the top and the muon channel on the bottom, before scale factors are determined.

- 396 • At least 2 Jets ($p_T > 25 \text{ GeV}$)
- 397 • $\cancel{E}_T > 30 \text{ GeV}$ and $m_T^W > 30 \text{ GeV}$ (for events with electrons)
- 398 • $\cancel{E}_T > 20 \text{ GeV}$ and $\cancel{E}_T + m_T^W > 60 \text{ GeV}$ (for events with muons)
- 399 • Exactly 0 b-tagged jet (MV2c10, 77% Working point)
- 400 • Exactly 1 photons, $p_T > 50 \text{ GeV}$
- 401 • Photon isolation cuts: $\text{topo}E_T \text{cone}40 < 4 \text{ GeV}$
- 402 • Z mass cut $|m_{l\gamma} - m_Z|V > 5 \text{ GeV}$

403 Distributions of kinematic variables in the electron (muon) channels are shown in Figure 14 (15).

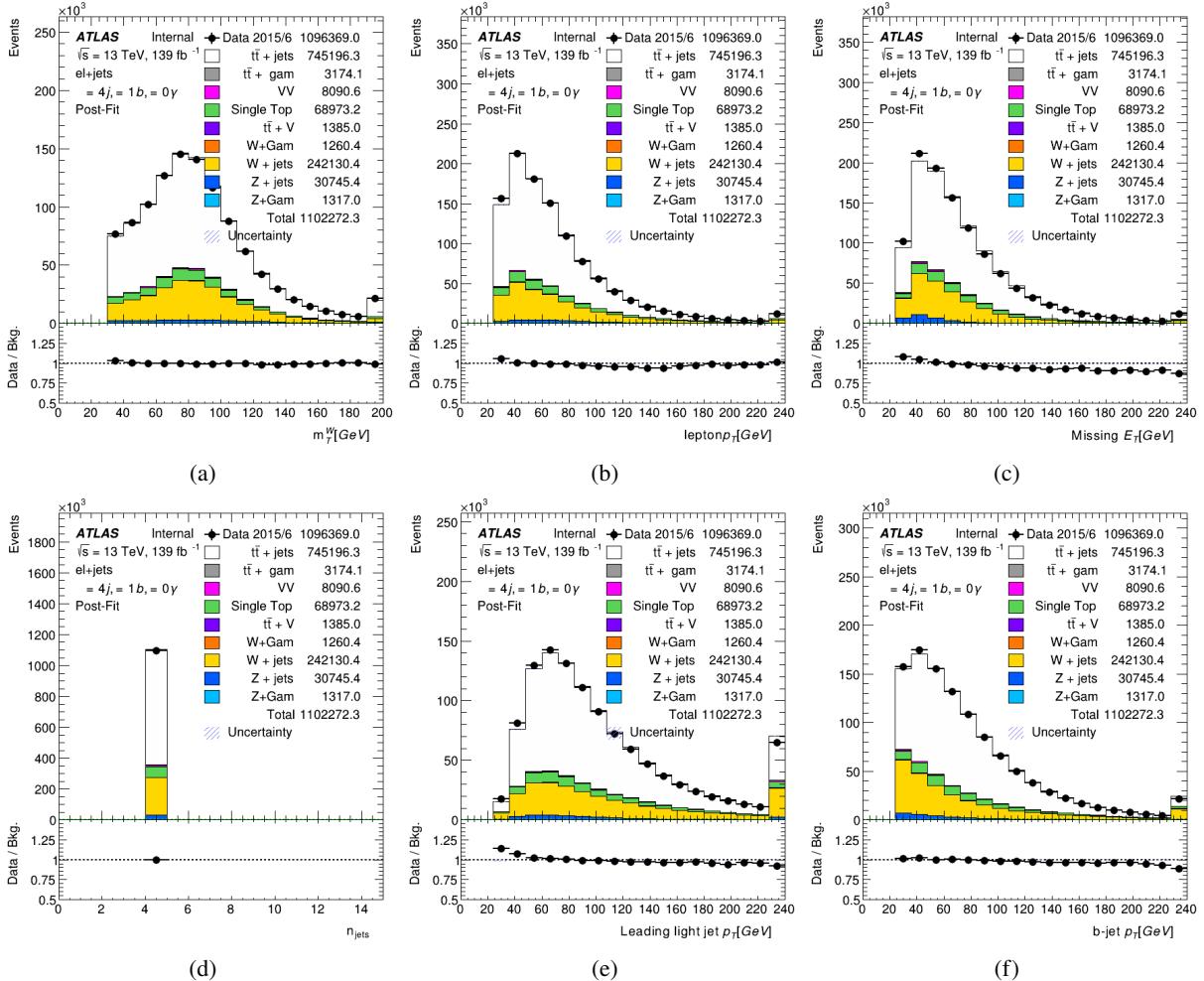


Figure 12: Event-level plots for the =4 jet validation region after scale factors have been applied in the electron channel. FCNC signal branching ratio is scaled to 1%.

Channel:	e+jets	μ +jets
W+jets+ γ SF	PUT IN VALUE	PUT IN VALUE

Table 5: Fit result W+jets+ γ normalization scale factors including both statistical and systematic uncertainties.

4.3.4 $t\bar{t} + \gamma$ Control Region

Another validation region was created for the other largest photon enriched samples, $t\bar{t} + \gamma$. The normalization for the $t\bar{t} + \gamma$ validation region enters as a free parameter into the final fit and the region selection is as follows:

- All of the Initial Event Selection as outlined in Section 4
- Exactly 1 lepton (electron or muon) $p_T > 25$ GeV

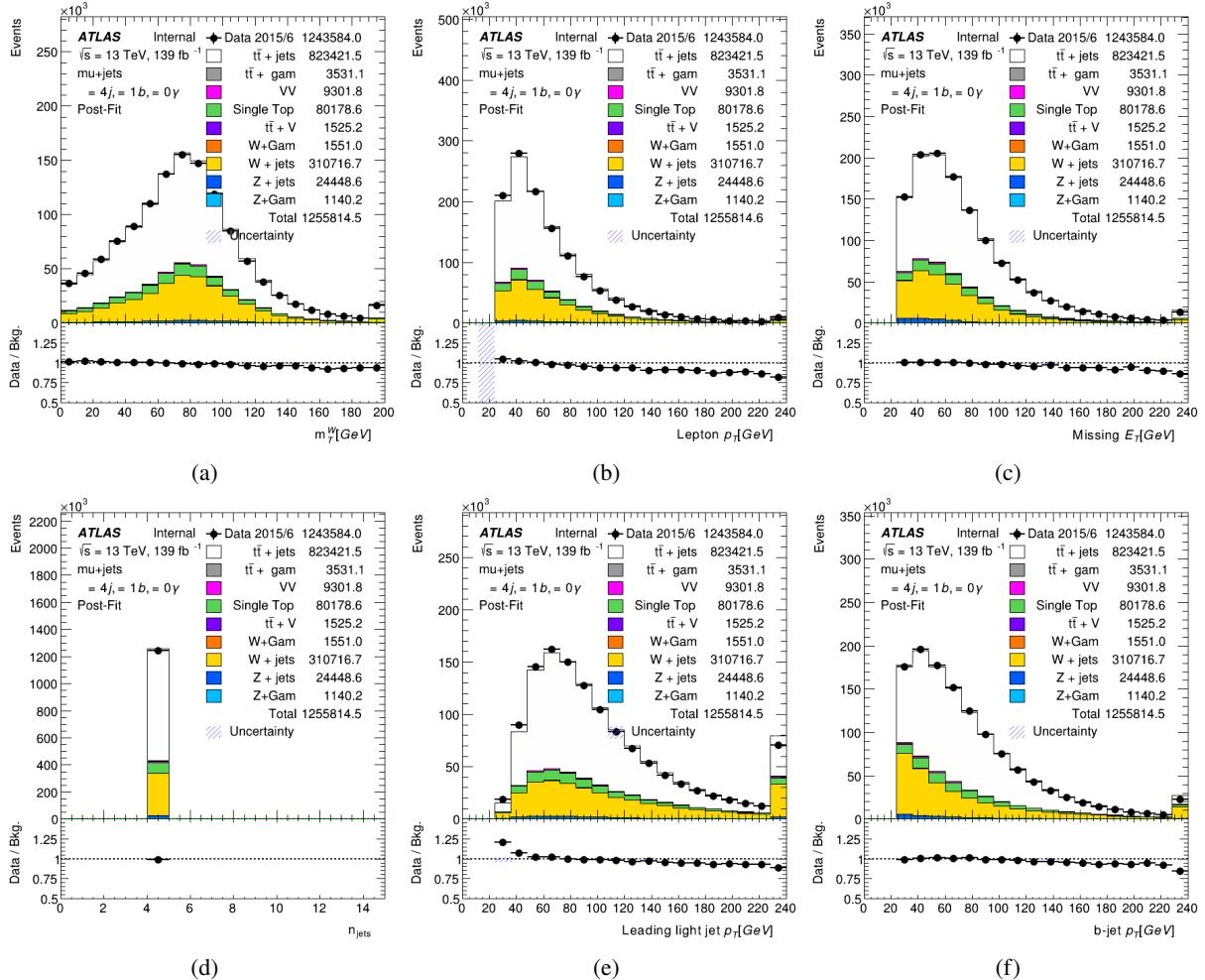


Figure 13: Event-level plots for the =4 jet validation region after scale factors have been applied in the muon channel. FCNC signal branching ratio is scaled to 1%.

- At least 4 Jets ($p_T > 25 \text{ GeV}$)
- $\cancel{E}_T > 30 \text{ GeV}$ and $m_T^W > 30 \text{ GeV}$ (for events with electrons)
- $\cancel{E}_T > 20 \text{ GeV}$ and $\cancel{E}_T + m_T^W > 60 \text{ GeV}$ (for events with muons)
- At least 1 b-tagged jet (MV2c10, 77% Working point)
- Exactly 1 photons, $p_T > 50 \text{ GeV}$
- Photon isolation cuts: $\text{topo}E_T \text{cone}40 < 4 \text{ GeV}$
- Reverse Neural Network Cut: $\text{NNOutput} < 0.93$ electron channel, $\text{NNOutput} < 0.92$ muon channel

417 Distributions of kinematic variables in the electron (muon) channels are shown in Figure 16 (17).

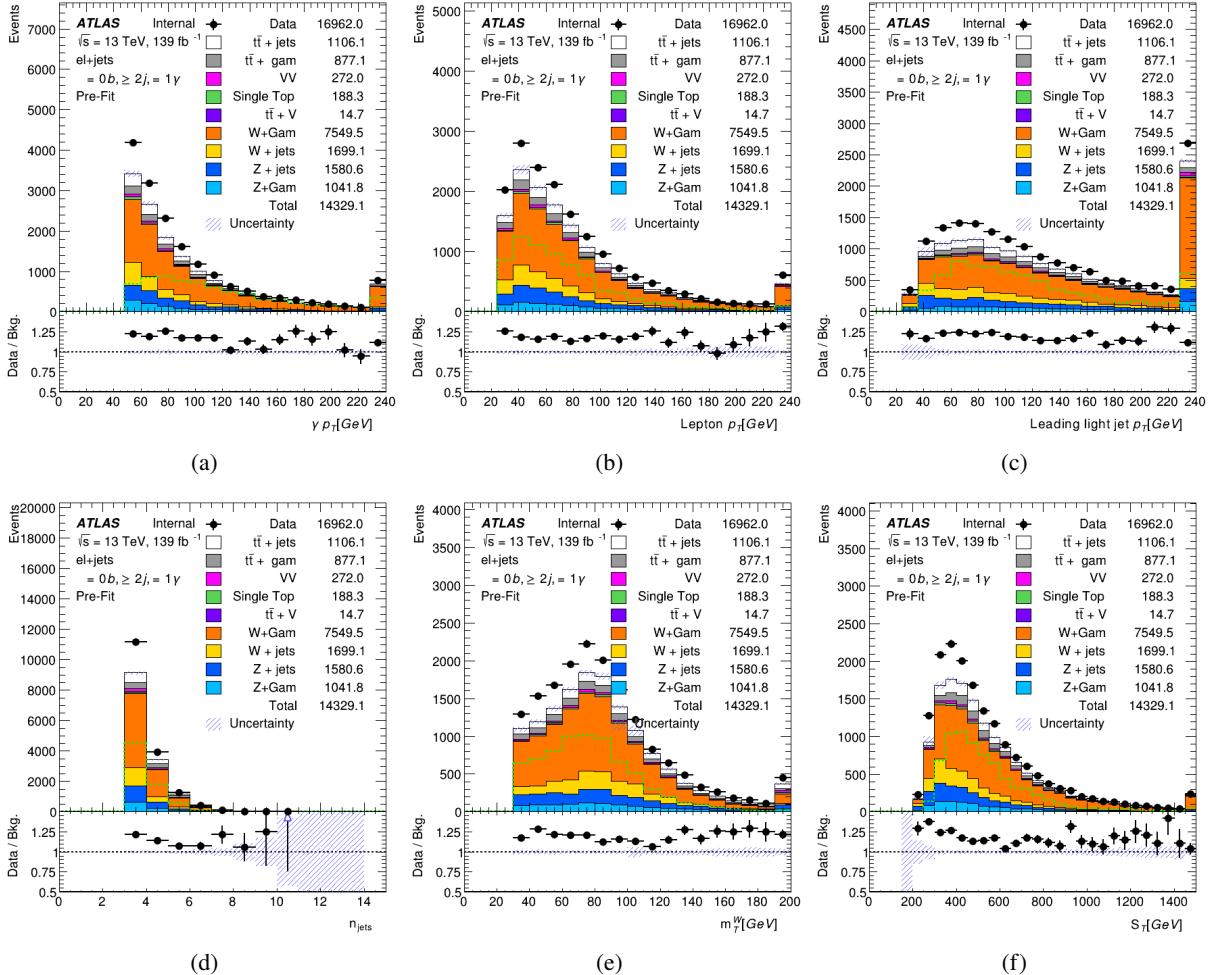


Figure 14: $W+\text{jets}+\gamma$ validation region plots for the electron channel. The FCNC signal sample is scaled to 1%.

Channel:	e+jets	$\mu+\text{jets}$
$t\bar{t} + \gamma$ SF	PUT IN VALUE	PUT IN VALUE

Table 6: Fit result $t\bar{t} + \gamma$ normalization scale factors including both statistical and systematic uncertainties.

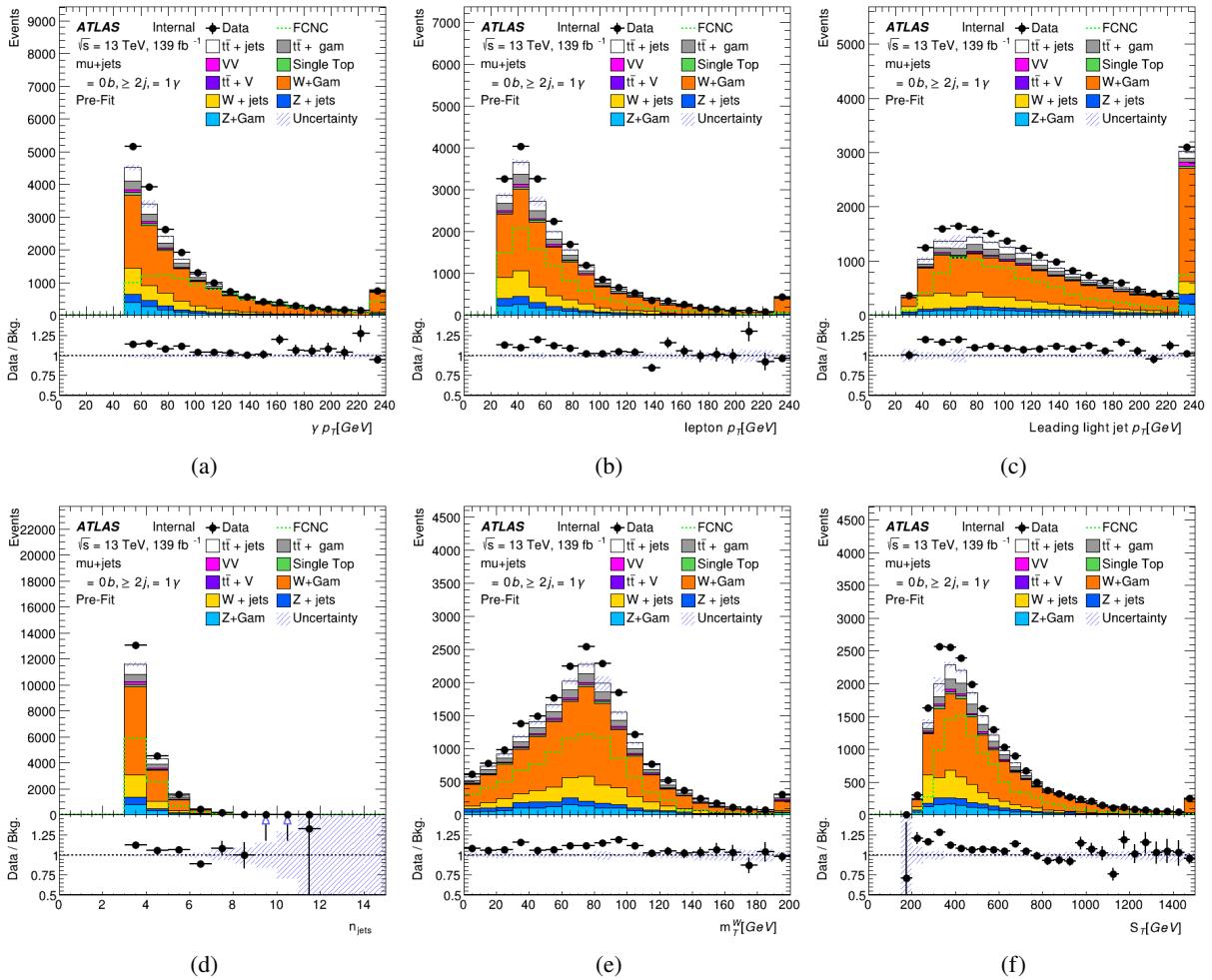


Figure 15: $W+jets+\gamma$ validation region plots for the muon channel. The FCNC signal sample is scaled to 1%.

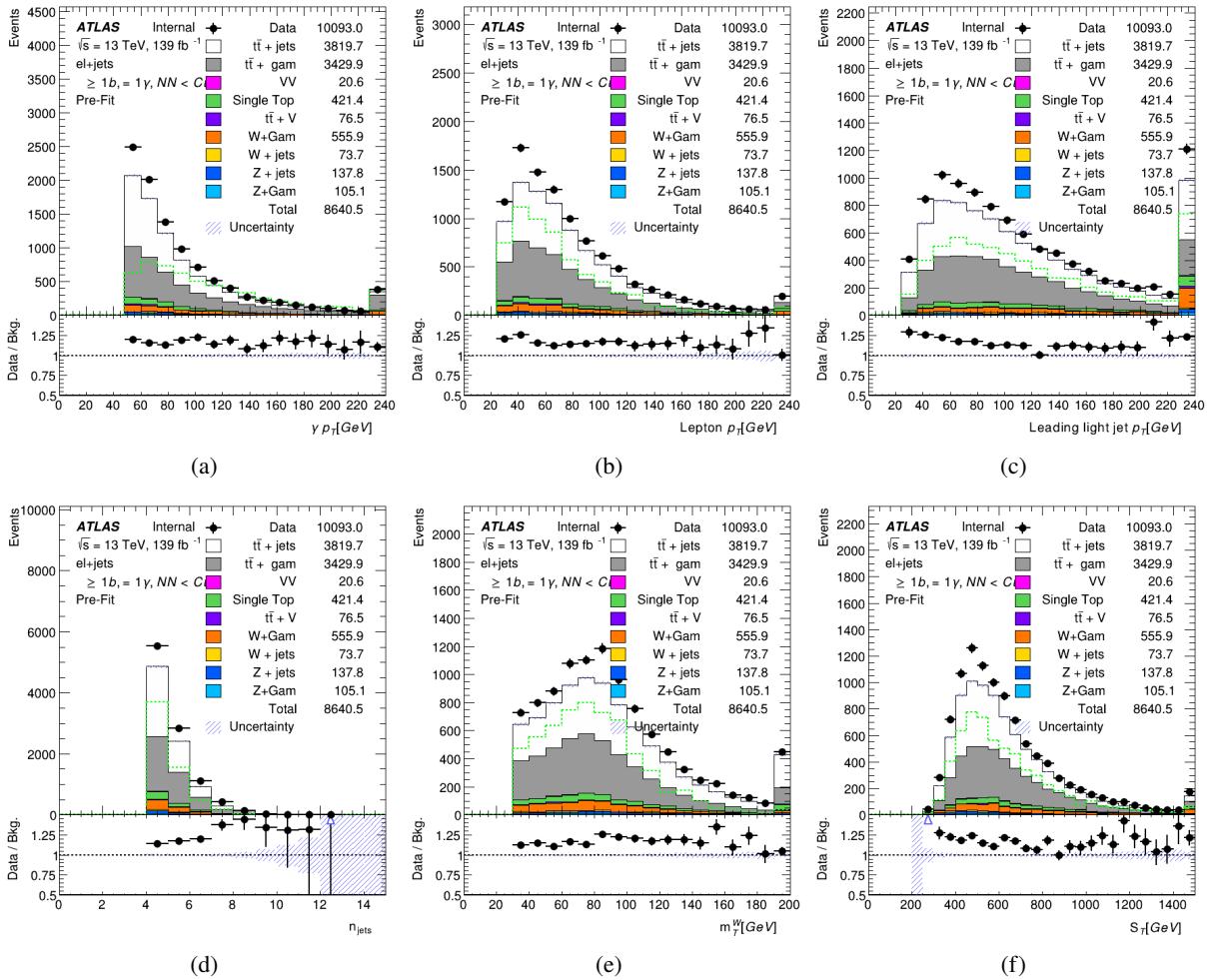


Figure 16: $t\bar{t} + \text{jets} + \gamma$ validation region plots for the electron channel. The FCNC signal sample is scaled to 1%.

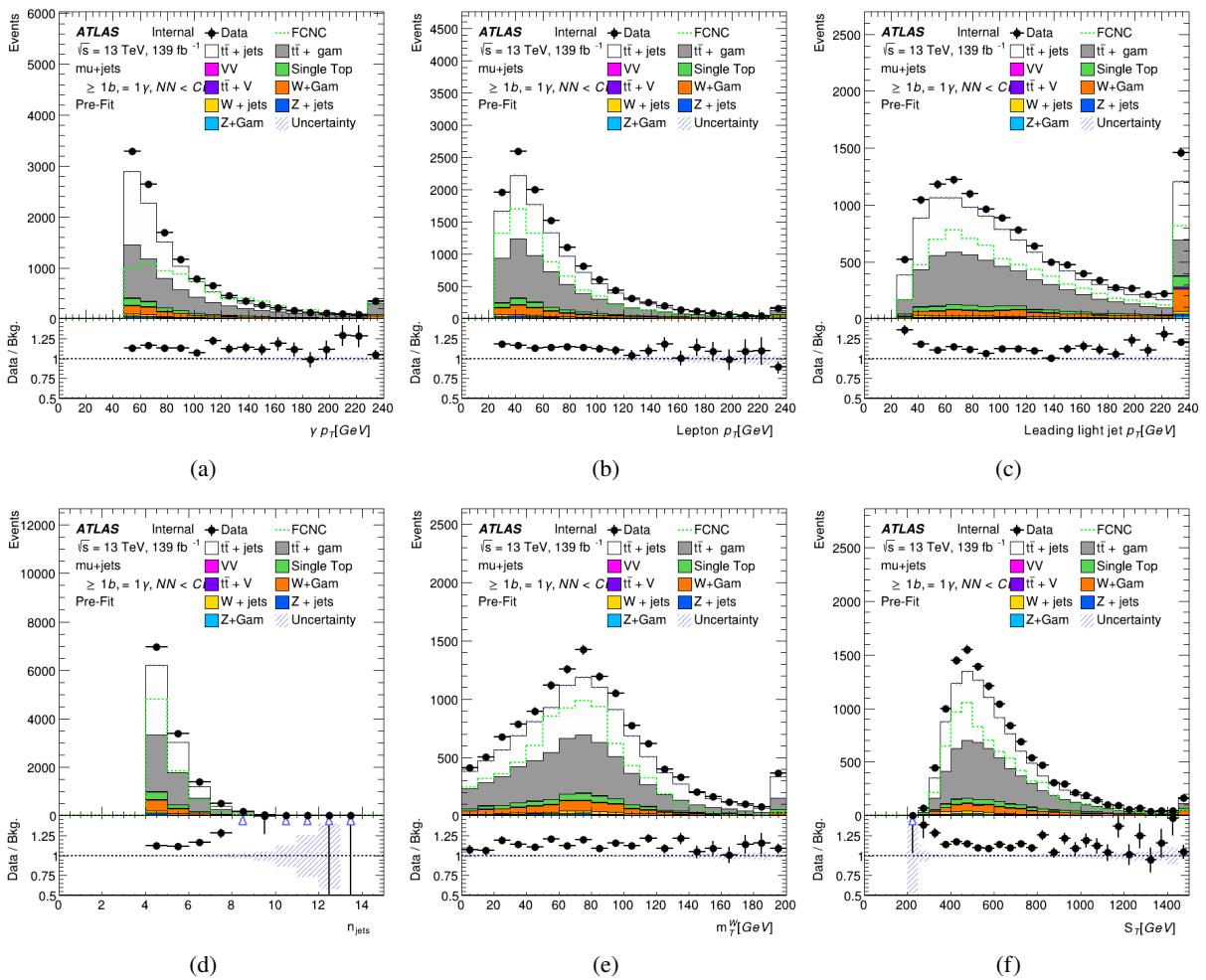


Figure 17: $t\bar{t}+\text{jets}+\gamma$ validation region plots for the muon channel. The FCNC signal sample is scaled to 1%.

⁴¹⁸ **5 Background Estimations**

419 6 Neural Network for Signal Background Separation

420 6.1 Event Classification: Neural Network Optimization

421 To help distinguish signal events from the majority of background events neural networks were employed
 422 for event classification. Neural networks are multivariate methods that take a variety of inputs and output a
 423 number between 0 and 1. The output value is a discriminating variable that will be used to classify events
 424 and determine which events make it into the final Signal Region selection. Signal-like events accumulate
 425 towards 1 while background-like events cluster around 0. Two neural networks are trained, one for the
 426 electron+jets final state and one for the muon+jets final state. This section will discuss the neural network
 427 studies completed and their uses in the search for FCNC events.

428 6.1.1 Input Variables

429 A wide variety of input variables to the neural network were studied in detail. Studies were done using only
 430 low level variables such as the kinematic variables (p_T , η , ϕ , E) of the physics objects in the signal region.
 431 This was done as a complex enough neural network should be able to figure out useful high level/event
 432 level variables (i.e. invariant masses, geometric separations) but in practice a combination of some of
 433 these low level variables and high level variables used as inputs to the neural network proved to give the
 434 best separation and projected limits. Using physical intuition to guide the neural network proved to be a
 435 valuable tool.

436 Combinations of 29 input variables were tested to start with however variables such as η and ϕ tend to
 437 not have significant weights in the neural network and are left out in favor of the the high level variables
 438 that include them (e.g., ΔR values). A measure of how different the variables are between signal and
 439 background is the Separation. Table 7 shows the separation values for the variables that are inputs to the
 440 final neural network. Comparisons between the shapes of the input variables for the μ +jets channel are
 441 shown in Figures 18, 19, and 20

$$\text{Separation} = \sum_i^{\text{bins}} \frac{n_{si} - n_{bi}}{n_{si} + n_{bi}}$$

442 Typically the kinematic variables with photon information have the biggest separation values. This is
 443 expected because the signal photon comes directly from the decay of a top quark and is much more energetic
 444 than background photons. Shape comparison plots for the e +jets channel and additional plots for other
 445 investigated variables are shown in Appendix ???. The largest difference in separation between the e +jets
 446 and μ +jets channels is the photon isolation value. This is due to the fact that all backgrounds are included
 447 and fake photon contamination from a large Z+jets background are expected. Both networks preform
 448 similarly in their separation of signal and background events. The network is able to learn and compensate
 449 for this behavior with the help of other variables that include the lepton and photon: $\Delta R_{l\gamma}$ and $m_{l\gamma}$.

450 The neural networks are trained on MC events that have a chance of being in the signal region after basic
 451 event level cuts and optimized for signal significance. Only events with 1 photon (> 15 GeV) and 1 bjet
 452 (MV2c10 77% working point) are classified by the neural network. The 77% working point was chosen by
 453 training the neural network on events with only 1 bjet at each working point: 70%, 77%, and 85% and

Variable	Separation e+jets	Separation μ +jets
$p_T(\gamma)$	22.97	24.01
$m_{q\gamma}$	22.65	28.31
γ_{iso}	18.62	41.32
m_{bW}	11.10	11.70
$m_{l\gamma}$	9.00	7.51
$\Delta R_{j\gamma}$	4.59	5.66
ΔR_{bl}	4.99	4.47
m_T^W	3.16	3.37
S_T	3.78	3.32
n_{jets}	1.70	2.03
χ_W^2	1.37	1.91
$p_T(q)$	2.46	2.82
$\Delta R_{l\gamma}$	1.40	1.19
E (lepton)	0.86	0.89
\not{E}_T	0.47	0.70
$p_T(b)$	0.51	0.53

Table 7: Separation of normalized variables between signal and background in the e+jets and μ +jets channels for the variables used as input to the final neural network.

454 picking the network and working point with the best estimated significance. The b-tagging neural network
 455 study is shown in Section 6.1.5

456 6.1.2 Architecture

457 A variety of architectures of dense neural networks are studied using KERAS[30] on top of the TENSORFLOW
 458 backend [31]. Each network has a number of input nodes equal to the number of input variables. Networks
 459 with one, two, and three hidden layers are investigated each with 20 nodes. The output layer contains
 460 only a single node. Every node in one layer is connected to every node in the next layer and the previous
 461 layer. Every connection is assigned a weight that is optimized during the training of the network. For
 462 every node in the network a value is computed using the weights and input values of the previous nodes
 463 using an activation function. Nodes with the highest output of this function are more important to the fit.
 464 The activation function used on the internal nodes in this search is the Rectified Linear Unit activation
 465 function.

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

466 The output layer uses the sigmoid function, $\sigma(x)$, as an activation function. The sigmoid function maps the
 467 output smoothly to the range (0,1).

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

468 Every training step the weights of each node are updated following an optimization algorithm, in this
 469 case the ADAM optimizer[32]. This optimizer follows the steepest gradient to reach the minimum of the
 470 parameter of interest called the loss function. The loss function used for these classification neural networks
 471 is the binary cross entropy:

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i))$$

472 where y is a binary indicator (0 or 1) if class label is the correct classification for observation and p is the
 473 predicted probability observation is the class label (0 or 1). The logarithmic nature of this loss function
 474 means it applies small values to correctly assigned event but more harshly punishes mismatching of events.
 475 Therefore having a similar number of signal and background events that get weighted similarly can improve
 476 the behavior of the network. In rare decay searches typically the amount of signal events is significantly
 477 smaller than the amount of background events in the training sample. Using the weight functionality in
 478 keras the total number of signal events can be scaled to be similar to the number of background events.

479 Weighting the signal events this way allows the network to separate the signal and background events in
 480 a way that is significantly less harsh than without the weights by taking advantage of the loss function
 481 being used. This improves the estimated significance of the neural network cut after the signal events are
 482 rescaled to their proper normalization values.

483 Various hyperparameters are used as inputs into the neural network as well as the optimizer used. The
 484 ADAM optimizer has a default learning rate of 0.001 which was not changed throughout these studies. The
 485 learning rate corresponds to the amount that weights are updated during training. A learning rate that is
 486 too large can mean the network never settles into a local minima as it is always missing the minima or
 487 at the very least it can take much longer to converge into a minima. As the neural network training for
 488 this search always converged quickly and to a similar value after being tested multiple different times the
 489 learning rate was not adapted.

490 Another hyperparameter of note is the batch size which defines the number of samples that are propagated
 491 through the network at once. The batch size is of crucial importance in how long the training of the network
 492 takes. A set of 1000 training samples with a batch size of 100 will propagate each set of 100 samples
 493 through the neural network every epoch, so 10 separate batches. A larger batch size means that each epoch
 494 of the training takes a shorter amount of time. However, as the weights are updated after each batch the
 495 network can take many more epochs to converge as the weights are being updated less frequently. A batch
 496 size of 100 was used while training the networks presented in this chapter. Larger batch sizes were tested
 497 with the only difference being the time each epoch took and the total time the network took to converge.

498 Epochs are the total number of times the network has been trained over the entire training set. All of the
 499 networks were allowed up to 200 epochs to converge with a KERAS patience value set to 50. The loss
 500 function minimization would be done every batch and after each epoch the best possible value of the loss
 501 function is found. If this value is better than any previous epoch the network is allowed to train for 50
 502 more epochs until 50 epochs have passed without finding a new minimum loss function value which then
 503 terminates the training. All models converge early and are terminated typically between epoch 80 and 120
 504 meaning the loss function was minimized between epoch 30 and 70.

505 One method employed to avoid overtraining the network dropout regularization was used on each of the
 506 hidden layers. Dropout has the effect of simulating a large number of networks with very different network
 507 structures by removing nodes randomly throughout the training. A dropout rate of 20% was used meaning

508 that for every batch 20% of the weights of the hidden layer nodes were set to 0. This forces the network to
 509 not become overly dependent on any given node and learning the data ‘by heart’ as opposed to recognizing
 510 the trends in the sample.

511 6.1.3 Training and Validation of Neural Networks

512 The input variables into the neural network are preprocessed using the ROBUSTSCALAR method implemented
 513 in **scikit-learn**[34]. The preprocessing is done so that the input variables exist on a similar scale. As
 514 the network is tasked with learning how to combine these inputs through a series of linear combinations
 515 and nonlinear activation function values a disparity in the scales of the input values can lead to awkward
 516 loss function topology that will focus on certain parameter gradients instead of treating them all similarly.
 517 Normalizing the values to a standard scale allows the network to learn the optimal parameters for each
 518 input node more quickly and efficiently. This means that less focus can be used on the optimization of
 519 the hyperparameters for the network as the scales of the inputs do not need to be learned by the network
 520 itself.

521 Each input variable in the neural network, x , is scaled by the following equation:

$$z = \frac{x - m}{q_3 - q_1}$$

522 where m is the median of the distribution, q_1 and q_3 are the first and third quartile. This changes the
 523 distribution of the input variable distributions to be centered around zero.

524 A second method to avoid overtraining the neural network is to make use of a train-test split to split the
 525 signal and background samples into 3 independent randomized sets before training the neural network.
 526 The samples are split into a training set of 64% of the samples, a test set containing 20% of the samples,
 527 and the remaining 16% are a validation set. The training and test sets are used during the training of the
 528 network while the validation set is used to compute performance of the trained neural network.

529 One measure of the performance of the network is the accuracy. The KERAS default accuracy measure is
 530 defined:

$$\text{accuracy} = \frac{N(\text{event}_{NN} \geq 0.5 | \text{signal}) + N(\text{event}_{NN} < 0.5 | \text{background})}{N(\text{signal}) + N(\text{background})}$$

531 where $N(\text{event}_{NN} \geq 0.5 | \text{signal})$ ($N(\text{event}_{NN} < 0.5 | \text{signal})$) is the number of signal (background) events
 532 with $P_{\text{signal}} \geq 0.5$ ($P_{\text{signal}} < 0.5$). Essentially, the accuracy is a measure of the mean of how often correct
 533 prediction values occur assuming a cut on the output of ≥ 0.5 .

534 6.1.4 Hidden Layer Studies

535 The general performance of the neural network was studied with a varying number of hidden layers (1, 2,
 536 and 3) in both the e +jets and μ +jets channels. All of the networks are trained on the same set of variables
 537 and with the same train-test split input data. For each of the channels the *Receiver Operating Characteristic*
 538 (ROC) curves are shown in Figure 22. The ROC curves show the value of $1 - \epsilon_{\text{bkg}}$ as a function of the true
 539 positive rate, ϵ_{signal} . A figure of merit is the Area Under the Curve (AUC) which is a measure of how close
 540 the resulting values are to the optimal value of unity.

541 The AUC for 2 hidden layers and 3 hidden layers are identical, to rounding errors, for both channels. As
 542 such the network with 2 hidden layers has been chosen as it is computationally simpler. The normalized
 543 neural network output values are shown in Figure 25. Adding a second hidden layer significantly improves
 544 the performance of the network but a third layer does not. The output shapes change slightly adding the
 545 third hidden layer due to the network learning differently about the same data. However, as the AUC shows,
 546 the performance of 2 and 3 hidden layers is identical. Figures 23 and 24 show the accuracy metric and
 547 the loss function as a function of the training epoch for the networks trained with 2 hidden layers. The
 548 accuracy plot behavior is expected as the validation data sets do not have dropout regularization applied to
 549 them. These networks are also trained without further reduction of Z+jets background meaning the e +jets
 550 sample has a larger background contamination that makes the validation testing more volatile. This is due
 551 to the increased number of similar events in that sample that can be more heavily dependent on specific
 552 weights across the network for identification.

553 The main metric used in choosing which network has the best physics reach is the significance:

$$\text{significance} = \frac{N_s}{\sqrt{N_s + N_b}}$$

554 where N_s is the number of signal events that pass the cut and N_b is the number of background events that
 555 pass the neural network cut. After the model has been fully trained it is tested on all of the Monte Carlo
 556 for signal and background. The signal samples are normalized to various branching ratios (in the range
 557 $10^{-5} \rightarrow 3 \times 10^{-3}$) and full LHC Run-2 Luminosity and the significance is calculated as a function of the cut
 558 on the output of the neural network $P(\text{signal})$. The network with the output cut for the smallest branching
 559 ratio with a maximum significance of 2 is chosen, a rough estimate of where the expected limit could be
 560 set. The significance as a function of the neural network output cut is shown in Figure 26.

561 6.1.5 B-Tagging Working Point Studies

562 The b-tagging working point selection was also done with similar neural network studies. Three neural
 563 networks were trained with the datasets using the jet information and total scaled events for each of the
 564 major b-tagging working points: 70%, 77%, and 85%. Changing the working point changes a number of
 565 things about the signal and background data sets such as which jets are b tagged and therefore which jets
 566 are combined into the higher level variables (e.g., $m_{q\gamma}$ and m_{Wb}). The total number of events that pass
 567 the preselection to the neural network are also changed for all of the datasets since the neural network are
 568 only trained on events with 1 b-tagged jet. Similar sets of plots to Section 6.1.4 will be presented in this
 569 section.

570 This selection of neural networks were trained in parallel with one, two, and three hidden layers. The only
 571 results shown are the 2 hidden layer outputs as they perform equally or better to the others as previously
 572 discussed. The accuracy and loss plots for these networks are shown in Figures 27 and 28. Following that
 573 the neural network output and significance plots are shown in Figures 29 and 30.

574 The result of these studies is the choice of using the 77% working point for b-tagged jets. The branching
 575 ratio with significance of 2 is found for each network and reported in Table 8.

B-Tag Working Point	e +jets Branching Ratio	μ +jets Branching Ratio
70%	1.25×10^{-5}	1.31×10^{-5}
77%	1.23×10^{-5}	1.18×10^{-5}
85%	1.27×10^{-5}	1.19×10^{-5}

Table 8: Branching ratio values with a significance of 2 after neural network optimization

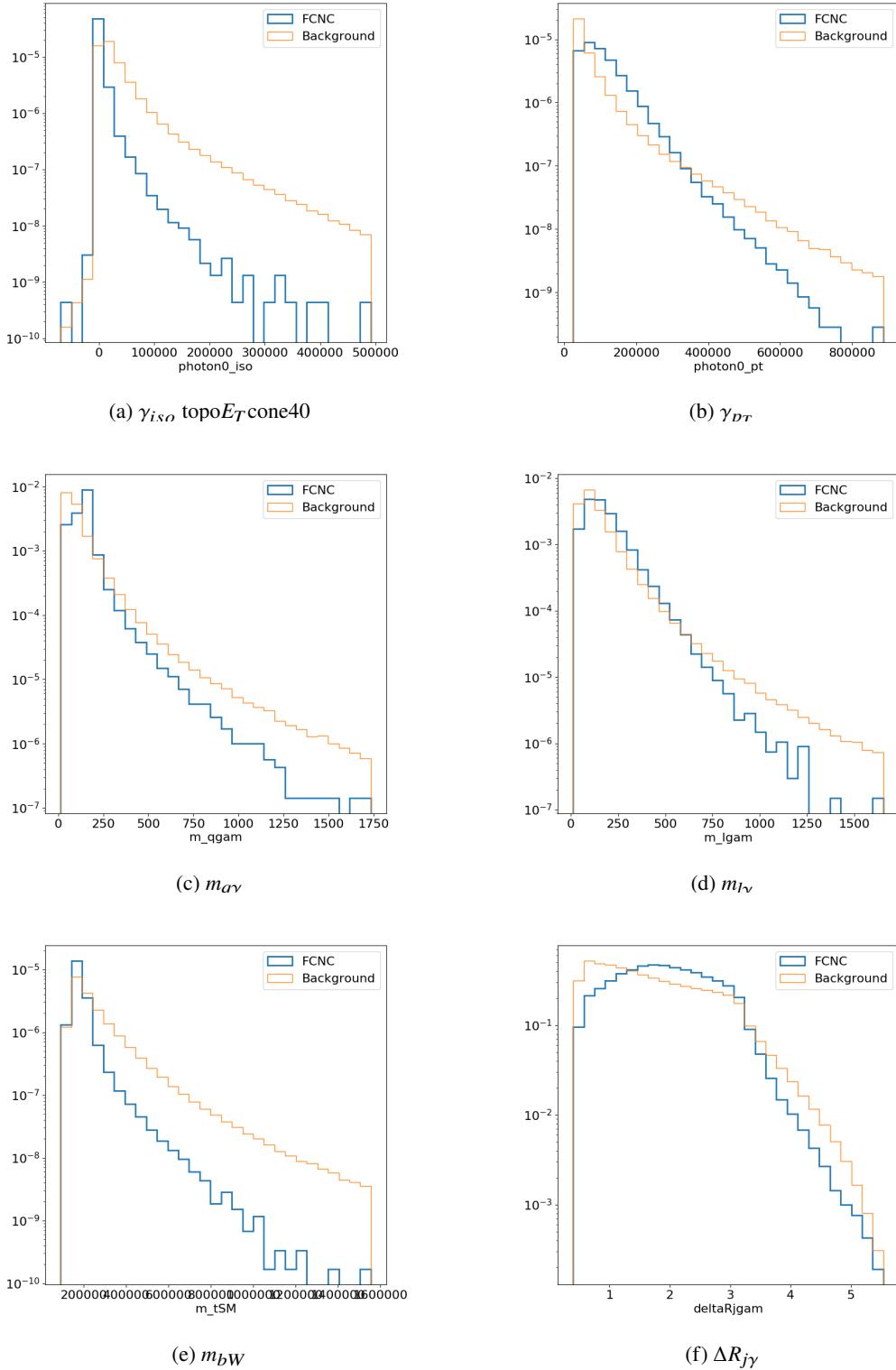


Figure 18: Normalized variables showing the shapes of neural network input variables for the μ +jets channel: γ_{iso} topo E_T cone40, γ_{pT} , $m_{q\gamma}$, $m_{l\gamma}$, m_{bW} , and $\Delta R_{j\gamma}$

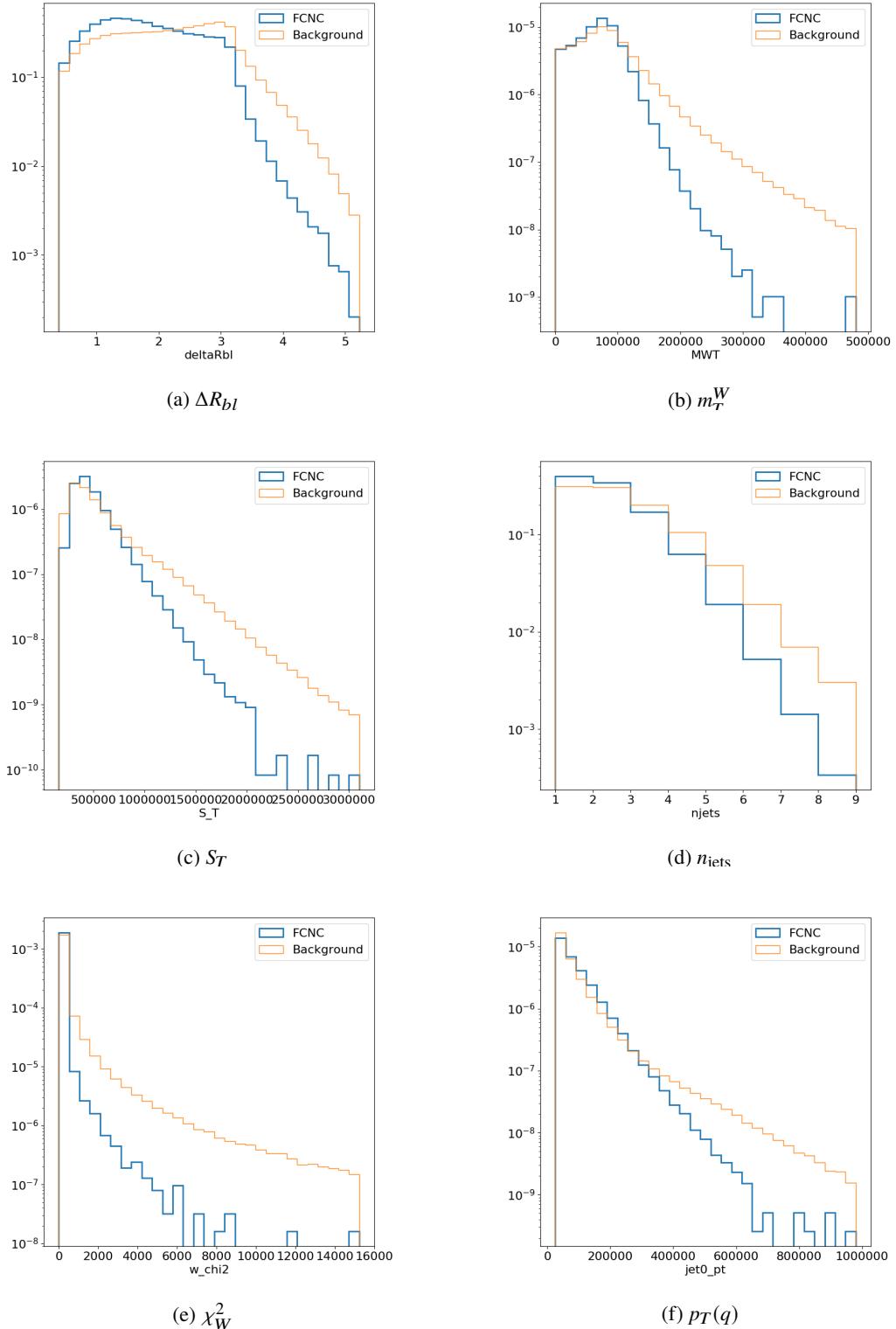


Figure 19: Normalized variables showing the shapes of neural network input variables for the $\mu+\text{jets}$ channel: ΔR_{bl} , m_T^W , S_T , n_{jets} , χ_W^2 , and $p_T(q)$

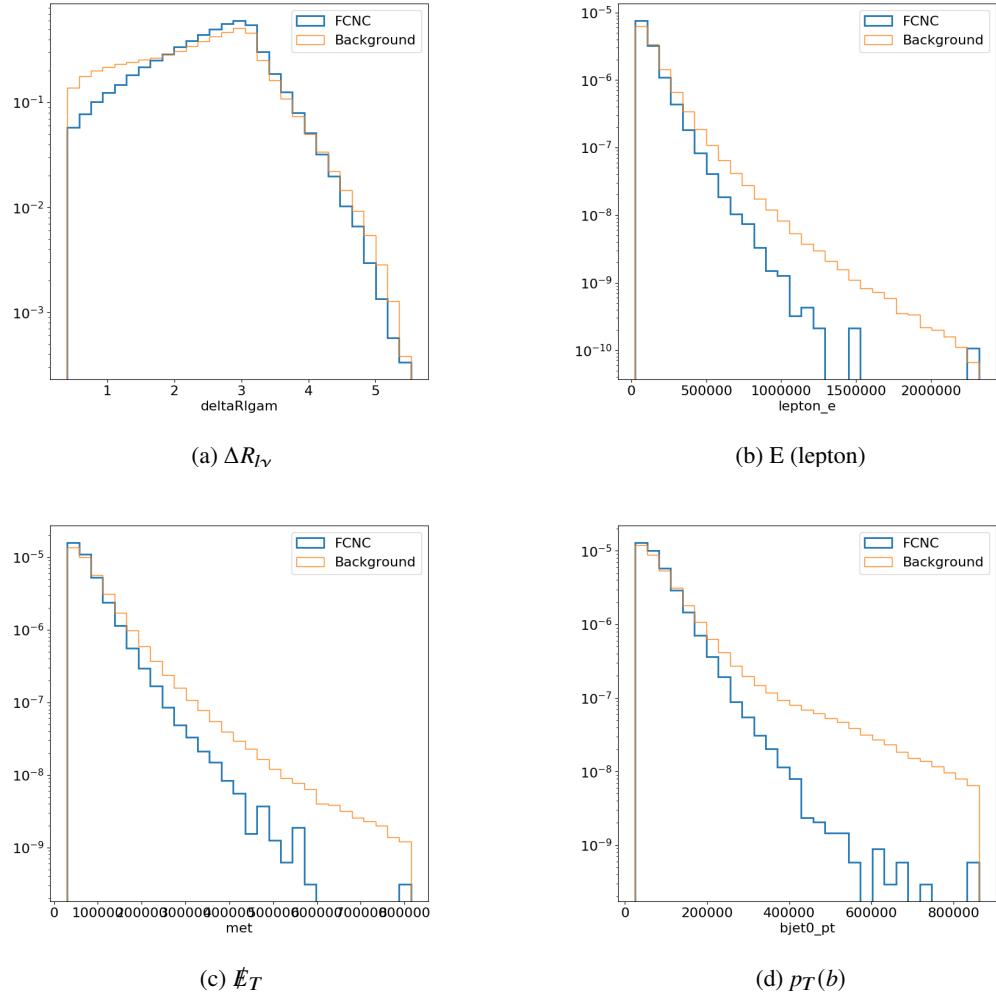


Figure 20: Normalized variables showing the shapes of neural network input variables for the $\mu + \text{jets}$ channel: $\Delta R_{l\gamma}$, E (lepton), \cancel{E}_T , and $p_T(b)$

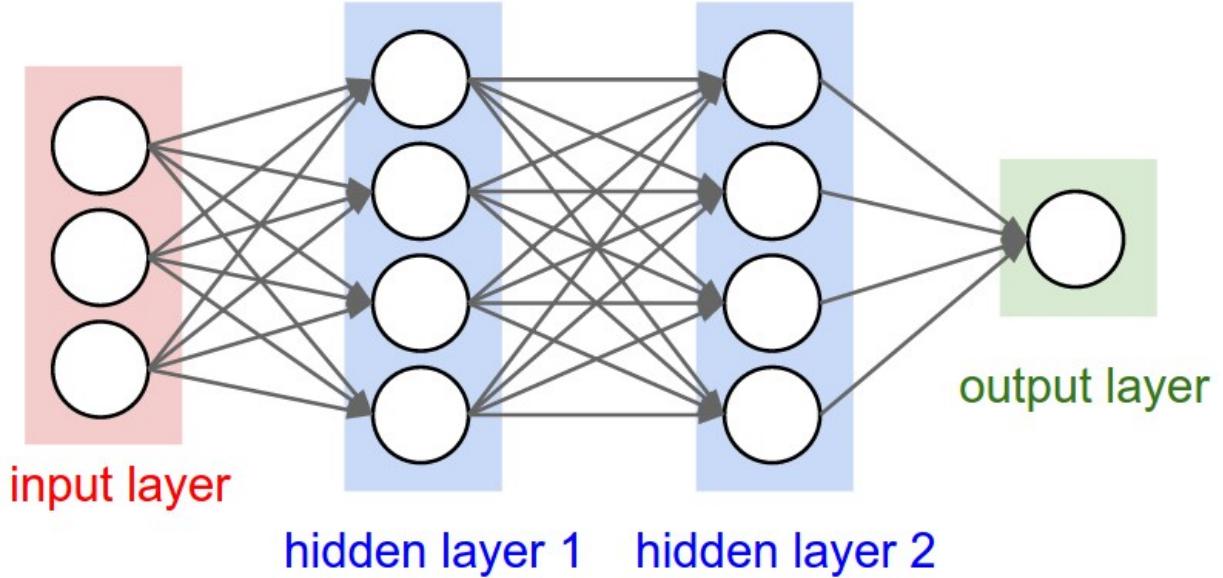


Figure 21: Pictoral representation of neural network architecture with 3 input variables, 2 hidden layers with 4 nodes each, and 1 output layer[33].

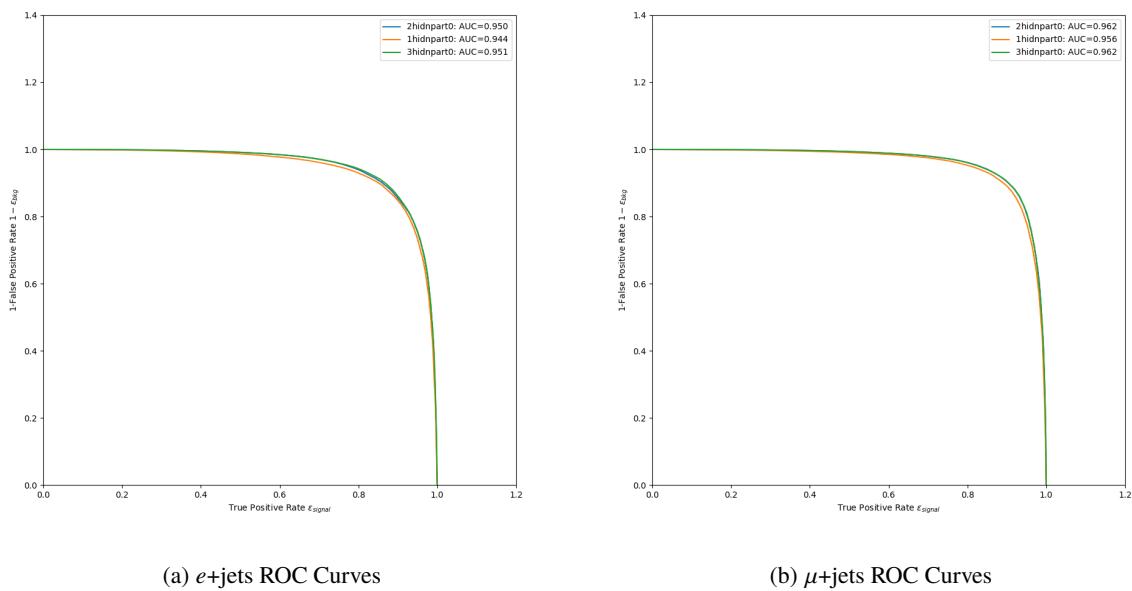


Figure 22: ROC Curves are shown for both search channels for a varying number of hidden layers. Orange lines correspond to one hidden layer, blue to 2 hidden layers and green to 3 hidden layers. The blue and green curves have near identical AUC values: 0.950 and 0.951 for the $e+jets$ case and 0.962 for the $\mu+jets$ cases.

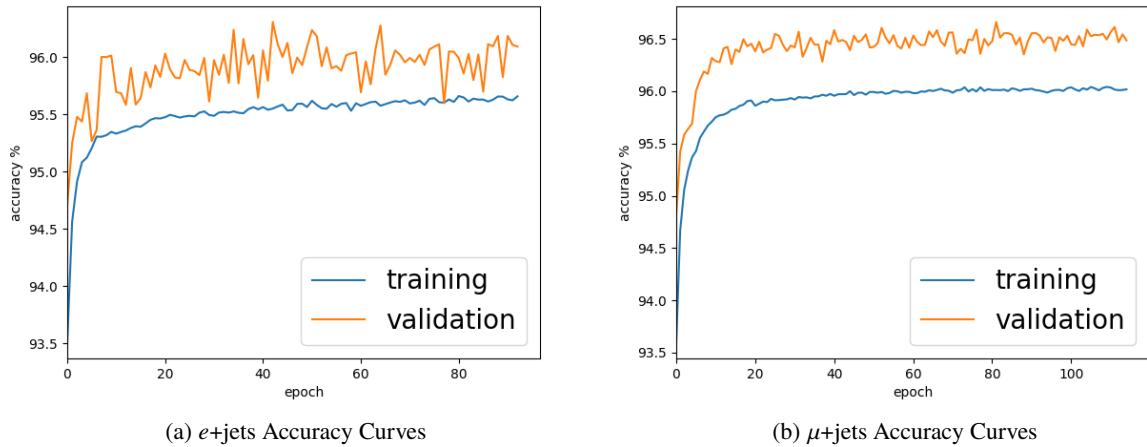


Figure 23: Accuracy plots for both channels for the 2 hidden layer neural network

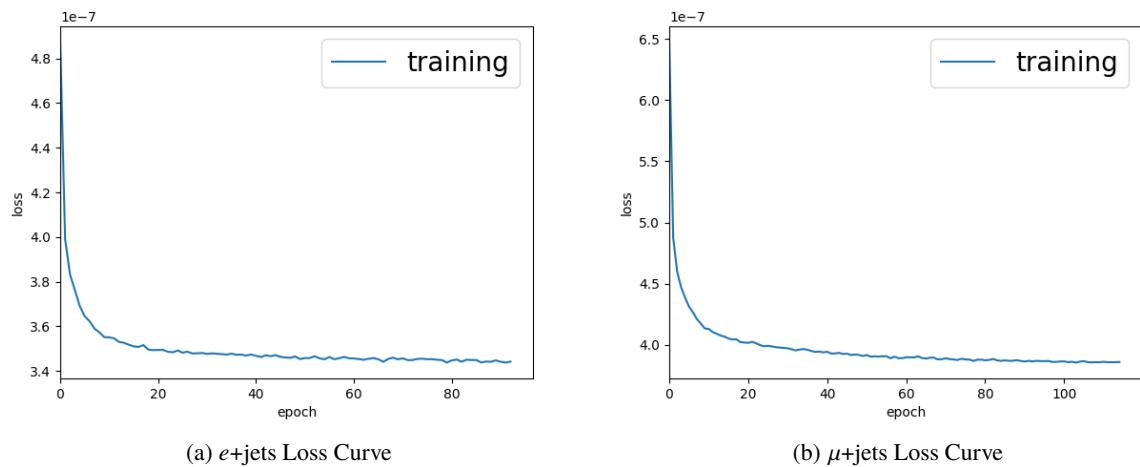


Figure 24: Loss plots for both channels for the 2 hidden layer neural network

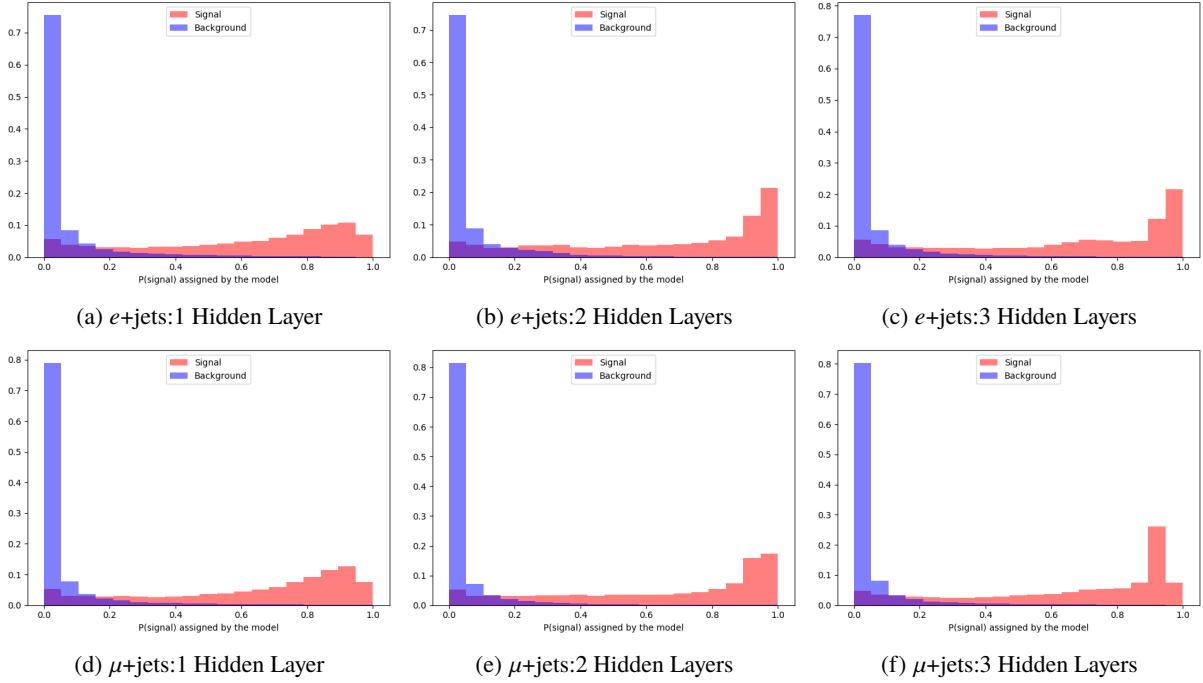


Figure 25: Normalized neural network output signal and background distribution plots are shown for both search channels for a varying number of hidden layers.

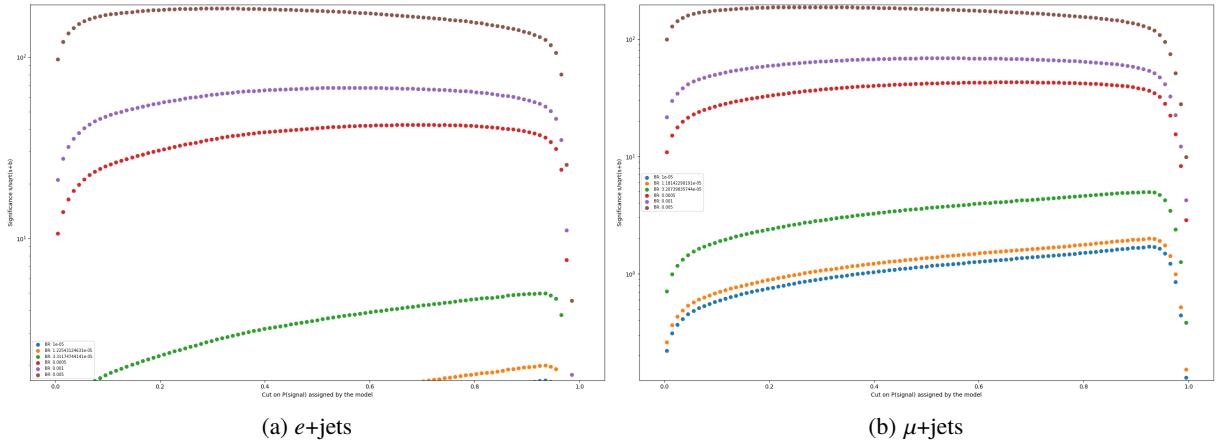


Figure 26: Significance plots for both channels for the 2 hidden layer neural network. The green points correspond to a branching ratio with a maximum significance of 5, the orange to a maximum significance of 2. The $e+\text{jets}$ ($\mu+\text{jets}$) branching ratio with max significance of 2 is 1.22×10^{-5} (1.18×10^{-5}). The blue, red, purple, and brown points correspond to branching ratios of 1×10^{-5} , 5×10^{-4} , 1×10^{-3} , and 5×10^{-3} , respectively.

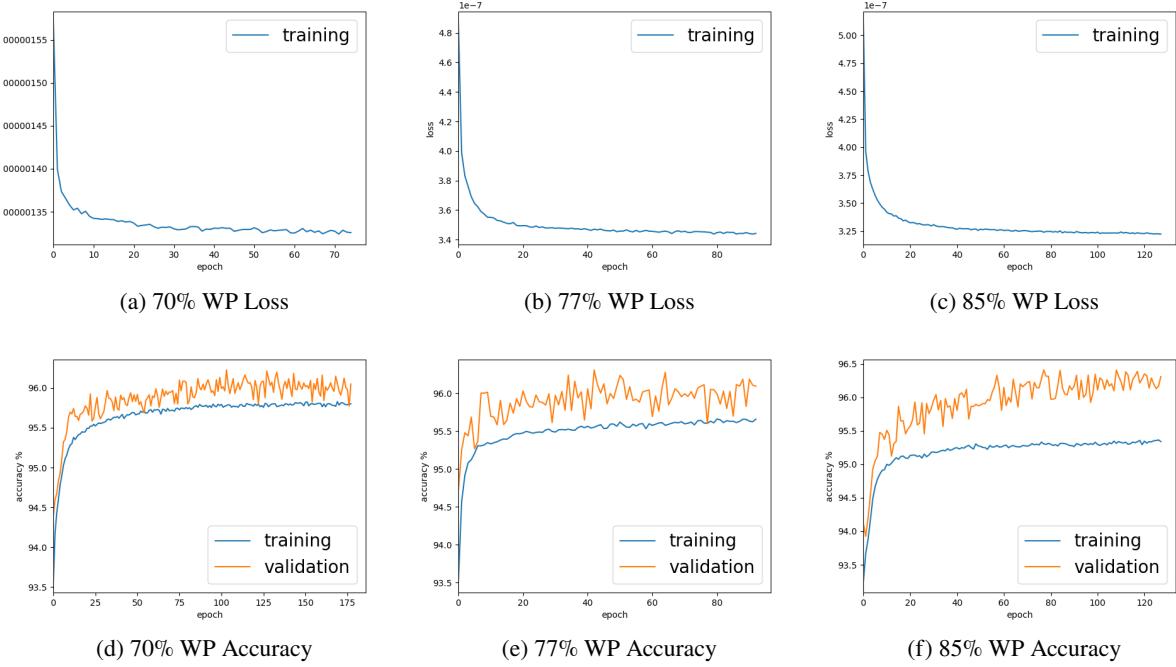


Figure 27: Accuracy and loss plots for the $e + \text{jets}$ channel at 70%, 77%, and 85% b-tagging working points.

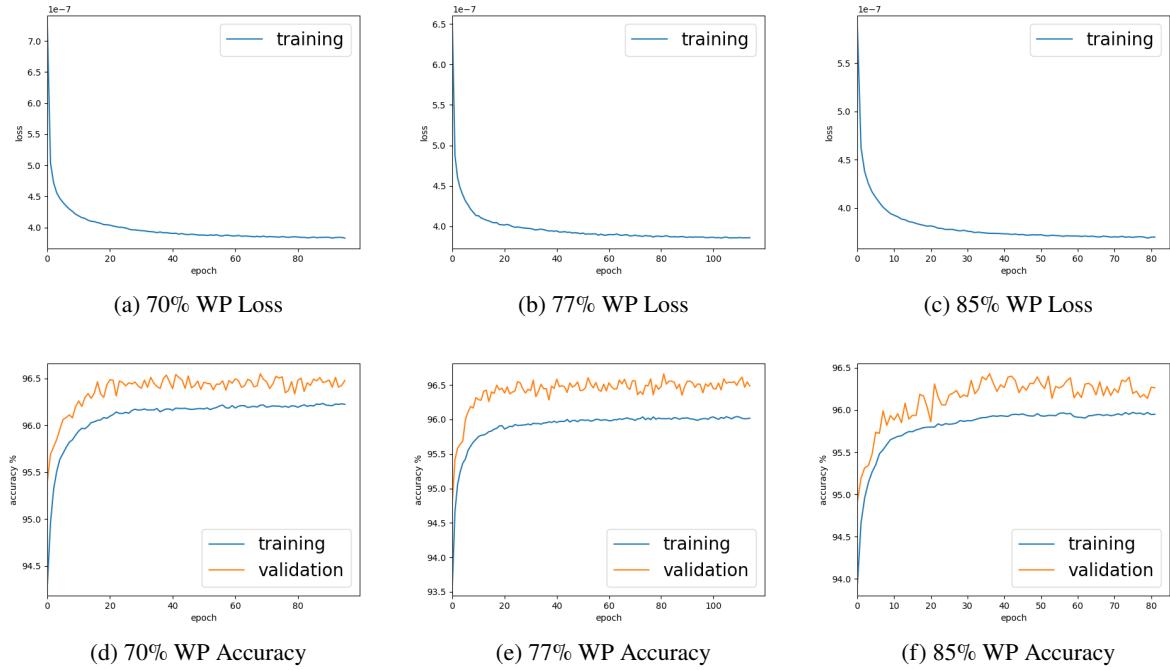


Figure 28: Accuracy and loss plots for the $\mu + \text{jets}$ channel at 70%, 77%, and 85% b-tagging working points.

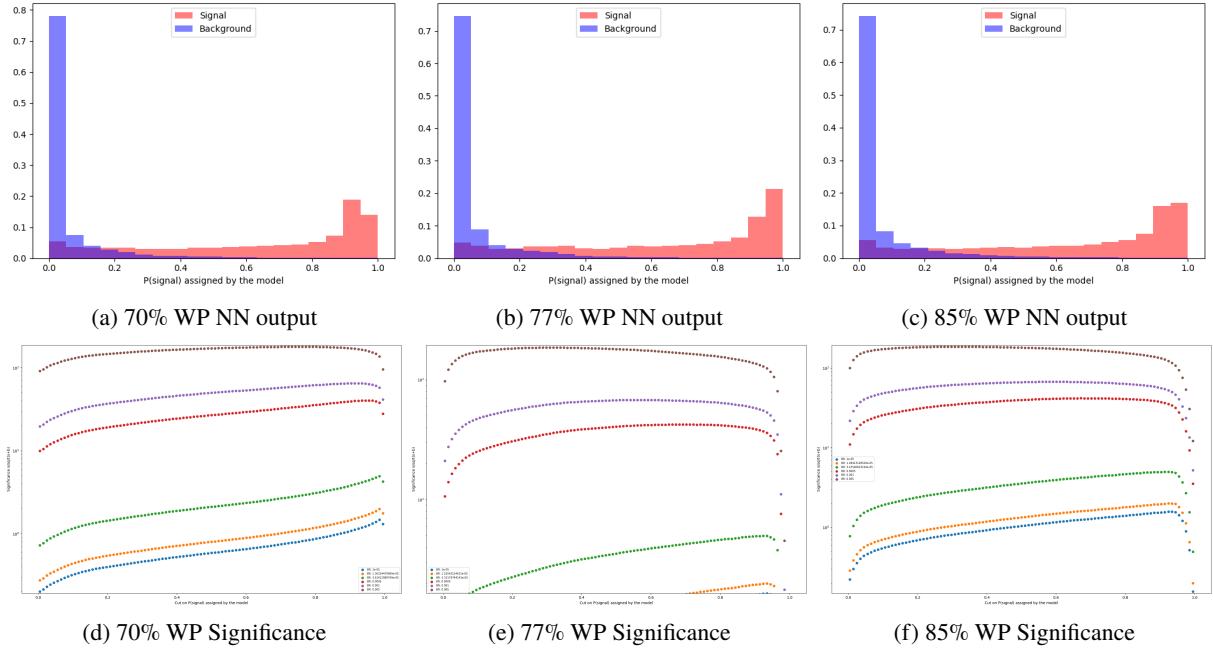


Figure 29: Neural network output and significance plots for the $e + \text{jets}$ channel at 70%, 77%, and 85% b-tagging working points.

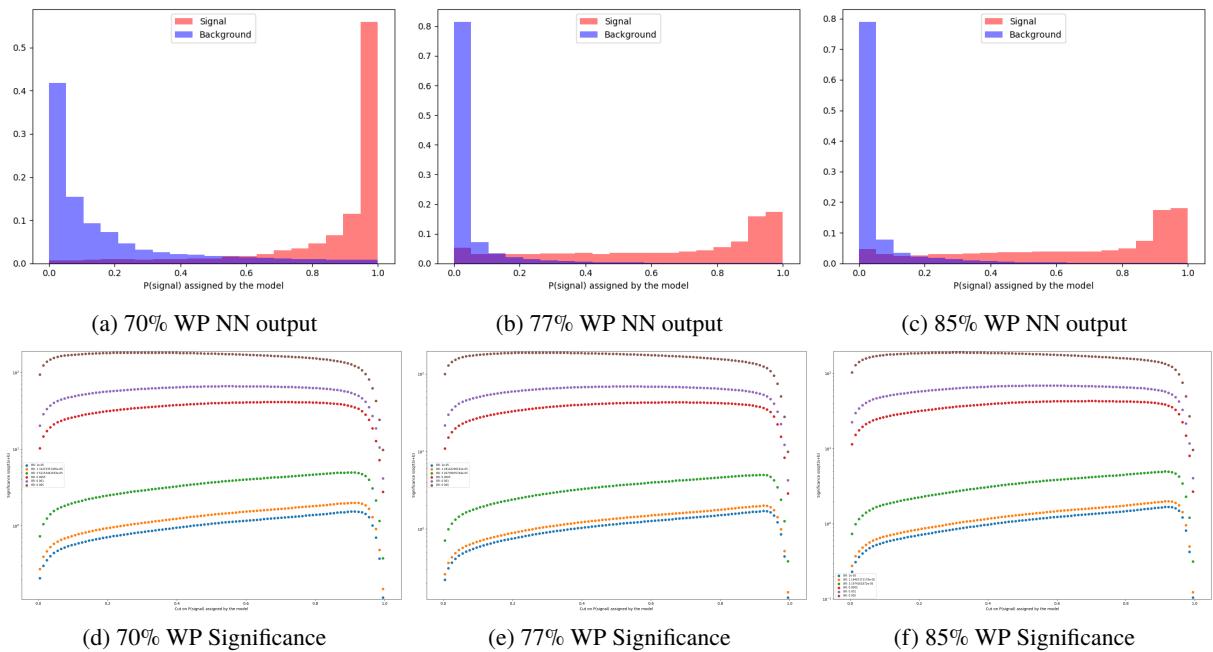


Figure 30: Neural network output and significance plots for the $\mu + \text{jets}$ channel at 70%, 77%, and 85% b-tagging working points.

⁵⁷⁶ **7 Analysis**

⁵⁷⁷ Place your results here.

578 8 Conclusion

579 Place your conclusion here.

580 References

- 581 [1] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961) 579–588.
- 582 [2] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508–509, [,160(1964)].
- 583 [3] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321–323, [,157(1964)].
- 584 [4] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264–1266.
- 585 [5] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C680519* (1968) 367–377.
- 586 [6] ATLAS Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B716** (2012) 1–29, [arXiv:1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214).
- 587 [7] CMS Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B716** (2012) 30–61, [arXiv:1207.7235 \[hep-ex\]](https://arxiv.org/abs/1207.7235).
- 588 [8] CDF Collaboration, F. Abe et al., *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) 2626–2631, [arXiv:hep-ex/9503002 \[hep-ex\]](https://arxiv.org/abs/hep-ex/9503002).
- 589 [9] ATLAS Collaboration, T. A. collaboration, *Measurement of the top quark mass in the $t\bar{t} \rightarrow$ lepton+jets channel from $\sqrt{s}=8$ TeV ATLAS data.*,
- 590 [10] Particle Data Group Collaboration, M. Tanabashi et al., *Review of Particle Physics*, *Phys. Rev.* **D98** (2018) 030001.
- 591 [11] J. J. Cao, G. Eilam, M. Frank, K. Hikasa, G. L. Liu, I. Turan, and J. M. Yang, *SUSY-induced FCNC top-quark processes at the large hadron collider*, *Phys. Rev.* **D75** (2007) 075021, [arXiv:hep-ph/0702264 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0702264).
- 592 [12] J. Cao, Z. Heng, L. Wu, and J. M. Yang, *R-parity violating effects in top quark FCNC productions at LHC*, *Phys. Rev.* **D79** (2009) 054003, [arXiv:0812.1698 \[hep-ph\]](https://arxiv.org/abs/0812.1698).
- 593 [13] G. Eilam, J. L. Hewett, and A. Soni, *Rare decays of the top quark in the standard and two Higgs doublet models*, *Phys. Rev.* **D44** (1991) 1473–1484, [Erratum: *Phys. Rev.* D59,039901(1999)].
- 594 [14] J. Pequenao, *Event Cross Section in a computer generated image of the ATLAS detector.*, Mar, 2008.
- 595 [15] ATLAS Collaboration, M. Aaboud et al., *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J.* **C79** (2019) 639, [arXiv:1902.04655 \[physics.ins-det\]](https://arxiv.org/abs/1902.04655).
- 596 [16] *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, *Eur. Phys. J.* **C74** (2014) 3071, [arXiv:1407.5063](https://arxiv.org/abs/1407.5063).

- [17] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV*, Aug, 2016, <https://cds.cern.ch/record/2203514>.
- [18] ATLAS Collaboration, S. Rettie, *Muon identification and performance in the ATLAS experiment*, PoS **DIS2018** (2018) 097.
- [19] ATLAS Collaboration, M. Aaboud et al., *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, Eur. Phys. J. **C79** (2019) 205, [arXiv:1810.05087 \[hep-ex\]](arXiv:1810.05087).
- [20] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_t jet clustering algorithm*, JHEP **04** (2008) 063, [arXiv:0802.1189 \[hep-ph\]](arXiv:0802.1189).
- [21] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, pp. , 270–279. 1998. [arXiv:hep-ph/9907280 \[hep-ph\]](arXiv:hep-ph/9907280).
https://inspirehep.net/record/484872/files/arXiv:hep-ph_9907280.pdf.
- [22] *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2015-029, CERN, Geneva, Jul, 2015. <http://cds.cern.ch/record/2037702>.
- [23] ATLAS Collaboration, T. A. collaboration, *Tagging and suppression of pileup jets*,.
- [24] S. Schramm, *HowToCleanJets2016 : Jet-Based Event Cleaning*, 2016.
<https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HowToCleanJets2016>.
- [25] CMS Collaboration, A. M. Sirunyan et al., *Measurement of b hadron lifetimes in pp collisions at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. **C78** (2018) 457, [arXiv:1710.08949 \[hep-ex\]](arXiv:1710.08949), [Erratum: Eur. Phys. J.C78,no.7,561(2018)].
- [26] N. Bartosik, *File:B-tagging diagram.png*, 2016.
https://commons.wikimedia.org/wiki/File:B-tagging_diagram.png.
- [27] ATLAS Collaboration, F. A. Di Bello, *Optimisation of the ATLAS b-tagging algorithms for the 2017-2018 LHC data-taking*, PoS **EPS-HEP2017** (2017) 733.
- [28] ATLAS Collaboration, M. Aaboud et al., *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, JHEP **08** (2018) 089, [arXiv:1805.01845 \[hep-ex\]](arXiv:1805.01845).
- [29] ATLAS Collaboration, M. Aaboud et al., *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. **C78** (2018) 903, [arXiv:1802.08168 \[hep-ex\]](arXiv:1802.08168).
- [30] F. C. et al., *Keras: The Python Deep Learning Library*, 2019. <https://keras.io>.
- [31] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*, [arXiv:1603.04467 \[cs.DC\]](arXiv:1603.04467).
- [32] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, [arXiv:1412.6980 \[cs.LG\]](arXiv:1412.6980).
- [33] A. Karpathy, *CS231n Convolutional Neural Networks for Visual Recognition*, 2015.
<http://cs231n.github.io/neural-networks-1/>.
- [34] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, J. Machine Learning Res. **12** (2011) 2825–2830, [arXiv:1201.0490 \[cs.LG\]](arXiv:1201.0490).

650 **List of contributions**

651 Jason Barkeloo

652 Jim Brau

All Aspects of Analysis and Note

Supervisor to Jason Barkeloo

653 Appendices

654 A Derivation Information (TOPQ1)

655 Preselection is applied to both data and MC samples using the derivation framework in order to reduce the
 656 xAOD sample size. TOPQ1 derivations were used and further skimmed for the specific n-tuples used in
 657 this analysis. TOPQ1 derivations are used for standard top analyses with at least one lepton and have the
 following definitions:

Particle	Definition
Electron	$ \eta < 2.5$ and Electrons.DFCommonElectronsLHLoose
Muon	$ \eta < 2.5$ and Muons.muonType=0 and Muons.DFCommonGoodMuon
Small-R Jet	$ \eta < 2.5$ and AntiKt4EMTopoJets.DFCommonJets_Calib
Large-R Jet	$ \eta < 2.5$ and AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets.DFCommonJets_Calib

Table 9: TOPQ1 Object Definitions

658
 659 In addition to these object definitions TOPQ1 Derivations require trigger skimming of ≥ 1 lepton with
 660 $p_T > 20\text{GeV}$.

661 B Complete List of Monte Carlo Samples Used

662 The following MC samples have been used to simulate the signal and various Standard Model backgrounds
 663 at center of mass energy of 13 TeV. MC16a, MC16d, and MC16e samples correspond to run conditions for
 664 years 2015/16, 2017, and 2018 respectively. The newest unprescaled p-tag samples at the time of writing
 665 were used for each sample.

Process	MC DSID	Generator	Generator Tune
Diboson processes, Full Simulation			
$ZZ \rightarrow llll$	364250	SHERPA2.2.2	NNPDF3.0 NNLO
$ZW \rightarrow llvv$	364253	SHERPA2.2.2	NNPDF3.0 NNLO
$WW/ZZ \rightarrow llvv$	364254	SHERPA2.2.2	NNPDF3.0 NNLO
$WZ \rightarrow lvvv$	364255	SHERPA2.2.2	NNPDF3.0 NNLO
$ZZ \rightarrow qqvv$	364355	SHERPA2.2.1	NNPDF3.0 NNLO
$ZZ \rightarrow qql\bar{l}$	364356	SHERPA2.2.1	NNPDF3.0 NNLO
$WZ \rightarrow qqvv$	363357	SHERPA2.2.1	NNPDF3.0 NNLO
$WZ \rightarrow qql\bar{l}$	363358	SHERPA2.2.1	NNPDF3.0 NNLO
$WW \rightarrow qql^-\nu$	363359	SHERPA2.2.1	NNPDF3.0 NNLO
$WW \rightarrow qql^+\nu$	363360	SHERPA2.2.1	NNPDF3.0 NNLO
$WZ \rightarrow qqlv$	363489	SHERPA2.2.1	NNPDF3.0 NNLO
$t\bar{t} + V$ processes			
$t\bar{t}W$	410155	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO
$t\bar{t}Z \rightarrow t\bar{t}vv$	410156	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO
$t\bar{t}Z \rightarrow t\bar{t}qq$	410157	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO
$t\bar{t}Z \rightarrow t\bar{t}ee$	410218	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO
$t\bar{t}Z \rightarrow t\bar{t}\mu\mu$	410219	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO
$t\bar{t}Z \rightarrow t\bar{t}\tau\tau$	410220	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO

Table 10: List of MC16a (r-tag 9364) samples used in this search for the $t\bar{t} + V$ and diboson processes. All samples are available and used for MC16d (r-tag 10201) and MC16e (r-tag 10724) unless otherwise stated.

Process	MC DSID	Generator	Generator Tune
Signal samples for FCNC in decay mode: $pp \rightarrow t\bar{t} \rightarrow blvq\gamma$, NLO, AFII			
$t\gamma$, anti-top	410980	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
$t\gamma$, top	410981	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
$t\gamma$, anti-top	410984	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
$t\gamma$, top	410985	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
$t\bar{t}$ processes, Full Simulation			
$t\bar{t}$ (NLO)	410470	POWHEG-BOX + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
$t\bar{t} + \gamma$ (LO)	410389	POWHEG-BOX + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
$t\bar{t}$ processes, Variation Samples, NLO, AFII			
$t\bar{t}$ (single lepton)	410557	POWHEG-BOX + HERWIG 7 + EvtGEN	MMHT2017lo68cl
$t\bar{t}$ (di-lepton)	410558	POWHEG-BOX + HERWIG 7 + EvtGEN	MMHT2017lo68cl
$t\bar{t}$ (single lepton)	410464	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO
$t\bar{t}$ (di-lepton)	410465	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO
$t\bar{t}$ (single lepton, $h_{\text{damp}} = 3m_{\text{top}}$)	410480	POWHEG-BOX + PYTHIA 8 + EvtGEN	A14NNPDF23LO
$t\bar{t}$ (di-lepton, $h_{\text{damp}} = 3m_{\text{top}}$)	410482	POWHEG-BOX + PYTHIA 8 + EvtGEN	A14NNPDF23LO
Single-top processes, NLO, Full Simulation			
t-channel, top	410658	POWHEG-BOX + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
t-channel, anti-top	410659	POWHEG-BOX + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
s-channel, top	410644	POWHEG-BOX + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
s-channel, anti-top	410645	POWHEG-BOX + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
tW-channel, top	410646	POWHEG-BOX + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
tW-channel, anti-top	410647	POWHEG-BOX + PYTHIA 8 + EvtGEN	A14 NNPDF23LO
Single-top processes, Variation Samples, NLO			
t-channel, anti-top, AFII	411032	POWHEG-BOX + HERWIG 7 + EvtGEN	MMHT2017lo68cl
t-channel, top, AFII	411033	POWHEG-BOX + HERWIG 7 + EvtGEN	MMHT2017lo68cl
t-channel, AFII	410560	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO
s-channel, anti-top, AFII	411034	POWHEG-BOX + HERWIG 7 + EvtGEN	MMHT2017lo68cl
s-channel, top, AFII	411035	POWHEG-BOX + HERWIG 7 + EvtGEN	MMHT2017lo68cl
s-channel, AFII	410561	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO
tW-channel, anti-top, AFII	411036	POWHEG-BOX + HERWIG 7 + EvtGEN	MMHT2017lo68cl
tW-channel, top, AFII	411037	POWHEG-BOX + HERWIG 7 + EvtGEN	MMHT2017lo68cl
tW-channel, anti-top, AFII	410654	POWHEG-BOX + HERWIG 7 + EvtGEN	MMHT2017lo68cl
tW-channel, top, AFII	410655	POWHEG-BOX + HERWIG 7 + EvtGEN	MMHT2017lo68cl
tW-channel, AFII	412002	MADGRAPH5_AMC@NLO + PYTHIA 8 + EvtGEN	A14NNPDF23LO

Table 11: List of MC16a (r-tag 9364) samples used in this search for the signal sample, $t\bar{t}$, and single-top processes. All samples are available and used for MC16d (r-tag 10201) and MC16e (r-tag 10724) unless otherwise stated.

Process	MC DSID	Generator	Generator Tune
W+Jets, Full Simulation			
$W \rightarrow ev, \max(H_T, p_T^V) < 70\text{GeV}, c \text{ veto, b veto}$	364170	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, \max(H_T, p_T^V) < 70\text{GeV}, c \text{ filter, b veto}$	364171	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, \max(H_T, p_T^V) < 70\text{GeV}, b \text{ filter}$	364172	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, c \text{ veto, b veto}$	364173	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, c \text{ filter, b veto}$	364174	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, b \text{ filter}$	364175	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, c \text{ veto, b veto}$	364176	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, c \text{ filter, b veto}$	364177	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, b \text{ filter}$	364178	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, c \text{ veto, b veto}$	364179	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, c \text{ filter, b veto}$	364180	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, b \text{ filter}$	364181	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, 500\text{GeV} < \max(H_T, p_T^V) < 1000\text{GeV}$	364182	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow ev, \max(H_T, p_T^V) > 1000\text{GeV}$	364183	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, \max(H_T, p_T^V) < 70\text{GeV}, c \text{ veto, b veto}$	364156	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, \max(H_T, p_T^V) < 70\text{GeV}, c \text{ filter, b veto}$	364157	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, \max(H_T, p_T^V) < 70\text{GeV}, b \text{ filter}$	364158	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, c \text{ veto, b veto}$	364159	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, c \text{ filter, b veto}$	364160	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, b \text{ filter}$	364161	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, c \text{ veto, b veto}$	364162	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, c \text{ filter, b veto}$	364163	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, b \text{ filter}$	364164	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, c \text{ veto, b veto}$	364165	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, c \text{ filter, b veto}$	364166	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, b \text{ filter}$	364167	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 500\text{GeV} < \max(H_T, p_T^V) < 1000\text{GeV}$	364168	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, \max(H_T, p_T^V) > 1000\text{GeV}$	364169	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, \max(H_T, p_T^V) < 70\text{GeV}, c \text{ veto, b veto}$	364184	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, \max(H_T, p_T^V) < 70\text{GeV}, c \text{ filter, b veto}$	364185	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, \max(H_T, p_T^V) < 70\text{GeV}, b \text{ filter}$	364186	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, c \text{ veto, b veto}$	364187	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, c \text{ filter, b veto}$	364188	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, b \text{ filter}$	364189	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, c \text{ veto, b veto}$	364190	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, c \text{ filter, b veto}$	364191	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, b \text{ filter}$	364192	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, c \text{ veto, b veto}$	364193	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, c \text{ filter, b veto}$	364194	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, b \text{ filter}$	364195	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 500\text{GeV} < \max(H_T, p_T^V) < 1000\text{GeV}$	364196	SHERPA2.2.1	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, \max(H_T, p_T^V) > 1000\text{GeV}$	364197	SHERPA2.2.1	NNPDF3.0 NNLO

Table 12: List of MC16a (r-tag 9364) samples used in this search for the W+jets background. All samples are available and used for MC16d (r-tag 10201) and MC16e (r-tag 10724). All processes are simulated for up to 2 partons at NLO and 4 partons at LO.

Process	MC DSID	Generator	Generator Tune
Z+Jets, Full Simulation			
$Z \rightarrow ee, \max(H_T, p_T^V) < 70\text{GeV}, \text{c veto, b veto}$	364114	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, \max(H_T, p_T^V) < 70\text{GeV}, \text{c filter, b veto}$	364115	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, \max(H_T, p_T^V) < 70\text{GeV}, \text{b filter}$	364116	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, \text{c veto, b veto}$	364117	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, \text{c filter, b veto}$	364118	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, \text{b filter}$	364119	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, \text{c veto, b veto}$	364120	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, \text{c filter, b veto}$	364121	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, \text{b filter}$	364122	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, \text{c veto, b veto}$	364123	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, \text{c filter, b veto}$	364124	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, \text{b filter}$	364125	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, 500\text{GeV} < \max(H_T, p_T^V) < 1000\text{GeV}$	364126	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow ee, \max(H_T, p_T^V) > 1000\text{GeV}$	364127	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, \max(H_T, p_T^V) < 70\text{GeV}, \text{c veto, b veto}$	364100	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, \max(H_T, p_T^V) < 70\text{GeV}, \text{c filter, b veto}$	364101	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, \max(H_T, p_T^V) < 70\text{GeV}, \text{b filter}$	364102	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, \text{c veto, b veto}$	364103	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, \text{c filter, b veto}$	364104	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, \text{b filter}$	364105	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, \text{c veto, b veto}$	364106	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, \text{c filter, b veto}$	364107	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, \text{b filter}$	364108	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, \text{c veto, b veto}$	364109	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, \text{c filter, b veto}$	364110	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, \text{b filter}$	364111	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 500\text{GeV} < \max(H_T, p_T^V) < 1000\text{GeV}$	364112	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, \max(H_T, p_T^V) > 1000\text{GeV}$	364113	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, \max(H_T, p_T^V) < 70\text{GeV}, \text{c veto, b veto}$	364128	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, \max(H_T, p_T^V) < 70\text{GeV}, \text{c filter, b veto}$	364129	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, \max(H_T, p_T^V) < 70\text{GeV}, \text{b filter}$	364130	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, \text{c veto, b veto}$	364131	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, \text{c filter, b veto}$	364132	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 70\text{GeV} < \max(H_T, p_T^V) < 140\text{GeV}, \text{b filter}$	364133	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, \text{c veto, b veto}$	364134	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, \text{c filter, b veto}$	364135	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 140\text{GeV} < \max(H_T, p_T^V) < 280\text{GeV}, \text{b filter}$	364136	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, \text{c veto, b veto}$	364137	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, \text{c filter, b veto}$	364138	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 280\text{GeV} < \max(H_T, p_T^V) < 500\text{GeV}, \text{b filter}$	364139	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 500\text{GeV} < \max(H_T, p_T^V) < 1000\text{GeV}$	364140	SHERPA2.2.1	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, \max(H_T, p_T^V) > 1000\text{GeV}$	364141	SHERPA2.2.1	NNPDF3.0 NNLO

Table 13: List of MC16a (r-tag 9364) samples used in this search for the Z+jets background. All samples are available and used for MC16d (r-tag 10201) and MC16e (r-tag 10724). All processes are simulated for up to 2 partons at NLO and 4 partons at LO.

Process	MC DSID	Generator	Generator Tune
Z+Jets+γ, Full Simulation			
$Z \rightarrow ee, 7\text{GeV} < p_T^\gamma < 15\text{GeV}$	366140	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow ee, 15\text{GeV} < p_T^\gamma < 35\text{GeV}$	366141	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow ee, 35\text{GeV} < p_T^\gamma < 70\text{GeV}$	366142	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow ee, 70\text{GeV} < p_T^\gamma < 140\text{GeV}$	366143	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow ee, p_T^\gamma > 140\text{GeV}$	366144	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 7\text{GeV} < p_T^\gamma < 15\text{GeV}$	366145	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 15\text{GeV} < p_T^\gamma < 35\text{GeV}$	366146	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 35\text{GeV} < p_T^\gamma < 70\text{GeV}$	366147	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, 70\text{GeV} < p_T^\gamma < 140\text{GeV}$	366148	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow \mu\mu, p_T^\gamma > 140\text{GeV}$	366149	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 7\text{GeV} < p_T^\gamma < 15\text{GeV}$	366150	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 15\text{GeV} < p_T^\gamma < 35\text{GeV}$	366151	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 35\text{GeV} < p_T^\gamma < 70\text{GeV}$	366152	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, 70\text{GeV} < p_T^\gamma < 140\text{GeV}$	366153	SHERPA2.2.4	NNPDF3.0 NNLO
$Z \rightarrow \tau\tau, p_T^\gamma > 140\text{GeV}$	366154	SHERPA2.2.4	NNPDF3.0 NNLO
W+Jets+γ, Full Simulation			
$W \rightarrow ev, 7\text{GeV} < p_T^\gamma < 15\text{GeV}$	364521	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow ev, 15\text{GeV} < p_T^\gamma < 35\text{GeV}$	364522	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow ev, 35\text{GeV} < p_T^\gamma < 70\text{GeV}$	364523	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow ev, 70\text{GeV} < p_T^\gamma < 140\text{GeV}$	364524	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow ev, p_T^\gamma > 140\text{GeV}$	364525	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 7\text{GeV} < p_T^\gamma < 15\text{GeV}$	364526	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 15\text{GeV} < p_T^\gamma < 35\text{GeV}$	364527	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 35\text{GeV} < p_T^\gamma < 70\text{GeV}$	364528	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, 70\text{GeV} < p_T^\gamma < 140\text{GeV}$	364529	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow \mu\nu, p_T^\gamma > 140\text{GeV}$	364530	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 7\text{GeV} < p_T^\gamma < 15\text{GeV}$	364531	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 15\text{GeV} < p_T^\gamma < 35\text{GeV}$	364532	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 35\text{GeV} < p_T^\gamma < 70\text{GeV}$	364533	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, 70\text{GeV} < p_T^\gamma < 140\text{GeV}$	364534	SHERPA2.2.2	NNPDF3.0 NNLO
$W \rightarrow \tau\nu, p_T^\gamma > 140\text{GeV}$	364535	SHERPA2.2.2	NNPDF3.0 NNLO

Table 14: List of MC16a (r-tag 9364) samples used in this search for the W+jets+ γ and Z+jets+ γ backgrounds. All samples are available and used for MC16d (r-tag 10201) and MC16e (r-tag 10724). All processes are simulated for up to 1 parton at NLO and 3 partons at LO.

666 C Complete List of Monte Carlo Samples Used**667 D Additional Shape Comparison Plots: $\mu+jets$ channel**

668 Various additional plots are shown in this appendix from the neural network creation and studies. Figure 31
669 and 32 show additional shape comparisons in variables which are not included in the final neural network
670 model as they do not significantly change the fit values. In the cases of p_T or E variables with the higher
671 separation value were used as there is a large correlation between the two values and the other is shown in
672 this appendix. ΔR_{jb} was not included as the other 3 ΔR values had higher separation values and they are
673 all related to each other as they are the geometrically related.

674 The neutrino reconstruction is done using a minimization of

$$\chi^2_\nu = \chi^2_{bW} + \chi^2_W$$

675 . All three were investigated for their separation values and the χ^2_W value had the largest separation.

676 E Shape Comparison Plots: $e+jets$ channel

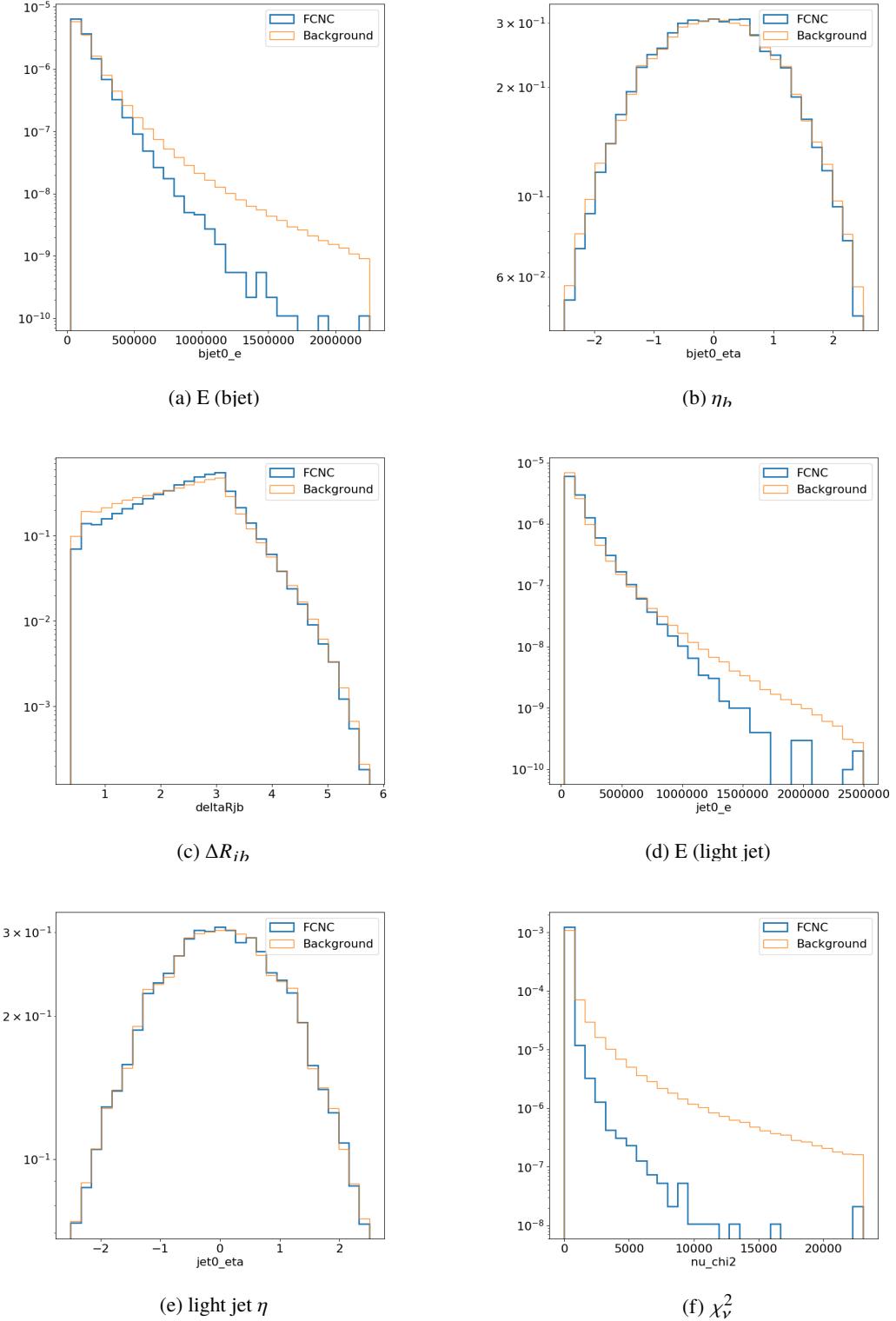


Figure 31: Normalized variables showing the shapes of neural network input variables for the $\mu + \text{jets}$ channel: [E (bjet), η_b , ΔR_{jb} , E (light jet), light jet η , and χ^2_ν the total χ^2 fit value]

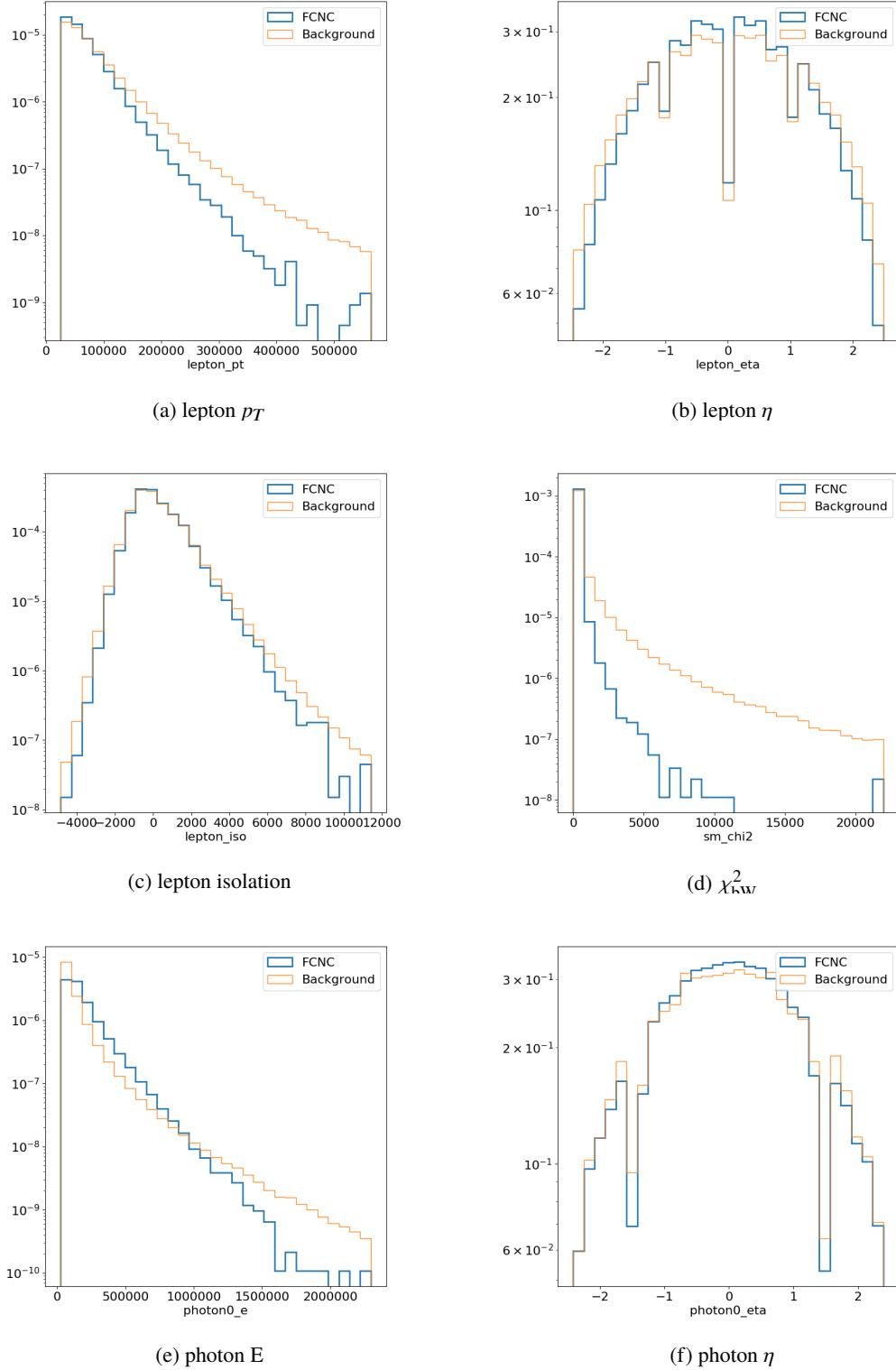


Figure 32: Normalized variables showing the shapes of neural network input variables for the $\mu+\text{jets}$ channel: [lepton p_T , lepton η , lepton isolation , χ^2_{bW} the bW χ^2 value from neutrino reconstruction ,photon E, and photon η .

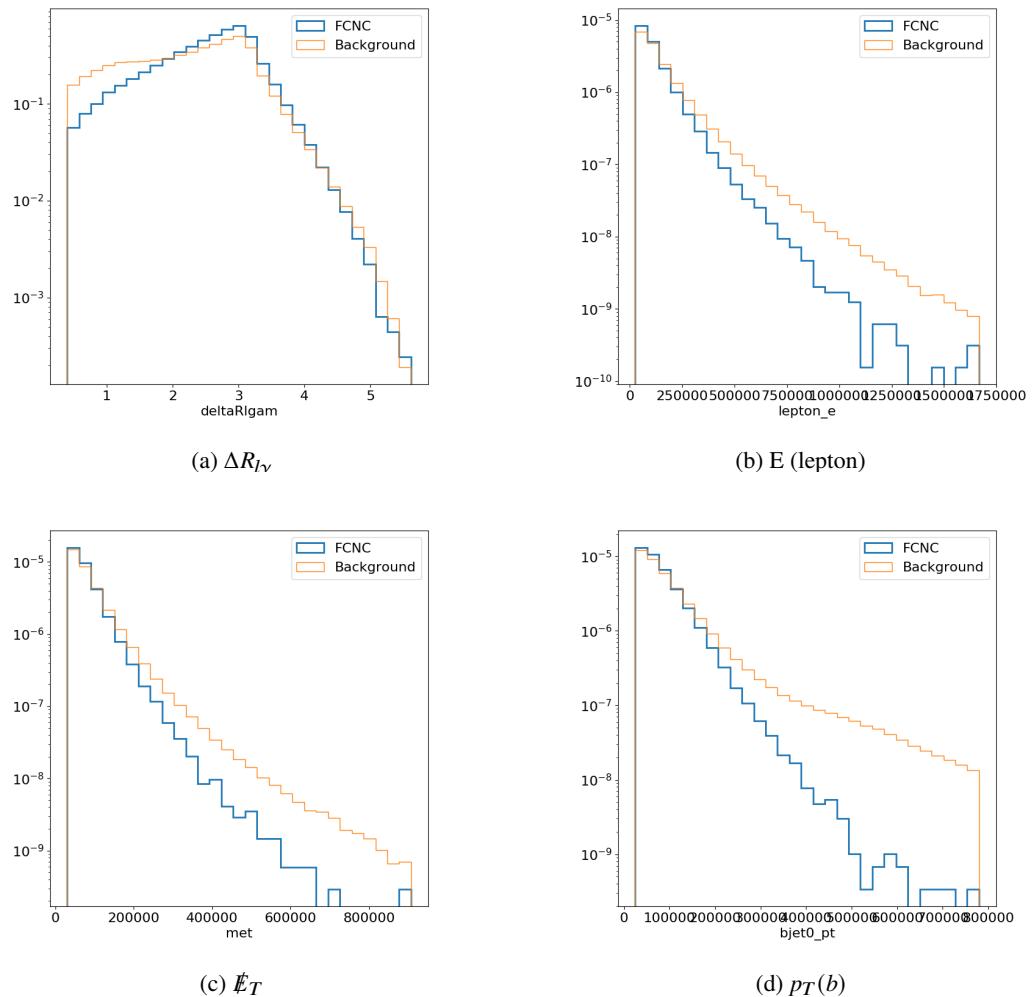


Figure 33: Normalized variables showing the shapes of neural network input variables for the $e + \text{jets}$ channel: $\Delta R_{l\gamma}$, E (lepton), \cancel{E}_T , and $p_T(b)$

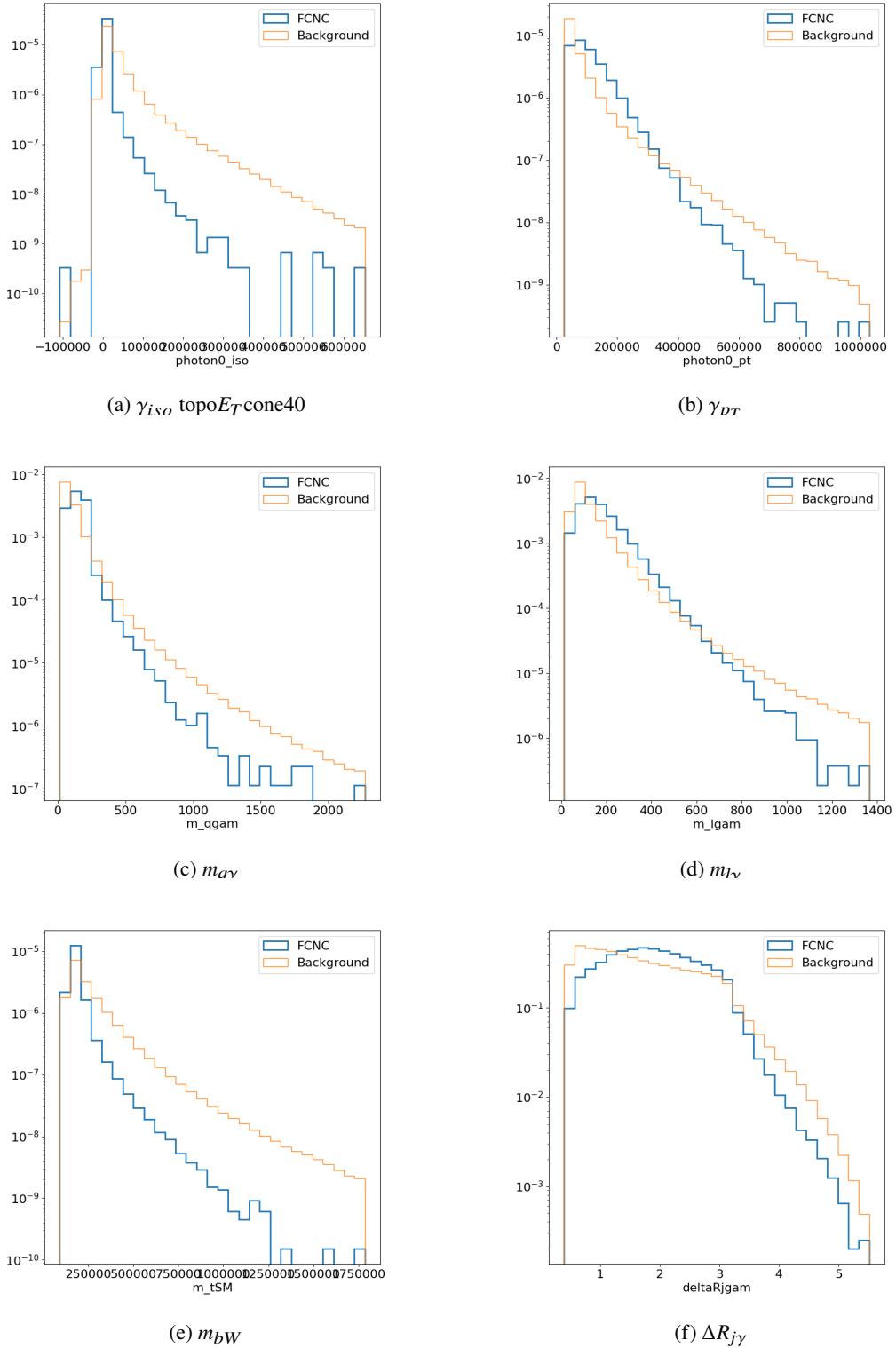


Figure 34: Normalized variables showing the shapes of neural network input variables for the $e+jets$ channel: γ_{iso} topo E_T cone40, γ_{pT} , $m_{q\gamma}$, $m_{l\gamma}$, m_{bW} , and $\Delta R_{j\gamma}$

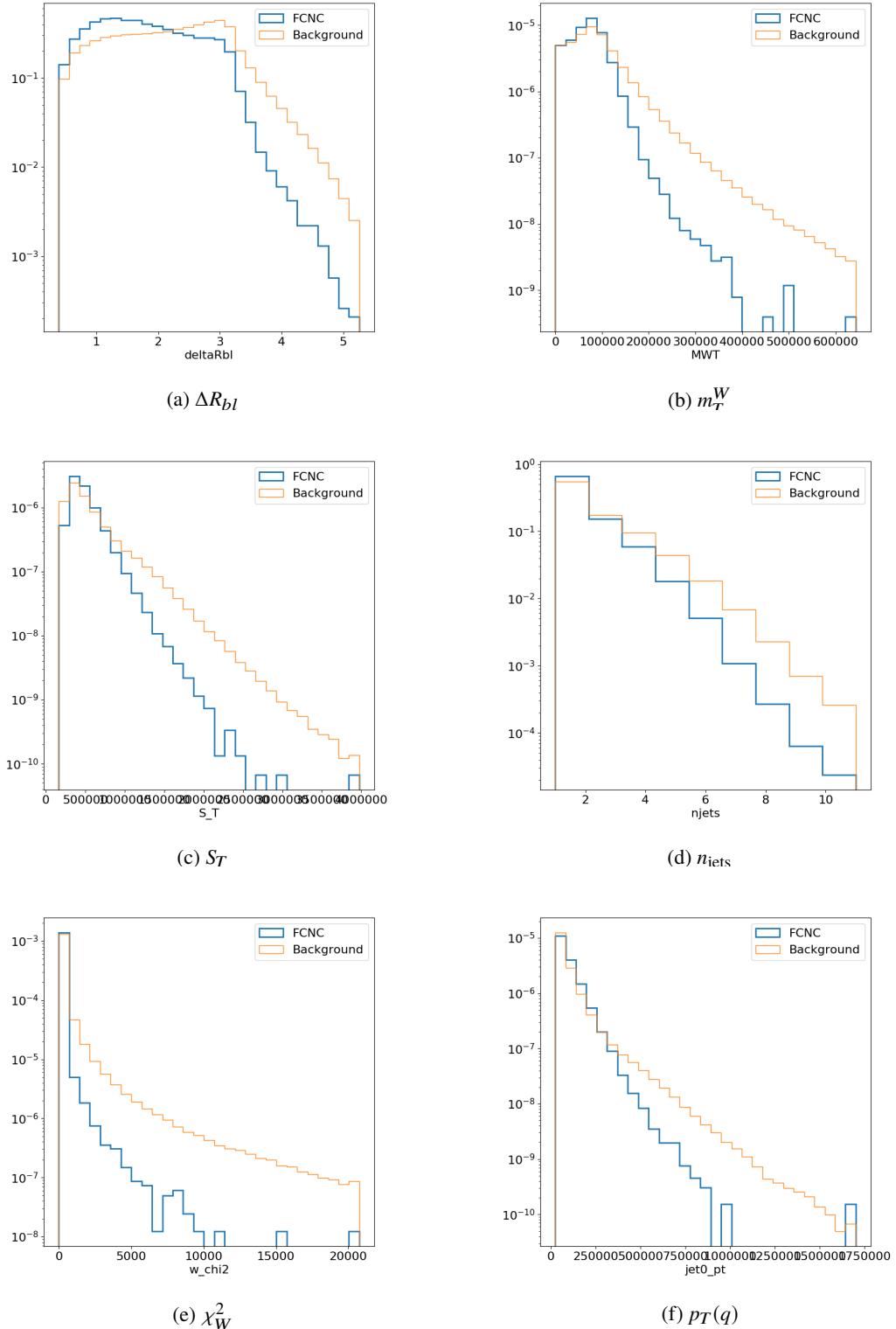


Figure 35: Normalized variables showing the shapes of neural network input variables for the $e+\text{jets}$ channel: ΔR_{bl} , m_T^W , S_T , n_{jets} , χ^2_W , and $p_T(q)$

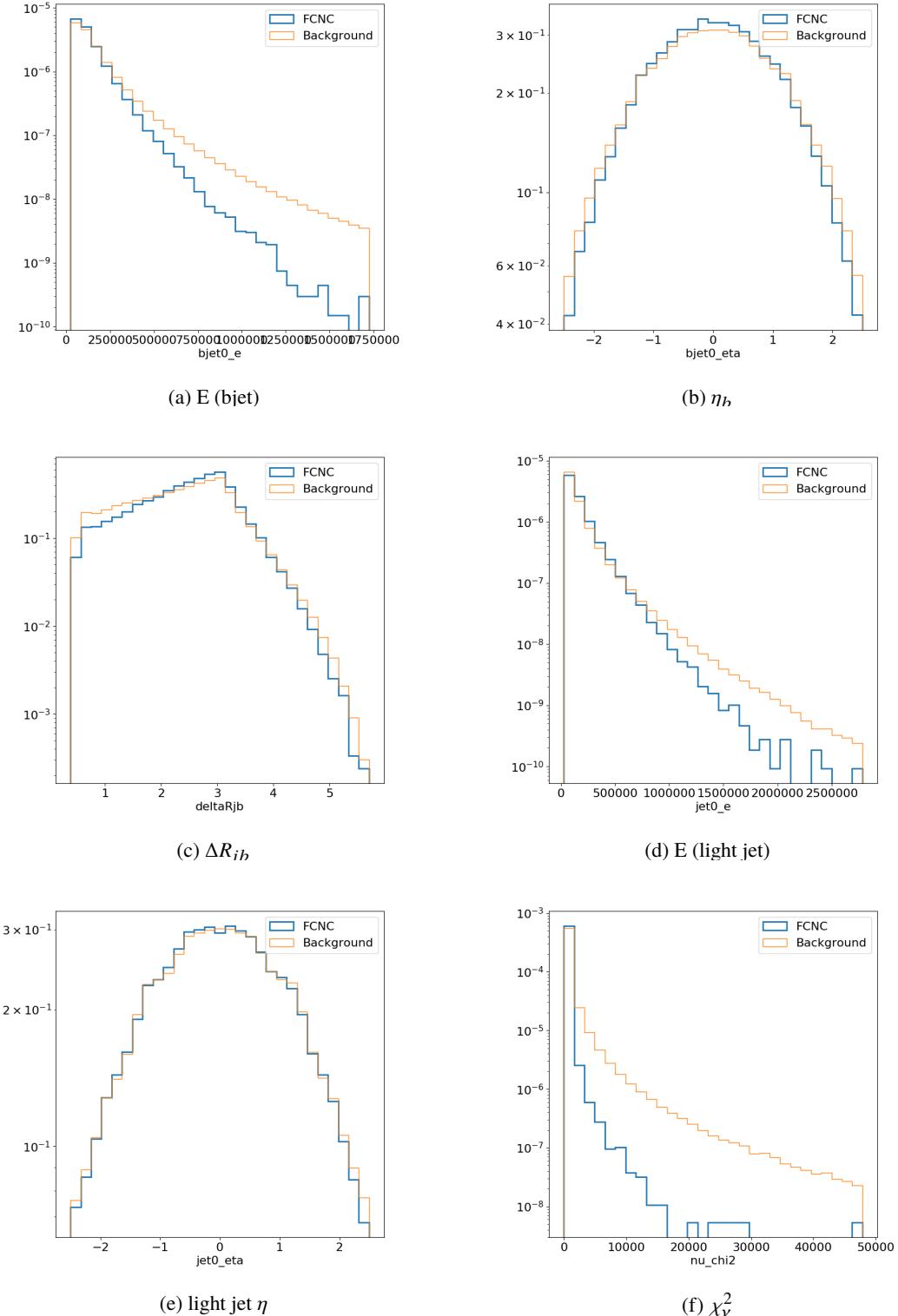


Figure 36: Normalized variables showing the shapes of neural network input variables for the $e+\text{jets}$ channel: [E (bjet), η_b , ΔR_{jb} , E (light jet), light jet η , and χ^2_ν the total χ^2 fit value]

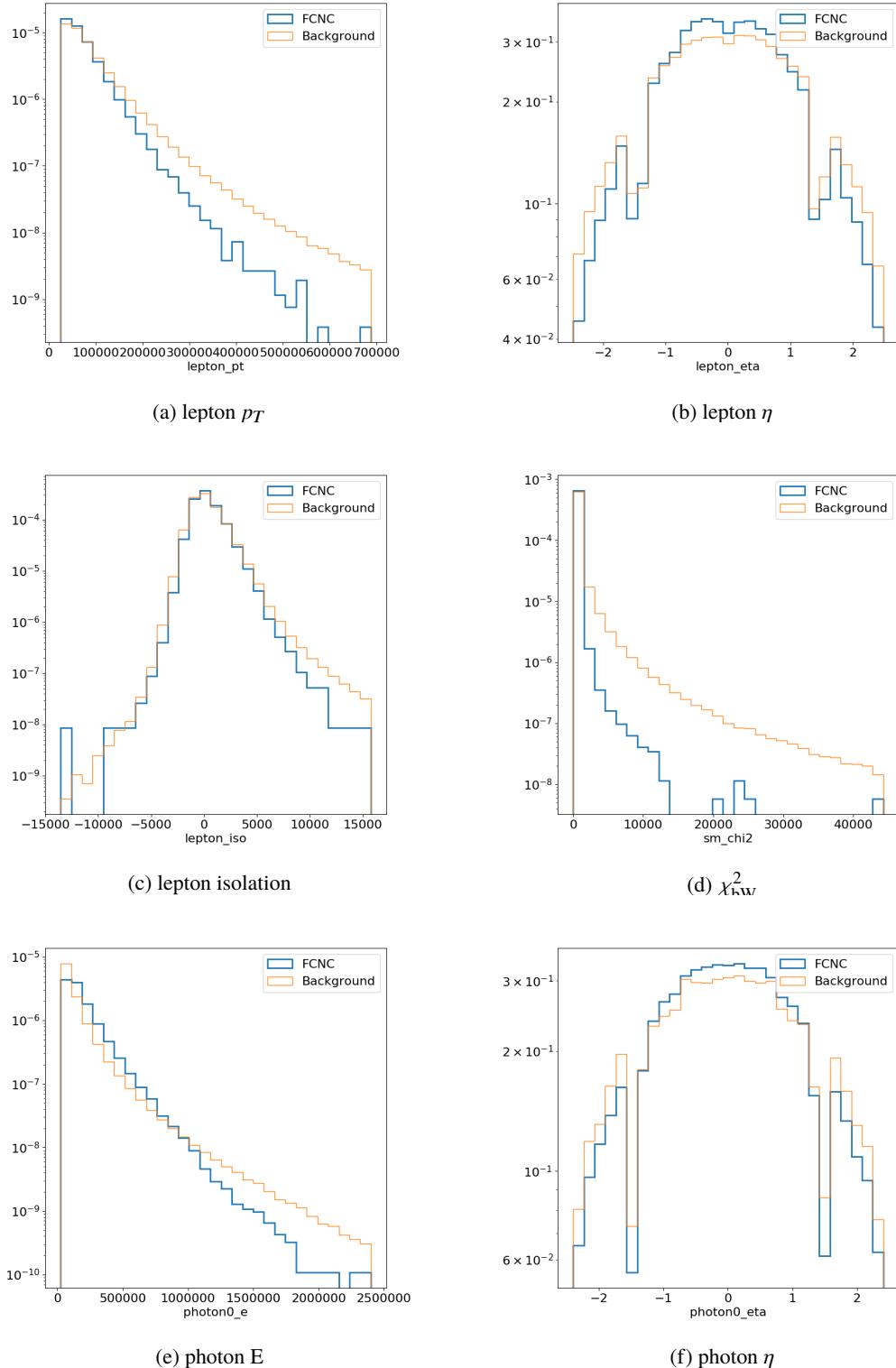


Figure 37: Normalized variables showing the shapes of neural network input variables for the $e+jets$ channel: [lepton p_T , lepton η , lepton isolation , χ^2_{bW} the bW χ^2 value from neutrino reconstruction, photon E, and photon η .