WWZ Analysis: BDT Summary

Keegan Downham, UCSB

WWZ Group Meeting: 26 September 2023

Baseline (Cut-Based Analysis)

Which bins are the most sensitive?

No changes in selection since numbers were shared w/ Kelci

SR	Summary			Composition of N_{total}				
Bins	N_{total}	NonResonant WWZ	ZHWWZ	ZZ	ttZ	Higgs	WZ	Other
Bin1	1.33 ± 0.21	0.65 ± 0.01	3.05 ± 0.01	0.62 ± 0.02	0.32 ± 0.03	0.27 ± 0.19	0.11 ± 0.08	0.01 ± 0.01
Bin2	0.96 ± 0.11	0.74 ± 0.01	1.35 ± 0.01	0.62 ± 0.02	0.38 ± 0.03	-0.06 ± 0.10	0 ± 0	0.01 ± 0.01
Bin3	1.39 ± 0.22	1.48 ± 0.01	0.35 ± 0.01	0.39 ± 0.02	0.83 ± 0.05	-0.06 ± 0.16	0.18 ± 0.13	0.04 ± 0.04
Bin4	3.60 ± 0.23	5.14 ± 0.01	0.15 ± 0.01	0.50 ± 0.02	2.51 ± 0.09	0.45 ± 0.17	0.12 ± 0.12	0.02 ± 0.01

Table 19: Yields in $e\mu$ SR bins (as of September 1st, 2023)

SR	Summary			Composition of N_{total}				
Bins	N_{total}	NonResonant WWZ	ZHWWZ	ZZ	ttZ	Higgs	WZ	Other
Bin5	3.25 ± 0.22	2.33 ± 0.01	0.95 ± 0.01	1.32 ± 0.03	1.32 ± 0.06	0.61 ± 0.21	0 ± 0	0.01 ± 0.01
Bin6	6.18 ± 0.36	1.97 ± 0.01	1.52 ± 0.01	4.58 ± 0.05	1.12 ± 0.06	0.42 ± 0.35	0 ± 0	0.06 ± 0.04
Bin7	3.15 ± 0.18	0.57 ± 0.01	0.68 ± 0.01	2.78 ± 0.04	0.27 ± 0.03	0.10 ± 0.17	0 ± 0	0.01 ± 0.01

Table 20: Yields in $ee/\mu\mu$ SR bins (as of September 1st, 2023)

Training Region (TR) Selection

4-lepton analysis selection criteria	
Category	Cut
Preselection	
Triggers	Passes at least one dilepton ($ee/e\mu/\mu\mu$) trigger
Lepton ID	4 leptons (e/μ) passing "tight" requirements of TOP-UL-MVA ID
Lepton p_T	$p_T^{leading} > 25$ GeV, $p_T^{subleading} > 15$ GeV, $p_T^{3rd,4th} > 10$ GeV
Z-candidate leptons	same-flavor, opposite sign pair closest to m_Z
Z mass window	Z candidate leptons must satisfy $\lvert m_{ll} - m_Z \rvert < 10$ GeV
Additional lepton requirements	Muons (Electrons) must satisfy $\lvert IP_{3D}/\sigma_{IP_{3D}} \rvert < 4$ (and $I_{rel,03,all} < 0.2$)
QCD low mass resonance veto	Any opposite charge pair of leptons must have $m_{ll} > 12$ GeV
b-tagged jet veto	Require 0 b-tagged jets in event ($p_T > 20$ GeV, $\lvert\eta\rvert < 2.4$, DeepCSV loose WP, loose Jet ID)
Same-Flavor Channel	W-candidate leptons are the same flavor
Off-Z peak requirement	W candidate leptons must have $\lvert m_{\ell\ell}$ - $m_Z \rvert > 10$ GeV
Signal Regions - Cut Based	
Bin A	$p_T^{miss} > 120$ GeV
Bin B	70 GeV $< p_T^{miss} <$ 120 GeV, 40 GeV $< p_T^{4\ell} <$ 70 GeV
Bin C	70 GeV $< p_T^{miss} <$ 120 GeV, $p_T^{4\ell} > 70$ GeV
Opposite-Flavor Channel	W-candidate leptons are different flavor
Signal Regions - Cut Based	
Bin 1	$m_{T2} > 25$ GeV, 0 GeV $< m_{\ell\ell}^{W} cands <$ 40 GeV
Bin 2	$m_{T2} > 25$ GeV, 40 GeV $< m_{\ell\ell}^{W} cands <$ 60 GeV
Bin 3	$m_{T2} > 25$ GeV, 60 GeV $< m_{\ell\ell}^{W} cands <$ 100 GeV
Bin 4	$m_{\ell\ell}^{W} cands > 100$ GeV

- Training region selection is a subset of the full cut-based selection
- Training region selection is shown in purple
  - Note: no additional cuts applied for opposite-flavor channel
- Cut-based SRs are a combination of purple and black
  - I.e. preselection + channel specific selection

Yields after TR selection

	Summary			Composition of N_{total}				
	N_{total}	NonResonant WWZ	ZHWWZ	ZZ	ttZ	Higgs	WZ	Other
Opp Flav Yield	27.3601 ± 0.5570	9.2928 ± 0.0169	7.933 ± 0.0202	20.487 ± 0.110	4.6626 ± 0.1180	1.5297 ± 0.4641	0.5747 ± 0.2592	0.1058 ± 0.0406
Opp Flav Entries	45517	386138	562626	39470	5743	146	6	152
Same Flav Yield	681.024 ± 1.117	8.5219 ± 0.0162	7.9725 ± 0.0202	671.271 ± 0.617	4.3152 ± 0.1136	5.2241 ± 0.9202	0.1066 ± 0.0755	0.1073 ± 0.0404
Same Flav Entries	1660574	354713	566117	1654879	5316	227	2	150

Table 24: Yields in training region for all processes.

- ZZ and ttZ are the only backgrounds used in training
    - Other backgrounds are limited by low statistics

Training+Testing samples

- 4 BDTs are trained using TMVA
  - W candidate leptons are same flavor (ee/μμ) or opposite flavor (eμ)
    - WWZ vs Backgrounds (ZZ + ttZ)
    - ZH→WWZ vs Backgrounds (ZZ + ttZ)
- Training+Testing samples obtained by splitting full MC samples in half
  - Alternate putting events into training+testing samples when looping over events
- Full list of samples shown below (includes full Run 2 MC)

Process	Sample Names (NanoAOD v9)
WWZ	WWZJetsTo4L2Nu_4F_TuneCP5_13TeV-amcatnlo-pythia8_RunIISummer20UL1(6/7/8)*
ZH→WWZ	GluGluZH_HToWWTo2L2Nu_M-125_TuneCP5_13TeV-powheg-pythia8_RunIISummer20UL1(6/7/8)* HZJ_HToWWTo2L2Nu_ZTo2L_M-125_TuneCP5_13TeV-powheg-jhugen727-pythia8_RunIISummer20UL1(6/7/8)*
Backgrounds	ZZTo4L_TuneCP5_13TeV_powheg_pythia8_RunIISummer20UL1(6/7/8)* ZZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8_RunIISummer20UL1(6/7/8)* ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8_RunIISummer20UL1(6/7/8)* GluGluToContinToZZTo(2e2mu/2e2tau/2mu2tau/4e/4mu/4tau)_TuneCP5_13TeV-mcfm701-pythia8_RunIISummer20UL1(6/7/8)* TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8_RunIISummer20UL1(6/7/8)* TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8_RunIISummer20UL1(6/7/8)* TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8_RunIISummer20UL1(6/7/8)*

Splitting Cross Check

The table below shows the yields for the training and testing datasets

Sample	SF Training	SF Testing	Ratio (SF)	DF Training	DF Testing	Ratio (DF)
WWZ	4.25206 ± 0.011483	4.25563 ± 0.0114843	0.99916	4.63711 ± 0.011984	4.63956 ± 0.011984	0.99947
ZH	4.00546 ± 0.014288	7.30354 ± 0.018344	0.54843	3.98529 ± 0.014279	7.31073 ± 0.018378	0.54513
Σ(ZZ,ttZ)	336.588 ± 0.445	336.658 ± 0.445	0.99979	12.54460 ± 0.11383	12.60440 ± 0.11382	0.99526

Note: ZH is not split 1/1 → Performed some checks, still unsure why this is happening (however this really doesn't matter too much for the interpretation of results)

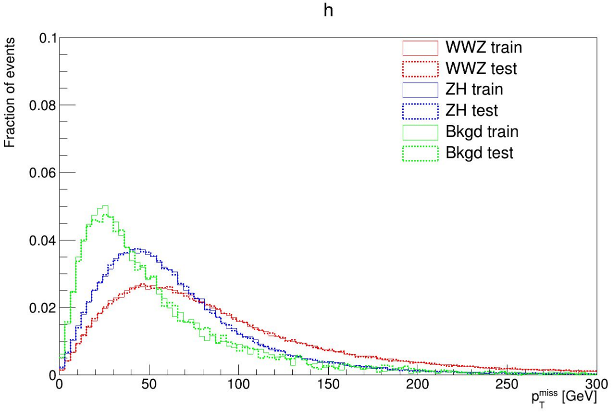
WWZ and Backgrounds have very similar statistics for training + testing datsets

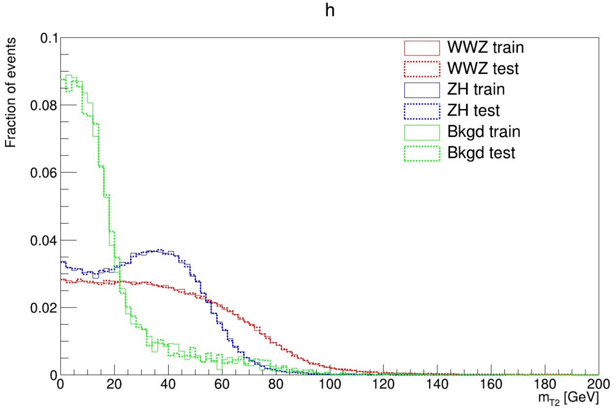
BDT Input Variables

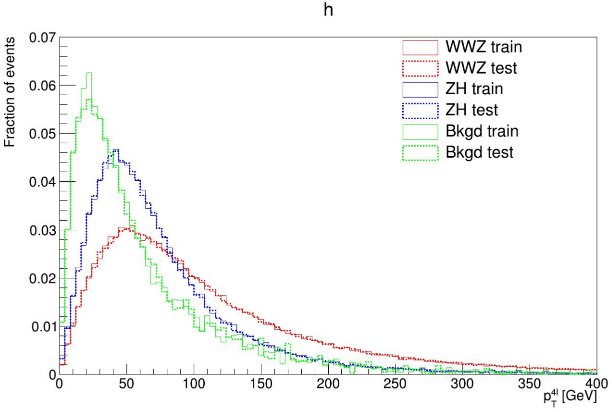
Input variables for 4-lepton MVA	
Variable	Brief Description
$m_{ll}^{W\,cands}$	Invariant mass of the W-lepton candidates
M_{T2}	Analogue of m_T for two semi-invisibly decaying particles
p_T^{miss}	Missing Transverse Energy (PuppiMET)
p_T^{4l}	Transverse momentum of the 4-lepton system
p_T^{Z1}	p_T of leading Z-candidate lepton
p_T^{Z2}	p_T of subleaing Z-candidate lepton
p_T^{W1}	p_T of leading W-candidate lepton
p_T^{W2}	p_T of subleading W-candidate lepton
$\sum_{lep,MET,jet} p_T$	Scalar sum of transverse energy in event
$\sum_{lep,MET} p_T$	Scalar sum of leptonic and missing transverse energy in event
$\Delta R(l^{Z1}, l^{Z2})$	Solid angle (in $\eta - \phi$ coordinates) difference between Z-candidate leptons
$\Delta R(l^{W1}, l^{W2})$	Solid angle (in $\eta - \phi$ coordinates) difference between W-candidate leptons
$\Delta R(WW, Z)$	Solid angle (in $\eta - \phi$ coordinates) difference between W candidate lepton system and Z boson
$\Delta\phi(WW, p_T^{miss})$	Angular separation between WW system and p_T^{miss}
$\Delta\phi(Z, p_T^{miss})$	Angular separation between Z boson and p_T^{miss}
$\Delta\phi(WWZ, p_T^{miss})$	Angular separation between the 4-lepton system and p_T^{miss}

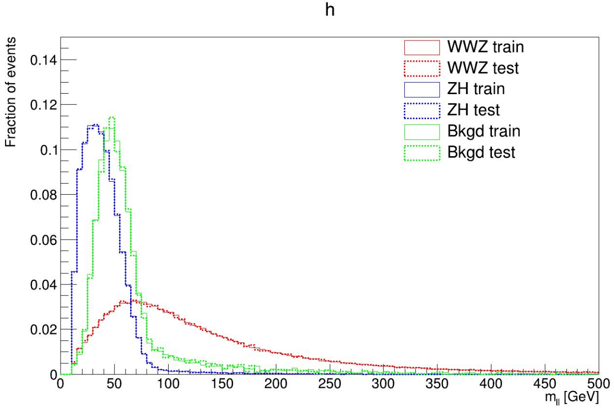
Table 15: Input variables for 4-lepton Boosted Decision Tree.

BDT Input Variables - Opposite Flavor



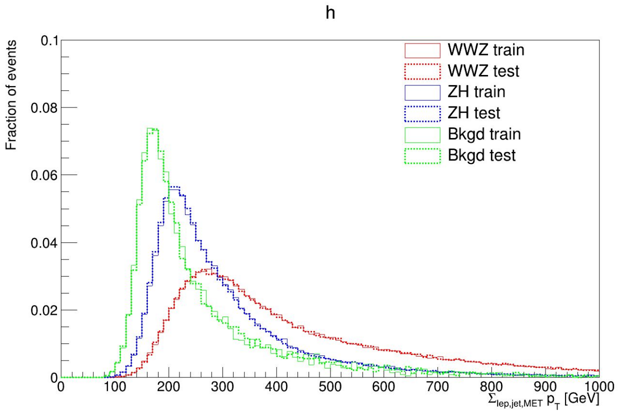


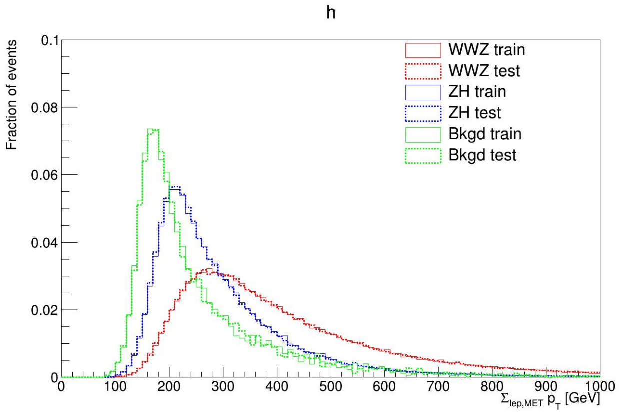


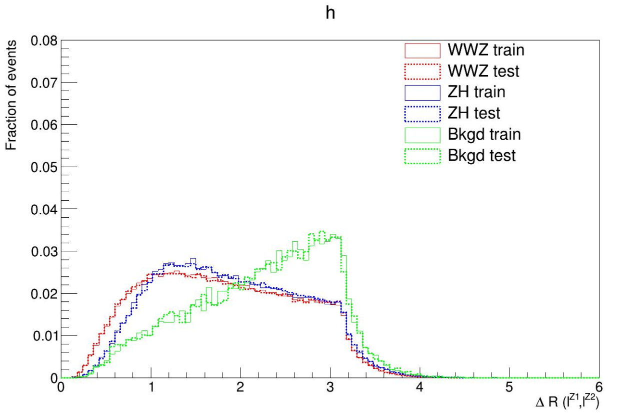


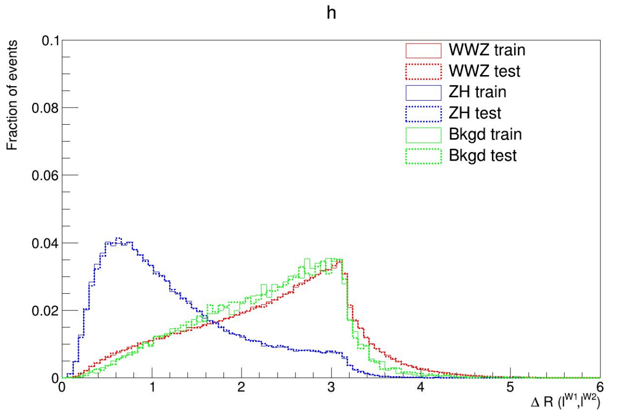
<u>All distributions normalized to 1!</u>

BDT Input Variables - Opposite Flavor (cont.)



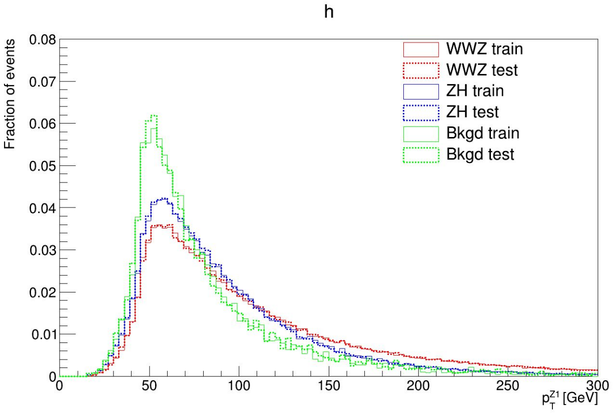


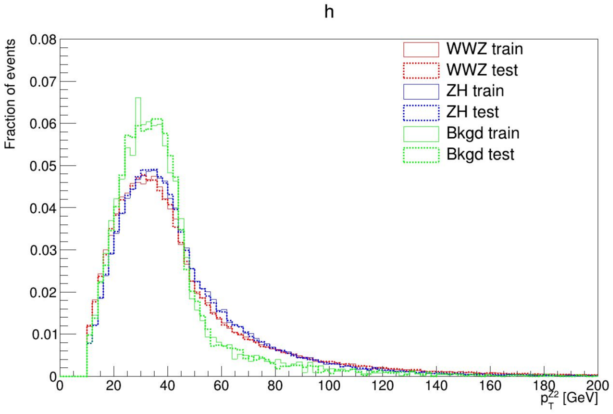


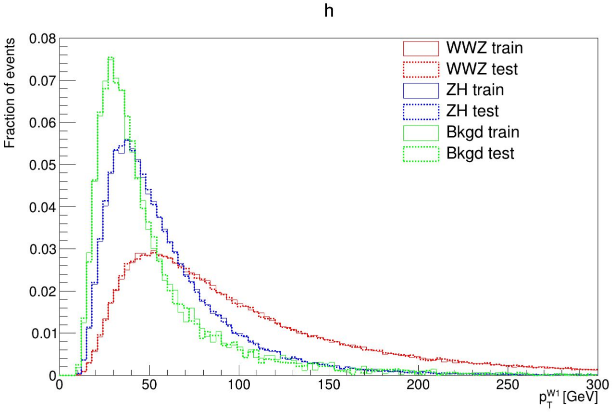


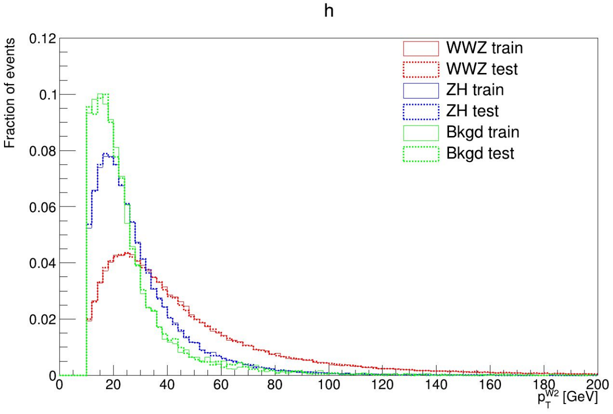
<u>All distributions normalized to 1!</u>

BDT Input Variables - Opposite Flavor (cont.)



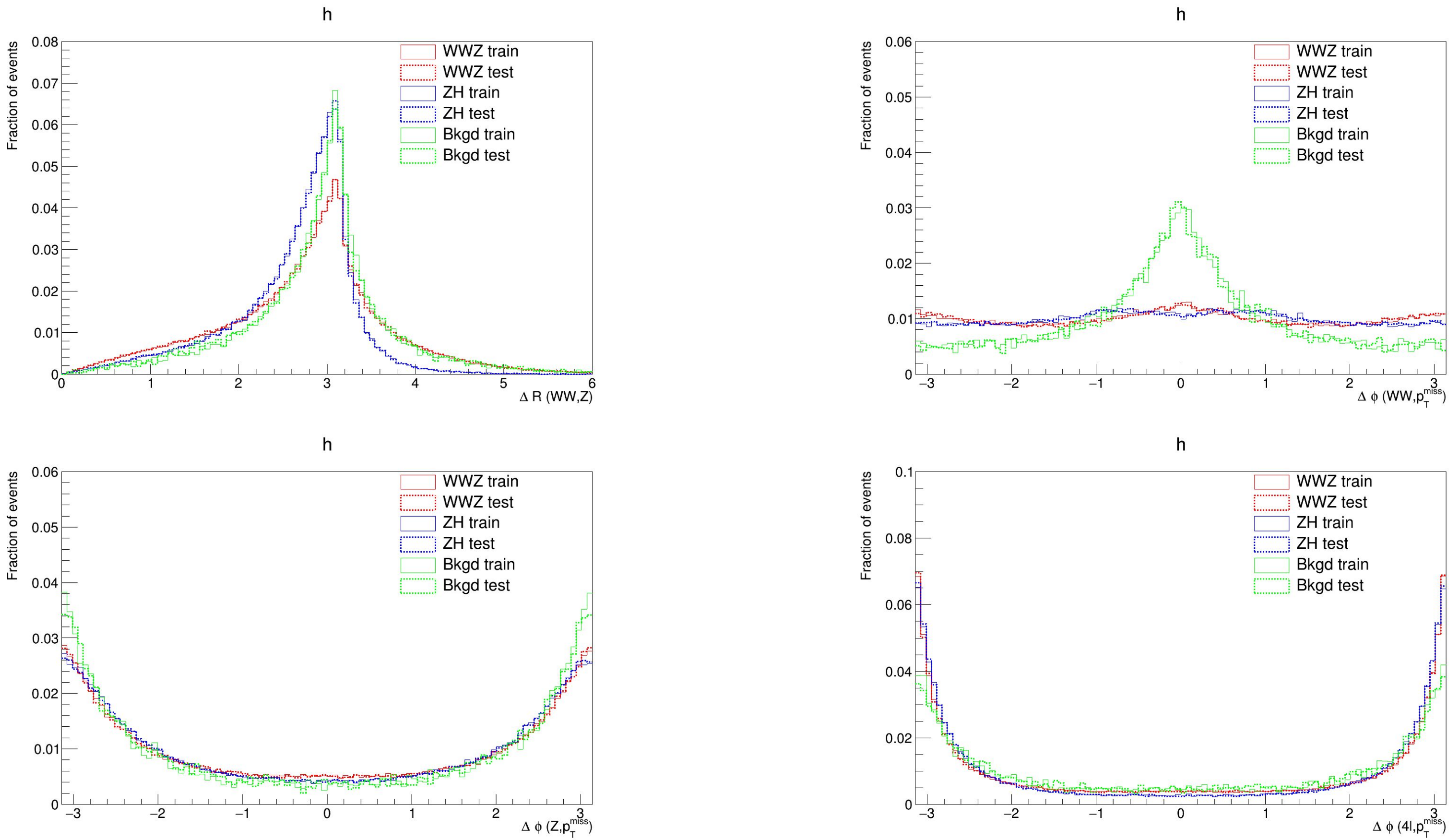






<u>All distributions normalized to 1!</u>

BDT Input Variables - Opposite Flavor (cont.)

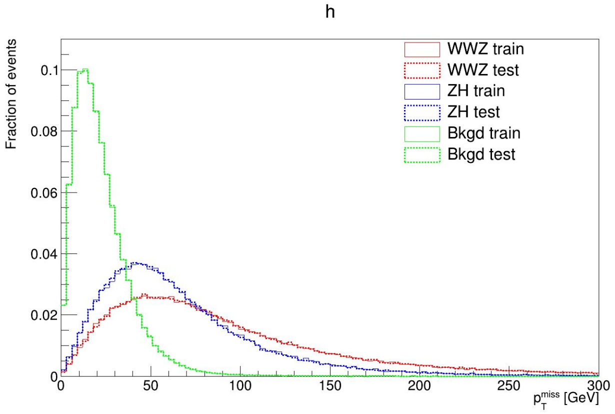


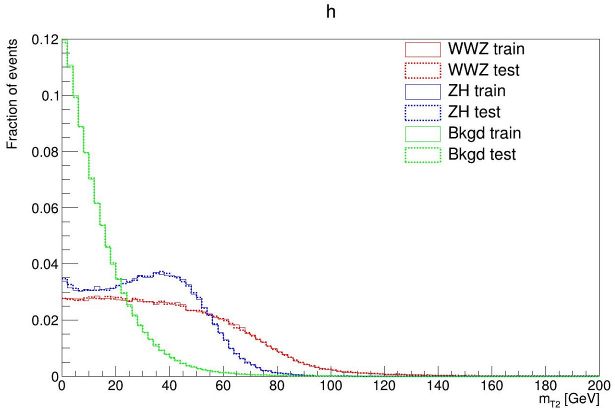
All distributions normalized to 1!

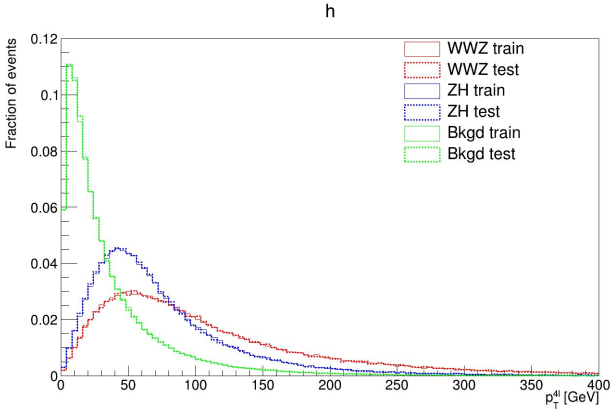
Variable Rankings - Opposite Flavor

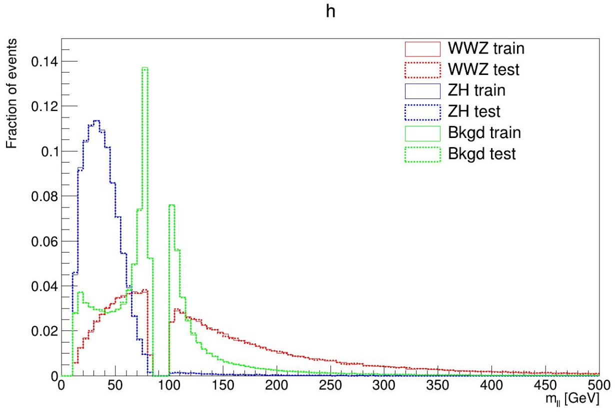
- Good variables
  - m_{ll}, m_{T2}, $dR(l^{W1},l^{W2})$
- Minor variables → vars we may want to drop eventually
  - $\Delta\phi$(4l, MET), $\Delta\phi$(Z,MET), STLepHad or STLep
- Bad variables →vars that may cause problems (i.e. bad data/MC)
  -

BDT Input Variables - Same Flavor



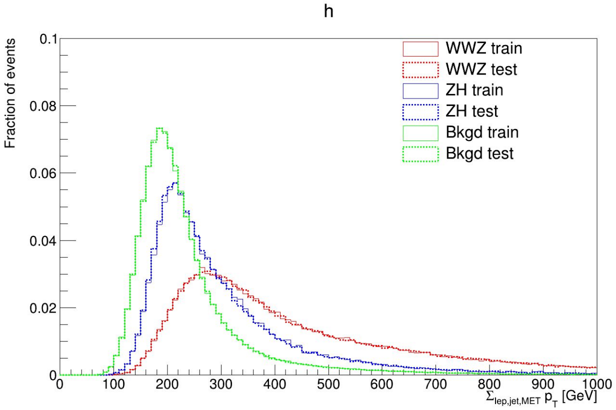


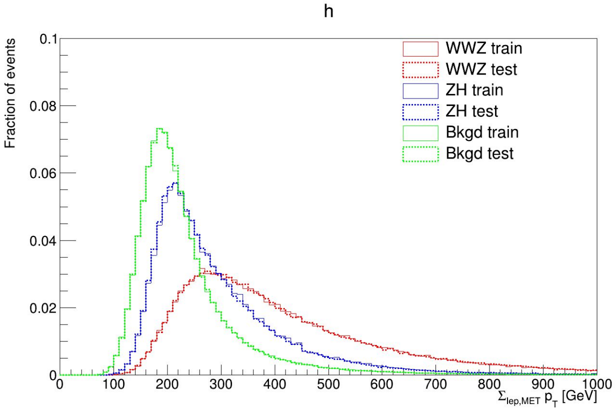


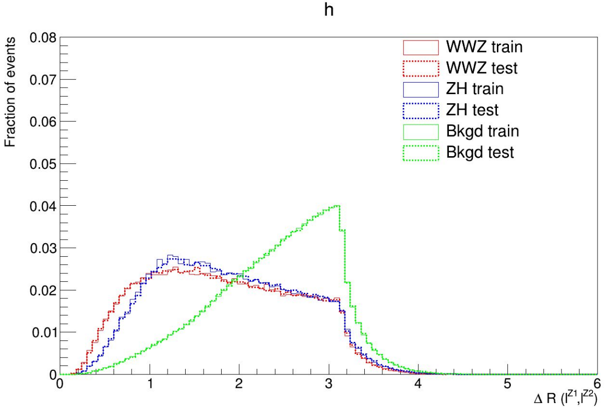


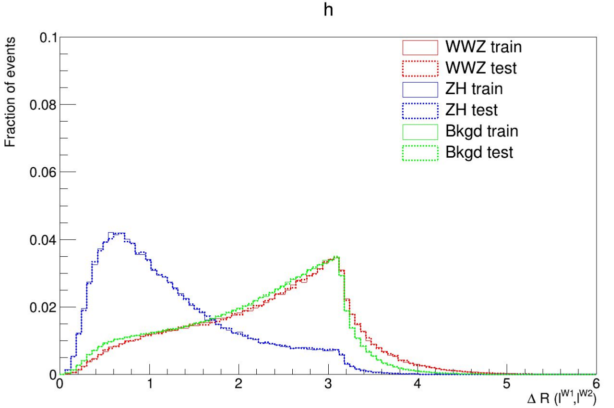
All distributions normalized to 1!

BDT Input Variables - Same Flavor (cont.)



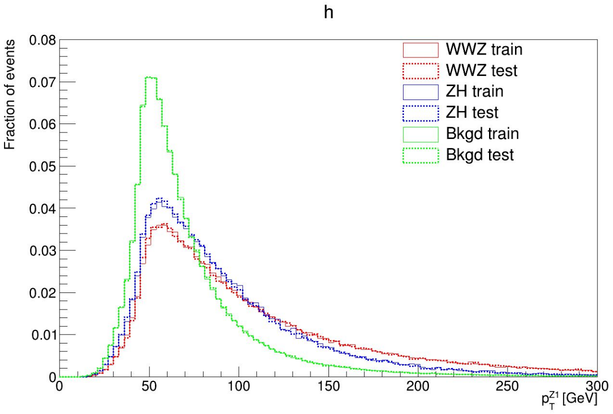


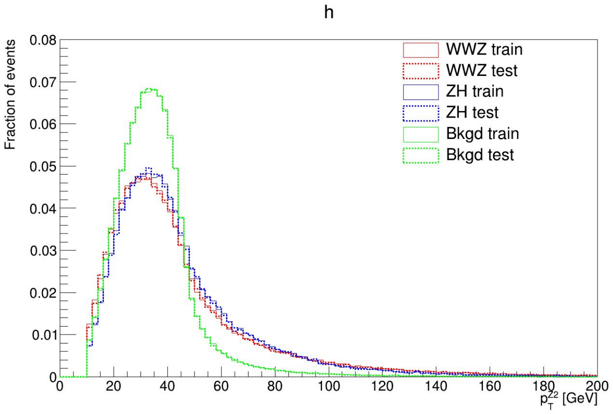


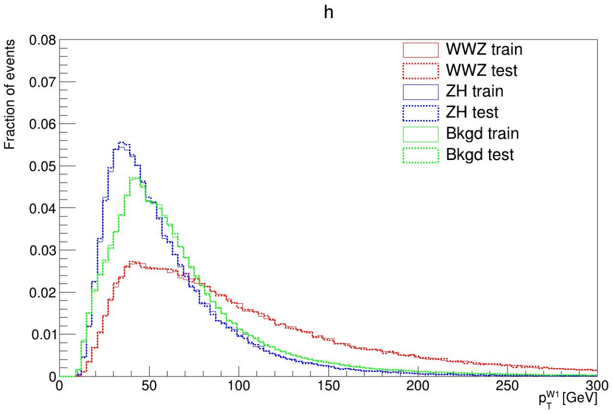


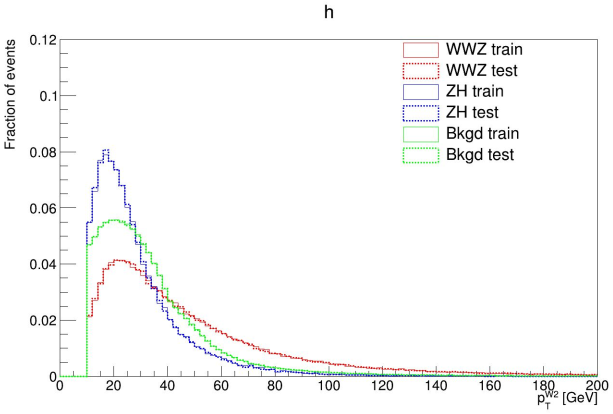
All distributions normalized to 1!

BDT Input Variables - Same Flavor (cont.)



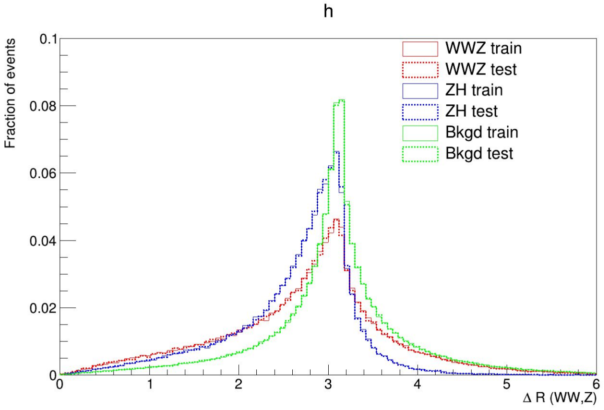


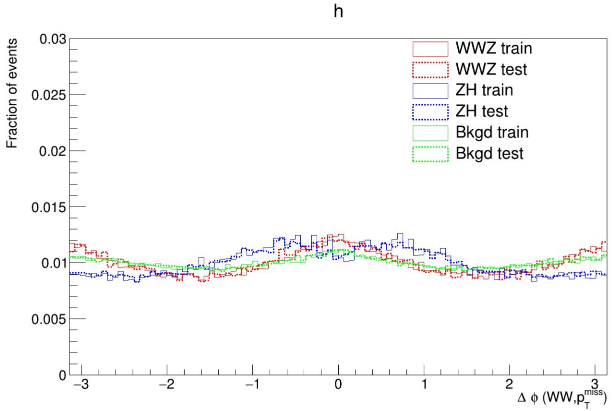


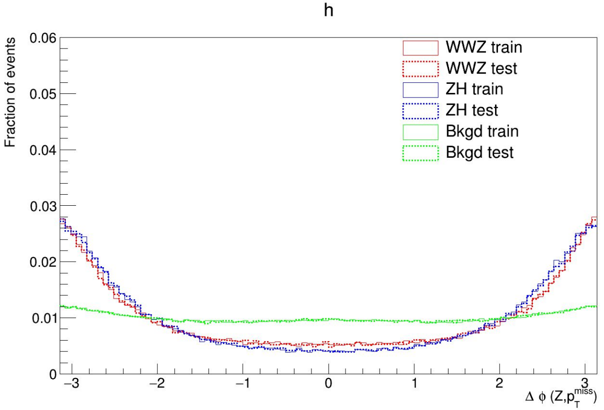


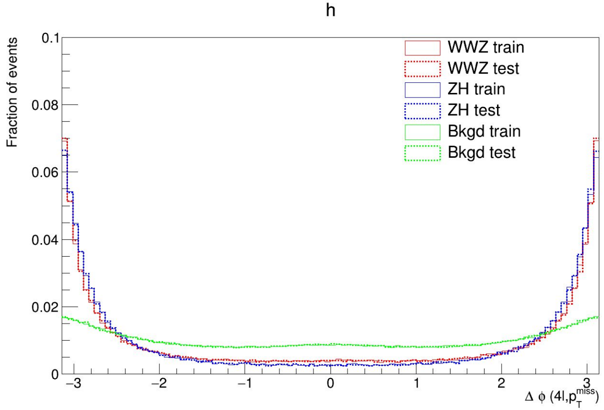
All distributions normalized to 1!

BDT Input Variables - Same Flavor (cont.)







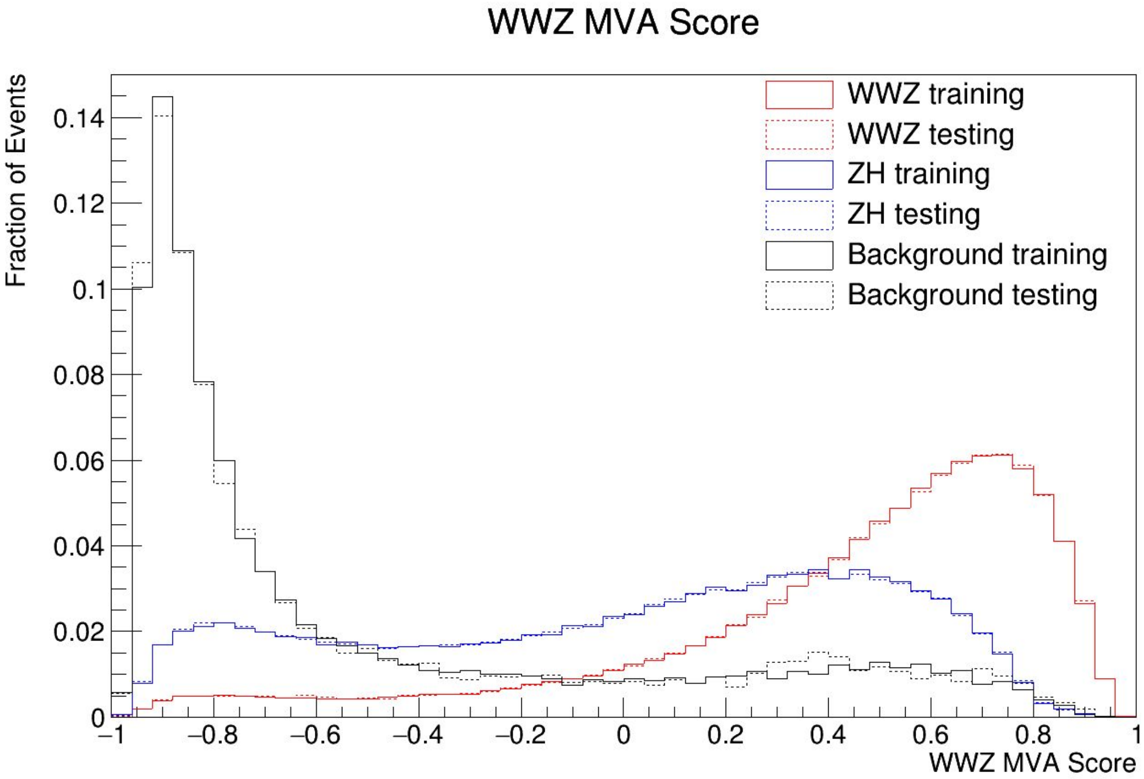


<u>All distributions normalized to 1!</u>

Variable Rankings - Same Flavor

- Good variables
  - MET, m_{T2}, p_T^{4l}
- Minor variables → may want to drop these eventually
  - $\Delta\phi$(WW,MET), $p_T^{leptons}$, STLepHad or STLep
- Bad variables → prone to data/MC mismodelling
  - m_{ll}

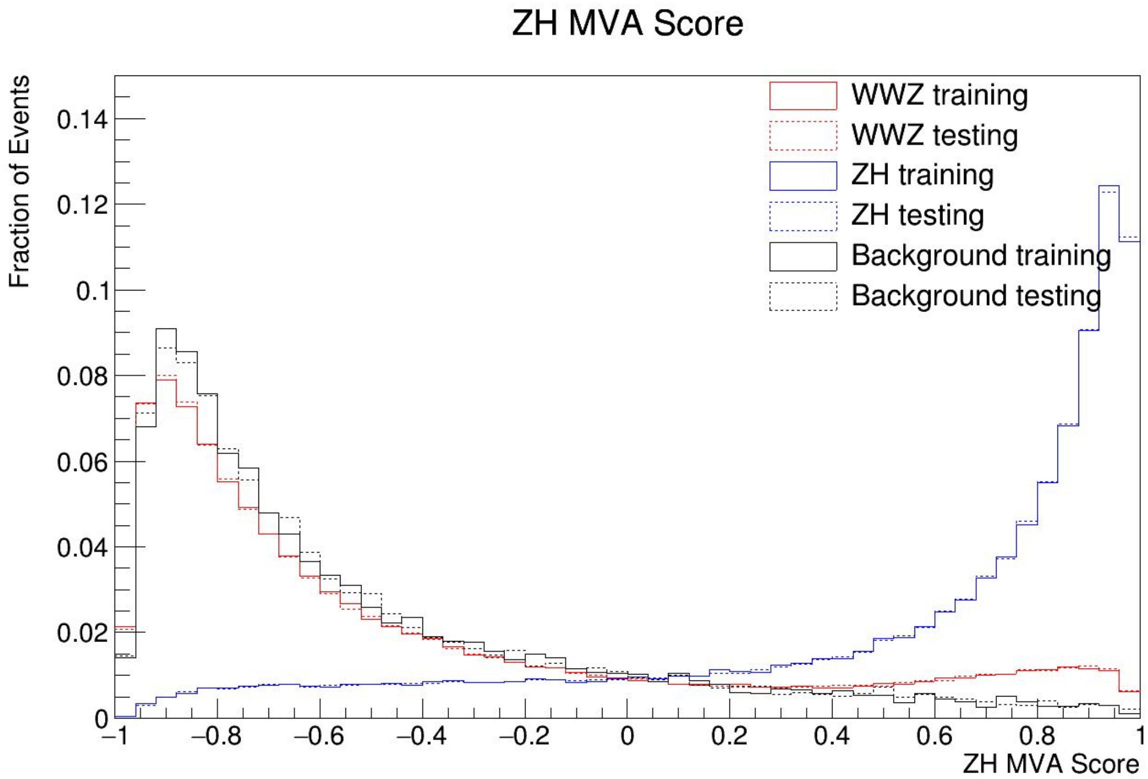
Validation Check - Opposite Flavor



Good shape agreement between training and testing datasets

Note: Background distribution has low stats compared to WWZ and ZH

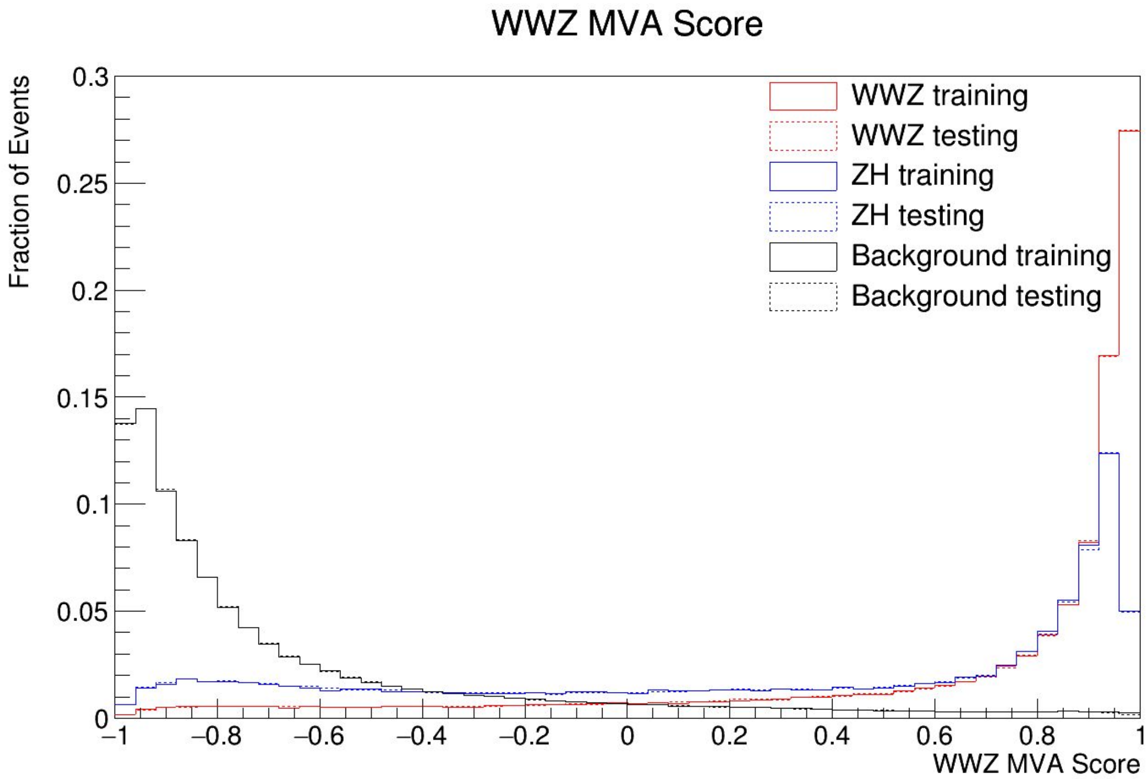
Validation Check - Opposite Flavor (cont.)



Good shape agreement between training and testing datasets

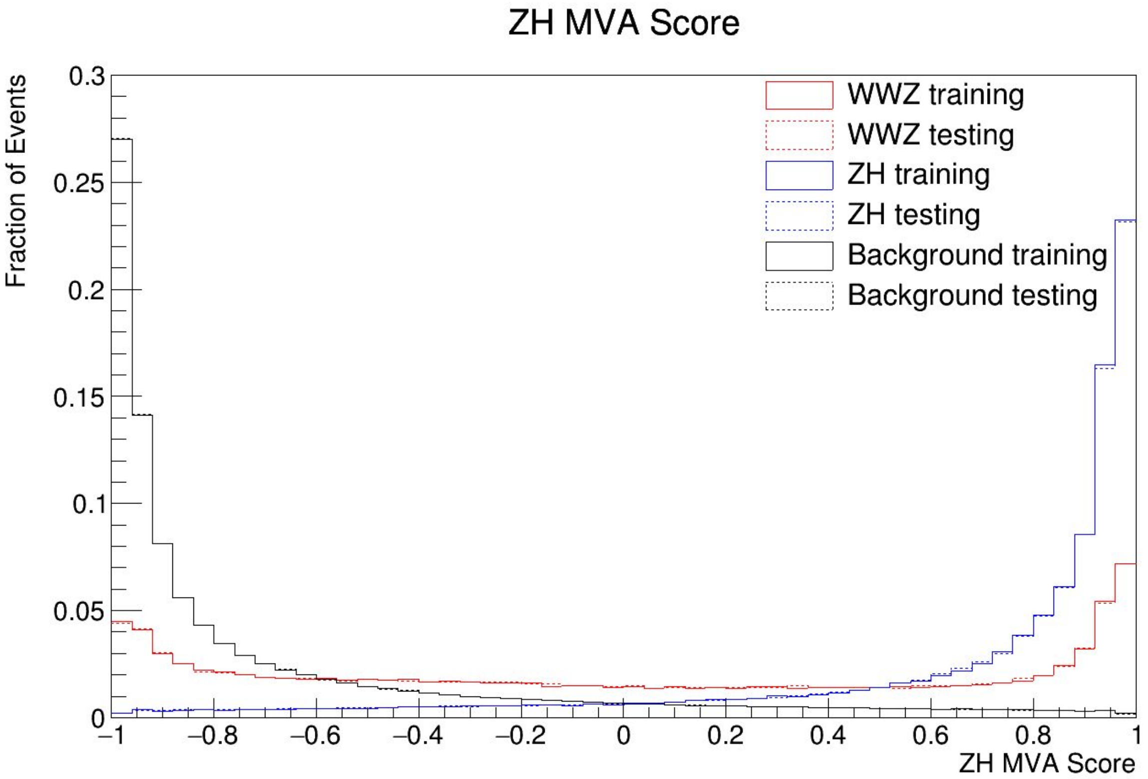
Note: Background distribution has low stats compared to WWZ and ZH

Validation Check - Same Flavor



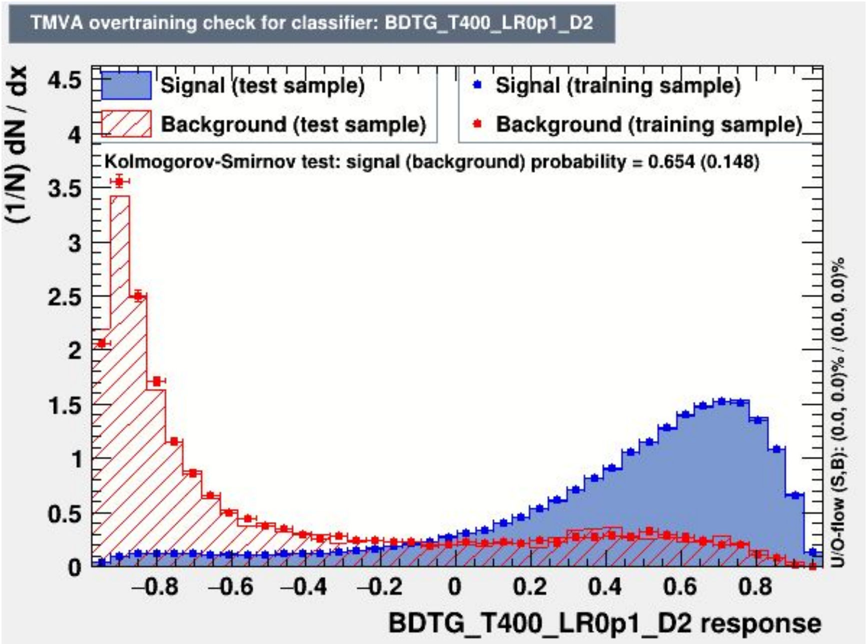
Good shape agreement between training and testing datasets

Validation Check - Same Flavor (cont.)

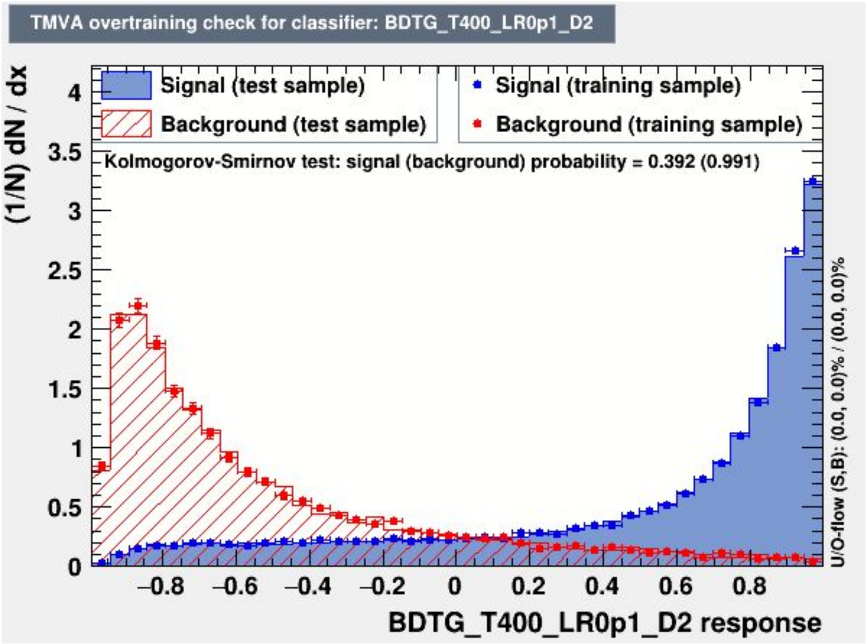


Good shape agreement between training and testing datasets

KS Test - Opposite Flavor



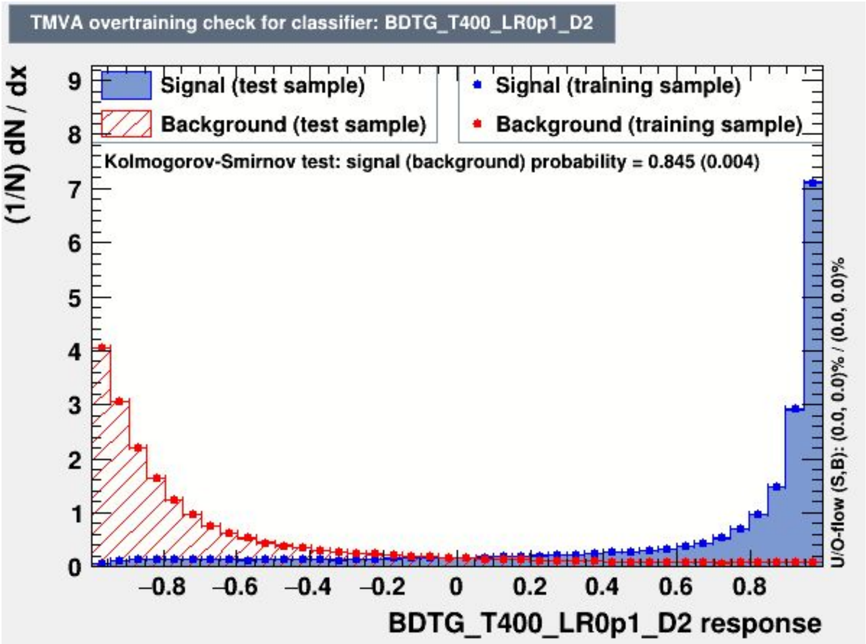
WWZ BDT



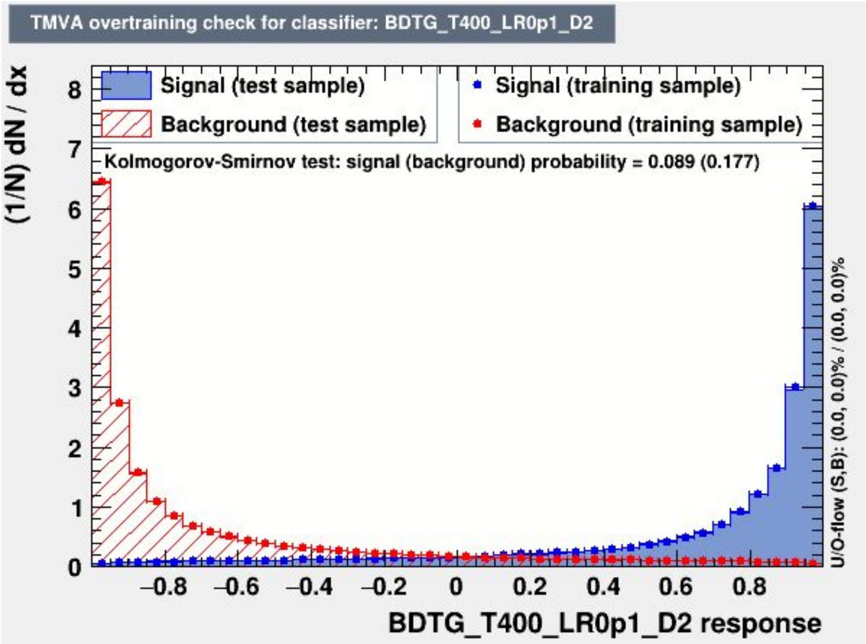
ZH BDT

→Don't seem to have overtraining

KS Test - Same Flavor



WWZ BDT



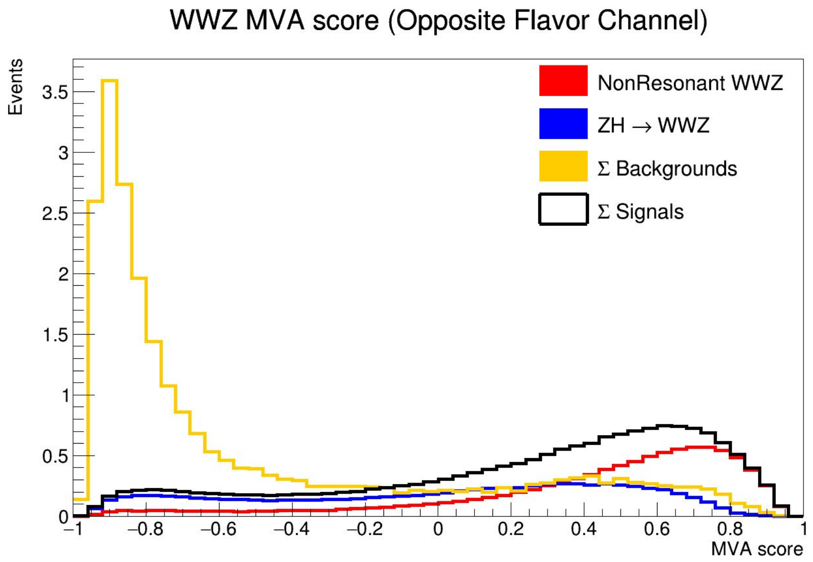
ZH BDT

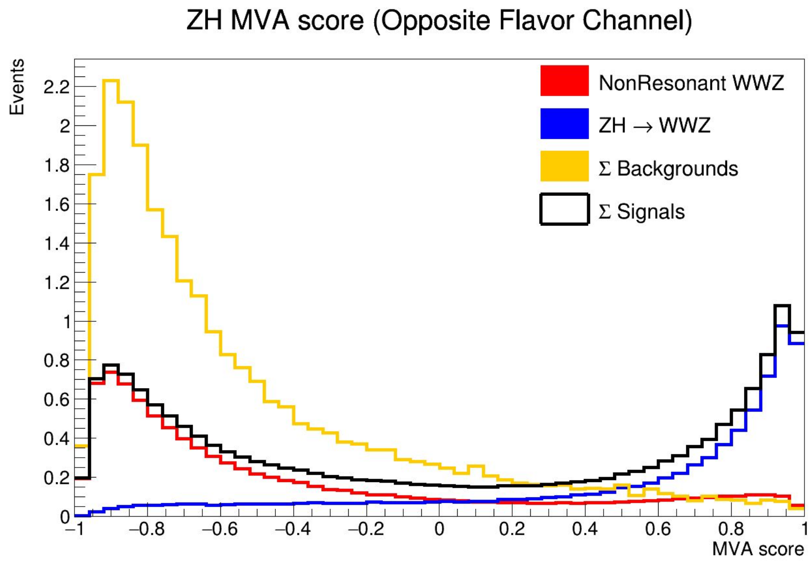
KS test seems to suggest overtraining → Not obvious visually

Merging of Testing+Training Events

- Can we merge testing+training?
  - Without overtraining, we can evaluate the MVA score for **all** events
    - If there is no bias towards the training dataset, then re-using the training data during evaluation will not bias the results
- Since there is no obvious overtraining, I suggest evaluating all (training+testing) events rather than scaling the testing dataset

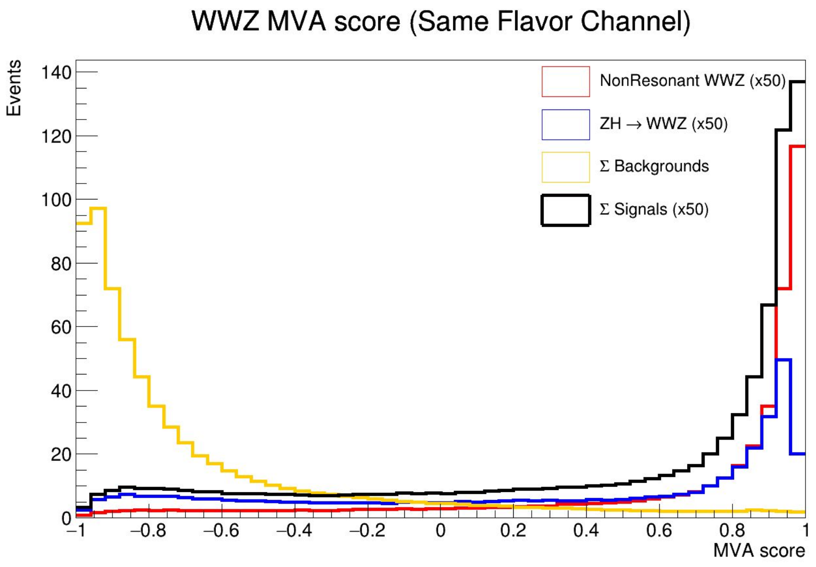
BDT MVA Distributions - Opposite Flavor

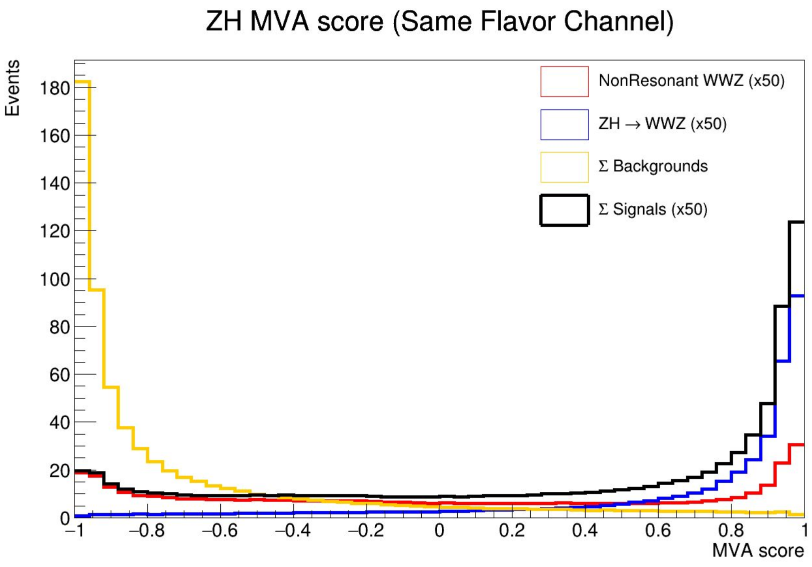




Comparing MVA distributions for signals and backgrounds
→ Looks reasonable

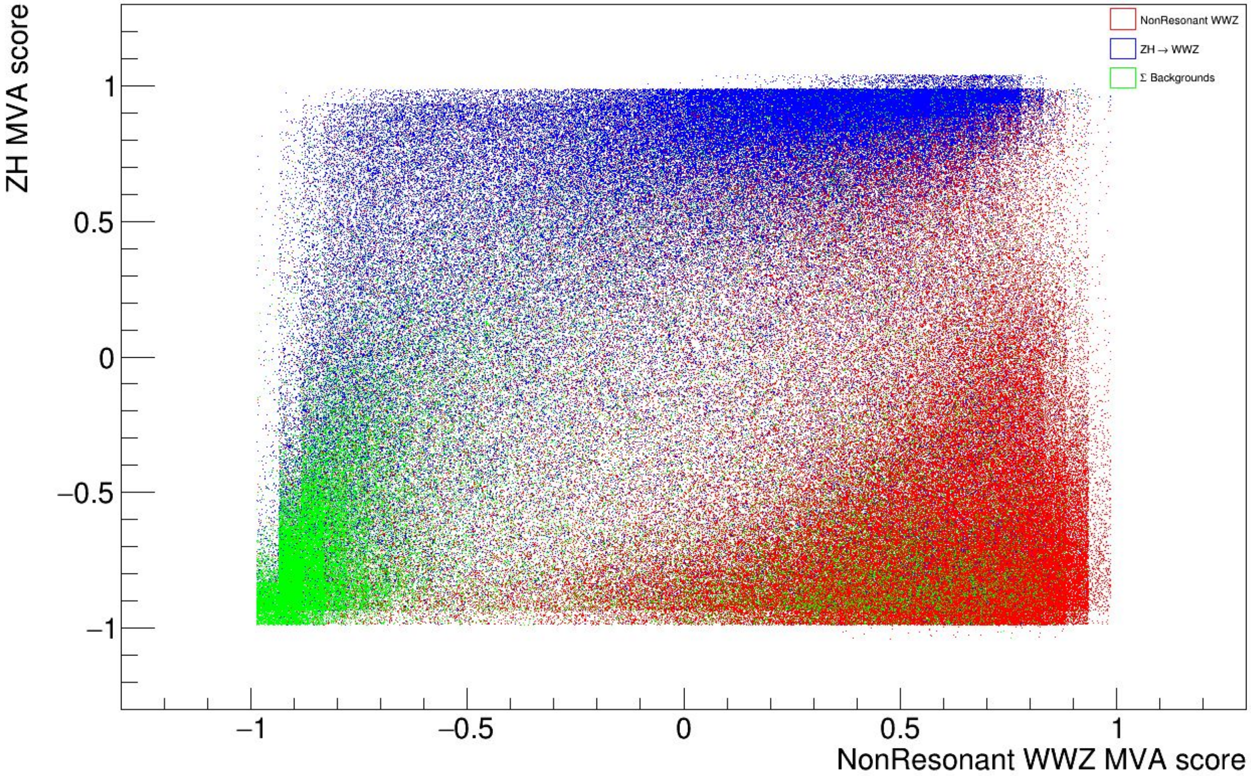
BDT MVA Distributions - Same Flavor





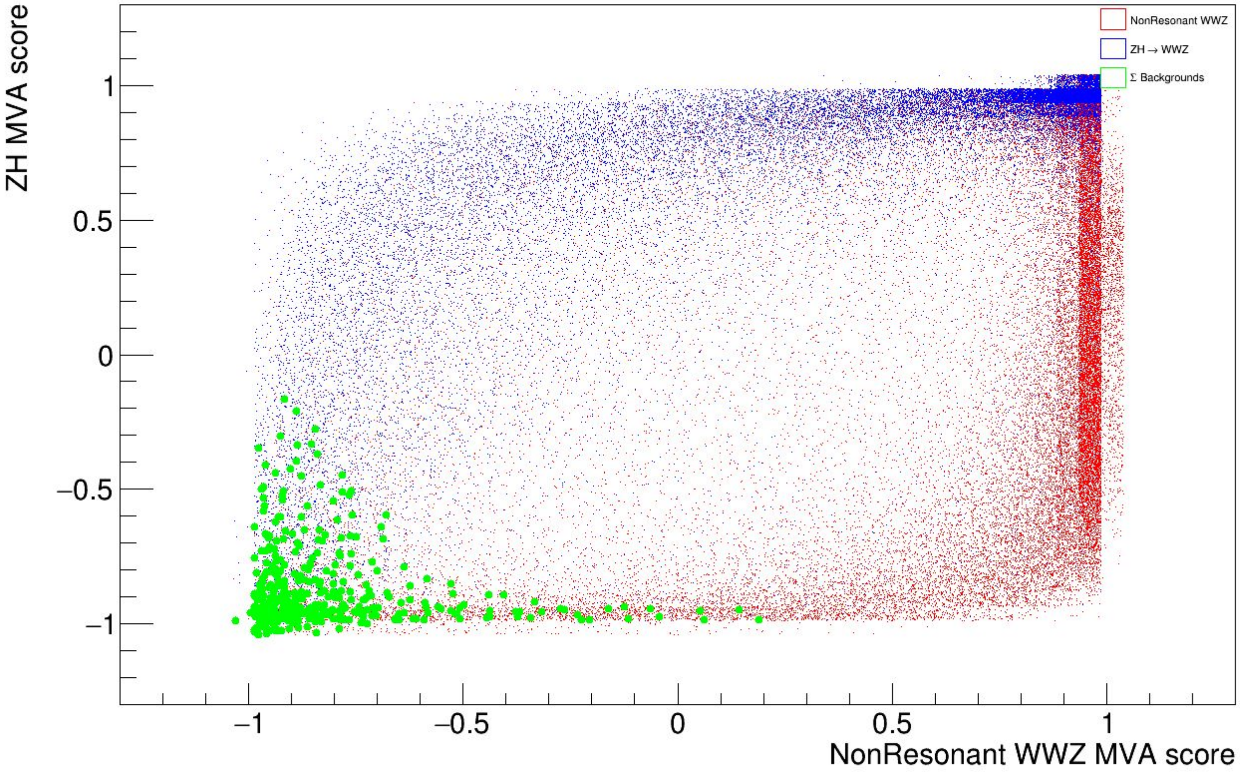
Again, this looks reasonable

BDT MVA outputs - 2D scatter plot (Opposite Flavor)



- Can identify regions that are “ZH-like” and “WWZ-like”
  - Top right: ZH-like
  - Bottom right: WWZ-like

BDT MVA outputs - 2D scatter plot (Same Flavor)



- Can identify 2 regions where signal tends to live
  - ZH: high ZH score and high WWZ score
  - WWZ: High WWZ score and (somewhat) uniform in ZH score

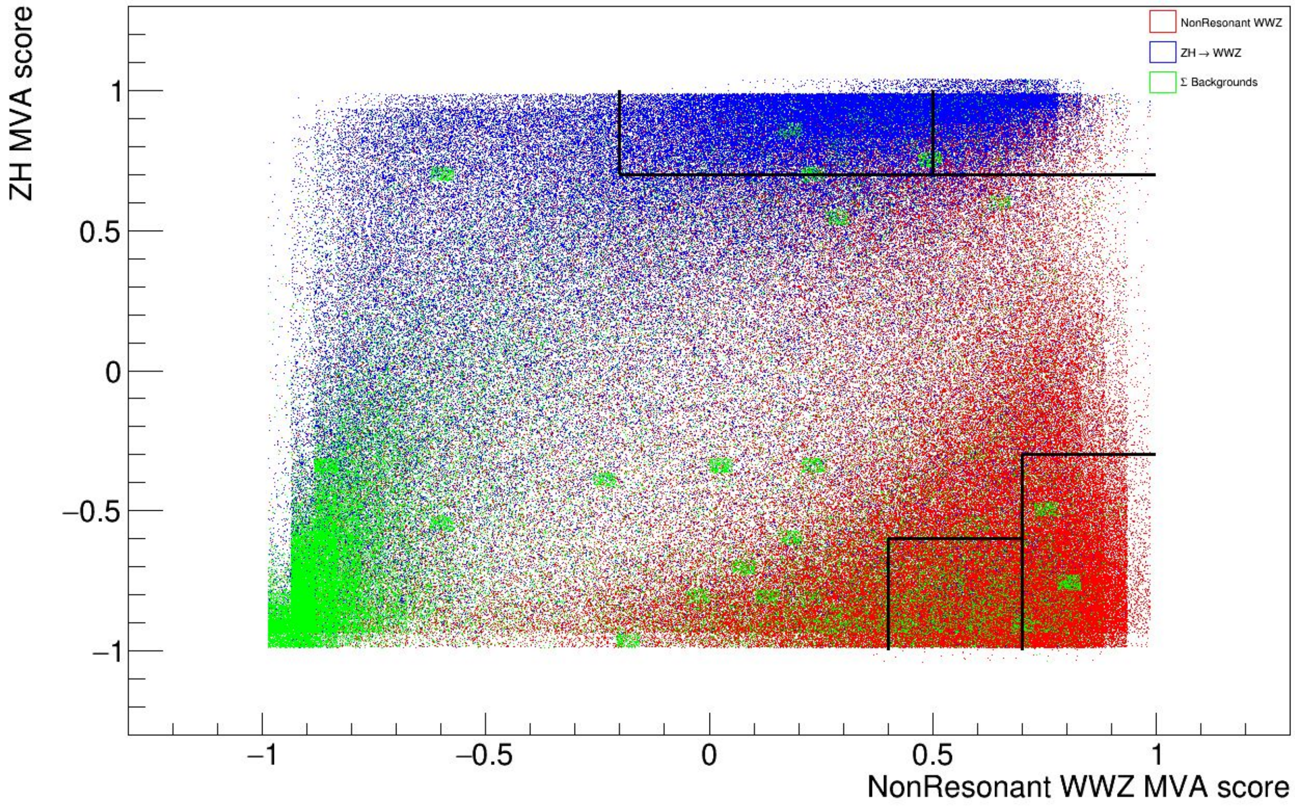
Binning for SRs

- Binning was done “by eye”
  - Idea: define regions that isolate the individual signals
    - Easier said than done → Considerable overlap between signals in 2D plots
- For each channel, define 2 SRs for each signal
  - 2 signals x 2 bins x 2 channels = 8 bins total
  - For each signal, 1 bin is “pure” in the signal while the other tends to be “mixed”

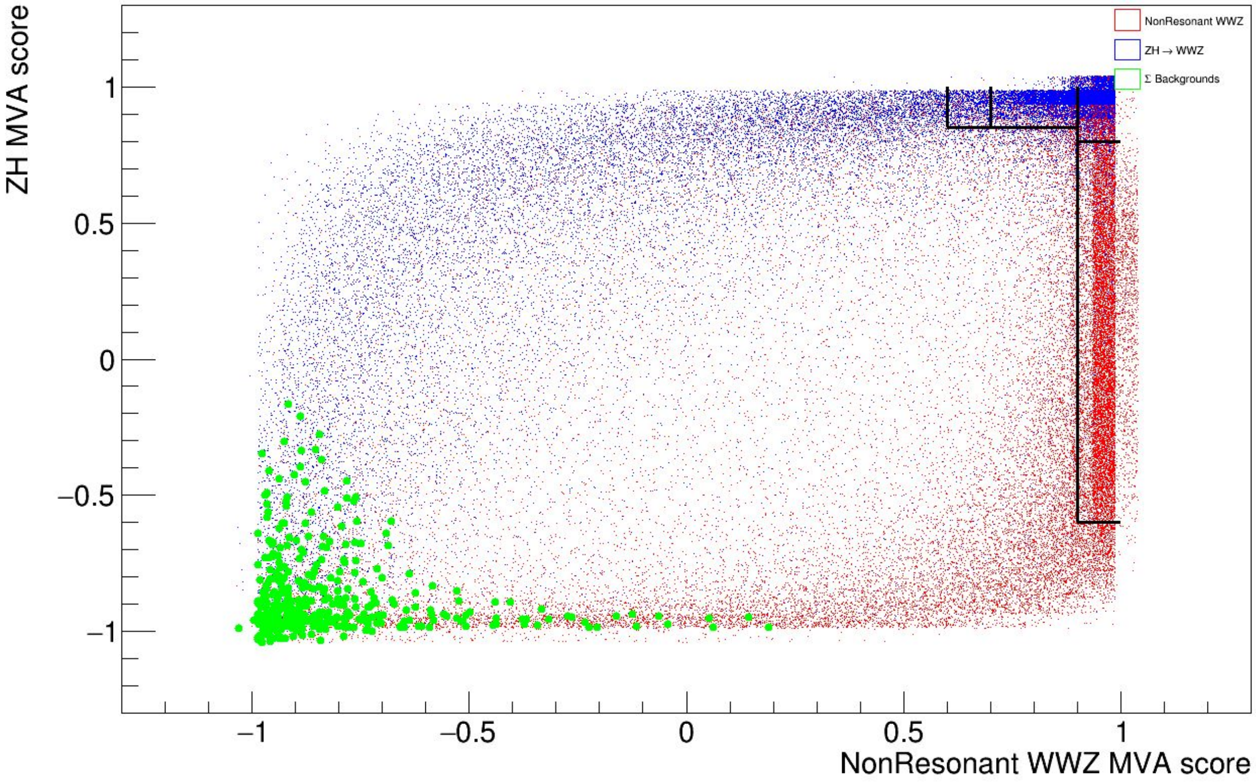
SR	SF SR 1	SF SR 2	SF SR 3	SF SR 4	OF SR 1	OF SR 2	OF SR 3	OF SR 4
WWZ Score	> 0.9	> 0.9	∈ (0.7,0.9)	∈ (0.6,0.7)	> 0.7	∈ (0.4,0.7)	> 0.5	∈ (-0.2,0.5)
ZH Score	> 0.8	∈ (-0.6,0.8)	> 0.85	> 0.85	< -0.3	< -0.6	> 0.7	> 0.7

(Binning shown on 2D plots on next 2 slides)

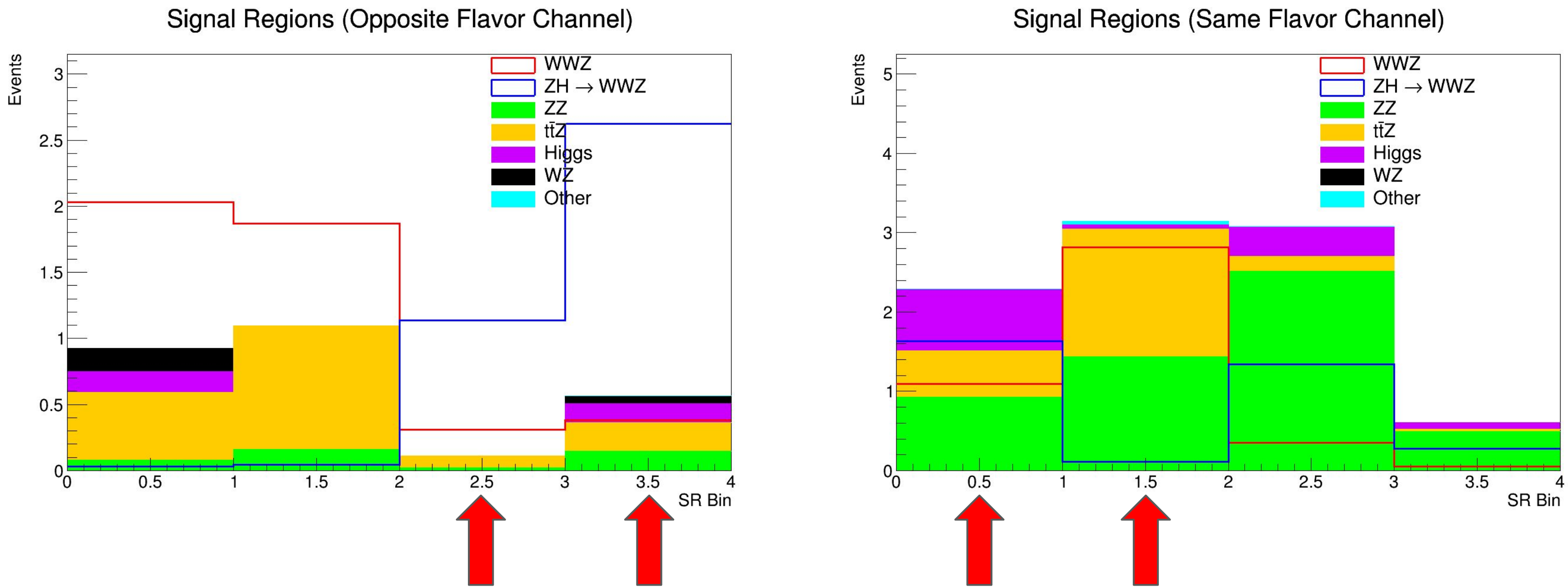
2D Signal Regions - Opposite Flavor



2D Signal Regions - Same Flavor



SR Yields - BDT Analysis



Arrows show 2 most sensitive bins in each channel

SR Yields - BDT Analysis (cont.)

Most sensitive bins (OF channel)

Most sensitive bins (SF channel)

SR Bins	Summary N_{total}	NonResonant WWZ	ZHWWZ	$S/\sqrt{B}$	Composition of N_{total} ZZ	ttZ	Higgs	WZ	Other
OF SR 1	0.93 ± 0.17	2.03 ± 0.01	0.032 ± 0.002	2.15	0.082 ± 0.007	0.51 ± 0.04	0.16 ± 0.10	0.18 ± 0.13	0.0001 ± 0.004
OF SR 2	1.06 ± 0.12	1.87 ± 0.01	0.044 ± 0.002	1.86	0.16 ± 0.01	0.94 ± 0.06	-0.046 ± 0.101	0 ± 0	0.011 ± 0.005
OF SR 3	0.11 ± 0.02	0.312 ± 0.003	1.135 ± 0.007	4.34	0.022 ± 0.004	0.090 ± 0.015	-0.002 ± 0.018	0 ± 0	0.001 ± 0.001
OF SR 4	0.56 ± 0.12	0.380 ± 0.004	2.62 ± 0.01	4.00	0.147 ± 0.009	0.215 ± 0.023	0.145 ± 0.099	0.056 ± 0.056	0.001 ± 0.005
SF SR 1	2.29 ± 0.27	1.09 ± 0.01	1.63 ± 0.01	1.80	0.93 ± 0.02	0.58 ± 0.04	0.78 ± 0.26	0 ± 0	0.009 ± 0.007
SF SR 2	3.15 ± 0.09	2.82 ± 0.01	0.114 ± 0.003	1.65	1.44 ± 0.03	1.61 ± 0.07	0.055 ± 0.037	0 ± 0	0.043 ± 0.037
SF SR 3	3.08 ± 0.26	0.354 ± 0.003	1.34 ± 0.01	0.97	2.52 ± 0.04	0.19 ± 0.02	0.37 ± 0.25	0 ± 0	0.003 ± 0.006
SF SR 4	0.61 ± 0.09	0.049 ± 0.001	0.274 ± 0.004	0.41	0.49 ± 0.02	0.032 ± 0.008	0.079 ± 0.087	0 ± 0	0.000 ± 0.002

Table 23: Yields in BDT signal regions

Most sensitive bins (OF channel)

Most sensitive bins (SF channel)

SR Bins	Summary N_{total}	NonResonant WWZ	ZHWWZ	$S/\sqrt{B}$	Composition of N_{total} ZZ	ttZ	Higgs	WZ	Other
OF Bin 1	1.33 ± 0.21	0.65 ± 0.01	3.05 ± 0.01	3.21	0.62 ± 0.02	0.32 ± 0.03	0.27 ± 0.19	0.11 ± 0.08	0.01 ± 0.01
OF Bin 2	0.96 ± 0.11	0.74 ± 0.01	1.35 ± 0.01	2.13	0.62 ± 0.02	0.38 ± 0.03	-0.06 ± 0.10	0 ± 0	0.01 ± 0.01
OF Bin 3	1.39 ± 0.22	1.48 ± 0.01	0.35 ± 0.01	1.55	0.39 ± 0.02	0.83 ± 0.05	-0.06 ± 0.16	0.18 ± 0.13	0.04 ± 0.04
OF Bin 4	3.60 ± 0.23	5.14 ± 0.01	0.15 ± 0.01	2.79	0.50 ± 0.02	2.51 ± 0.09	0.45 ± 0.17	0.12 ± 0.12	0.02 ± 0.01
SF Bin 1	3.25 ± 0.22	2.33 ± 0.01	0.95 ± 0.01	1.82	1.32 ± 0.03	1.32 ± 0.06	0.61 ± 0.21	0 ± 0	0.01 ± 0.01
SF Bin 2	6.18 ± 0.36	1.97 ± 0.01	1.52 ± 0.01	1.40	4.58 ± 0.05	1.12 ± 0.06	0.42 ± 0.35	0 ± 0	0.06 ± 0.04
SF Bin 3	3.15 ± 0.18	0.57 ± 0.01	0.68 ± 0.01	0.70	2.78 ± 0.04	0.27 ± 0.03	0.10 ± 0.17	0 ± 0	0.01 ± 0.01

Table 25: Yields in Cut-Based SR bins (as of September 1st, 2023)

Based on S/sqrt(B) considerations (purple), MVA bins are more sensitive!

Combine Result - MVA vs Cut-Based

- Take the MC-based yields from the previous slide
  - Use these to calculate the significance (Z) for the MVA and cut-based analyses

$$Z_{MVA} = 4.81\ \sigma$$

$$Z_{cut\text{-}based} = 4.56\ \sigma$$

→ **MVA outperforms cut-based analysis** (as expected)

This was achieved without “optimal” binning

BDT hyperparameters were also not “fully-optimized”

Thus…. **there is room for improvement on this result**

Improvements for the BDT

- Replace some “bad” training variables with variables that provide better discrimination
  - Also, figure out what these “better” variables are
- Tweak the BDT hyperparameters to find a combination that strikes a balance between (lack of) overtraining and discrimination power
- Figure out how to improve the 2D binning for the MVA SRs

To do

- Samples to add:
  - VVV (WWW, WZZ, ZZZ)
  - tWZ
- Reduce skim size
  - What progress has been made?
- Implement lepton SFs
- Background Estimation
  - ZZ, ttZ, WZ → 3 lepton + 1 fake

Backup

BDT Hyperparameters (TMVA)

- N_{trees} = 400
  - Number of trees in the forest
- MinNodeSize = 5%
  - Minimum fraction of the training data used to construct individual nodes
- Boost Type: Gradient Boost
  - Algorithm used for boosting
- Shrinkage = 0.1
  - Learning rate for the Gradient Boost Algorithm
- MaxDepth = 2
  - Maximum depth of decision tree
- SeparationType = SDivSqrtSPlusB
  - Node splitting is done by computing s/sqrt(s+b) and comparing to the nominal value for "signal" or "background"
- Ignore Negative Weights in Training