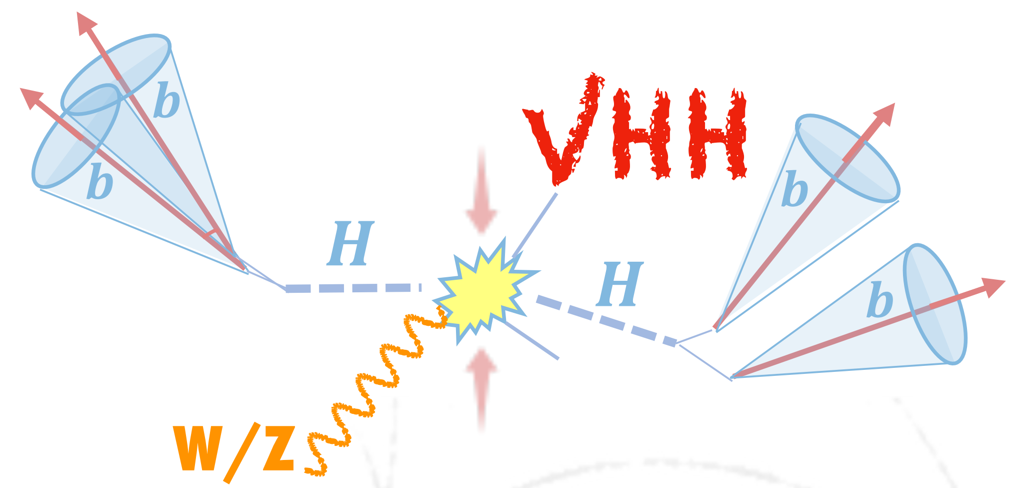
Search for Higgs boson pair production in $b\bar{b}b\bar{b}$ final state in association with a vector boson with the CMS detector

HIG-22-006



John Alison[2], Chayanit Asawatangtrakuldee[1], Agni Bethani[3], Yihui Lai[3], Chuyuan Liu[2], Chris Palmer[3], Xiaohu Sun[4], Long Wang[3], Licheng Zhang[4]

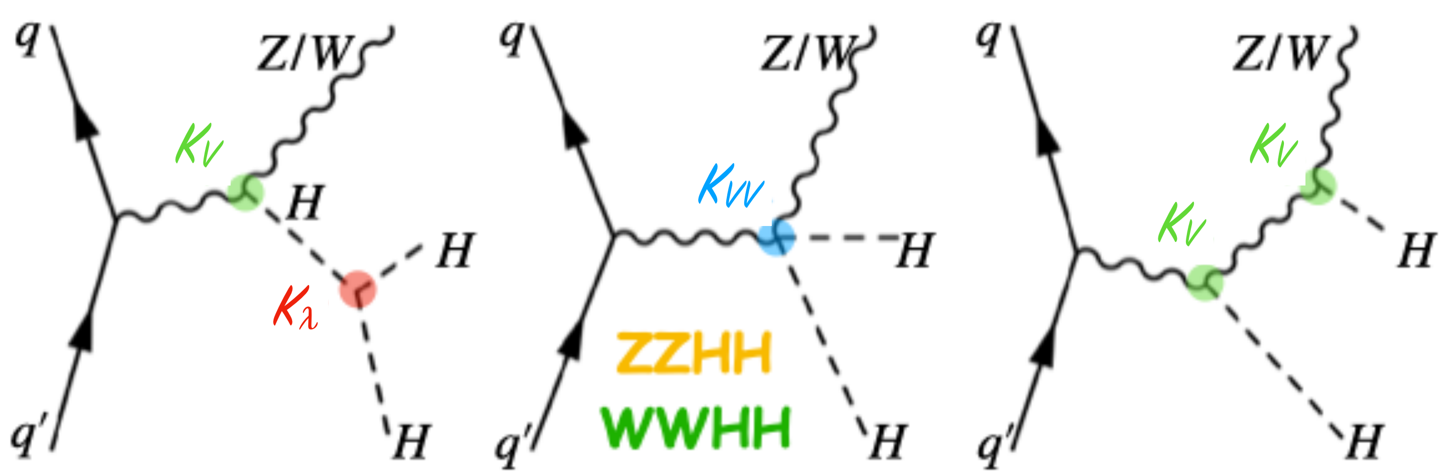
1. Chulalongkorn University
2. Carnegie-Mellon University
3. University of Maryland
4. Peking University

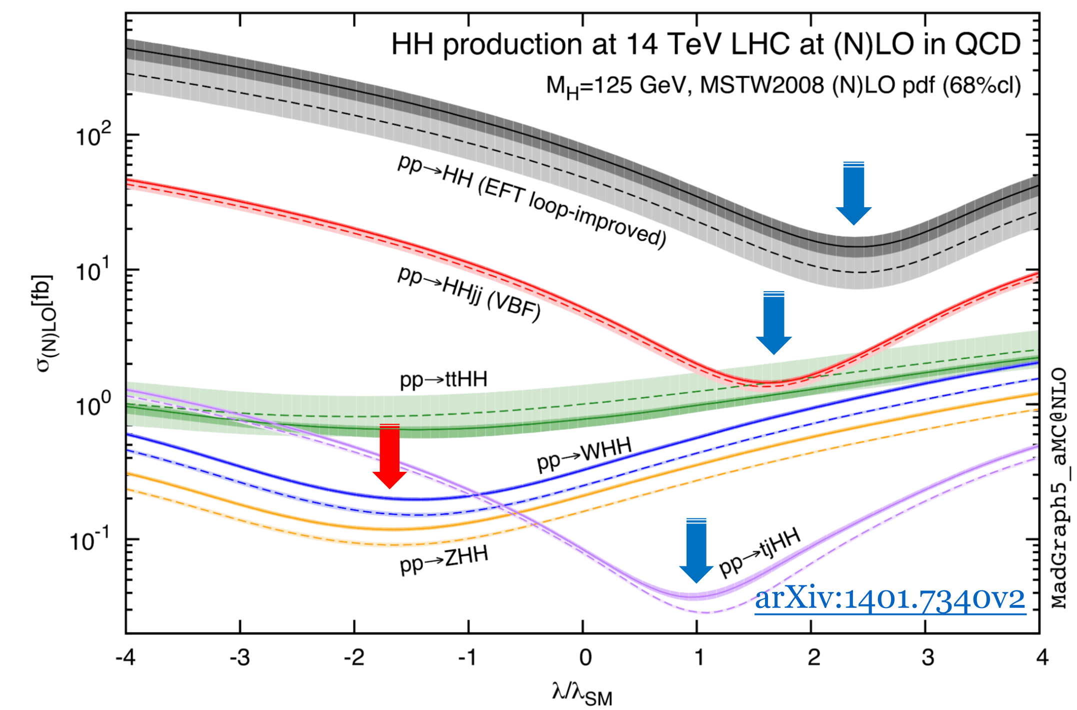
23rd-27th Nov. 2022
CLHCP, NanJing

On behalf of PKU CMS group and VHH4b analysis team

HH : direct investigation of EWSB and scalar sector properties

- We are studying HH production mode associated with one vector boson (VHH)
    - Focus on HH decay to 4b final states and leptonic decay and hadronic decay for V-bosons.
    - Complementary to ggF and VBF analyses.





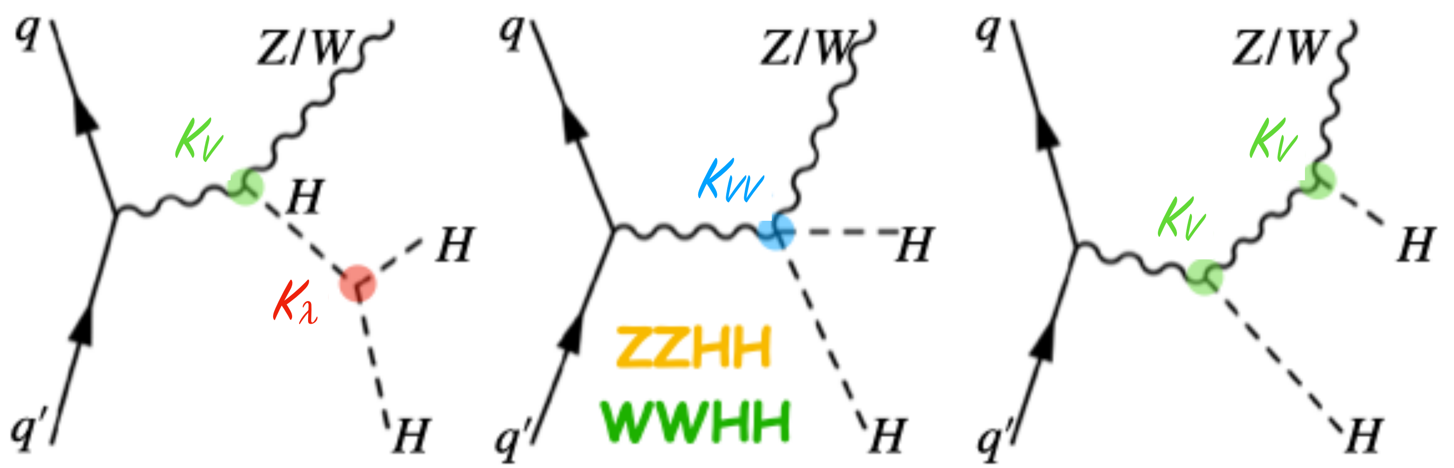
arXiv:1401.7340v2	$\sqrt{s}=8$ TeV (LO) NLO		$\sqrt{s}=13$ TeV (LO) NLO		$\sqrt{s}=14$ TeV (LO) NLO	
HH (EFT loop-improv.)	$(5.44^{+38\%}_{-26\%})$	$8.73^{+17+2.9\%}_{-16-3.7\%}$	$(19.1^{+33\%}_{-23\%})$	$29.3^{+15+2.1\%}_{-14-2.5\%}$	$(22.8^{+32\%}_{-23\%})$	$34.8^{+15+2.0\%}_{-14-2.5\%}$
$HHjj$ (VBF)	$(0.436^{+12\%}_{-10\%})$	$0.479^{+1.8+2.8\%}_{-1.8-2.0\%}$	$(1.543^{+9.4\%}_{-8.0\%})$	$1.684^{+1.4+2.6\%}_{-0.9-1.9\%}$	$(1.839^{+8.9\%}_{-7.7\%})$	$2.017^{+1.3+2.5\%}_{-1.0-1.9\%}$
$t\bar{t}HH$	$(0.265^{+41\%}_{-27\%})$	$0.177^{+4.7+3.2\%}_{-19-3.3\%}$	$(1.027^{+37\%}_{-25\%})$	$0.792^{+2.8+2.4\%}_{-10-2.9\%}$	$(1.245^{+36\%}_{-25\%})$	$0.981^{+2.3+2.3\%}_{-9.0-2.8\%}$
W^+HH	$(0.111^{+4.0\%}_{-3.9\%})$	$0.145^{+2.1+2.5\%}_{-1.9-1.9\%}$	$(0.252^{+1.4\%}_{-1.7\%})$	$0.326^{+1.7+2.1\%}_{-1.2-1.6\%}$	$(0.283^{+1.1\%}_{-1.3\%})$	$0.364^{+1.7+2.1\%}_{-1.1-1.6\%}$
W^-HH	$(0.051^{+4.2\%}_{-4.0\%})$	$0.069^{+2.1+2.6\%}_{-1.9-2.2\%}$	$(0.133^{+1.5\%}_{-1.7\%})$	$0.176^{+1.6+2.2\%}_{-1.2-2.0\%}$	$(0.152^{+1.1\%}_{-1.4\%})$	$0.201^{+1.7+2.2\%}_{-1.1-1.8\%}$
ZHH	$(0.098^{+4.2\%}_{-4.0\%})$	$0.130^{+2.1+2.2\%}_{-1.9-1.9\%}$	$(0.240^{+1.4\%}_{-1.7\%})$	$0.315^{+1.7+2.0\%}_{-1.1-1.6\%}$	$(0.273^{+1.1\%}_{-1.3\%})$	$0.356^{+1.7+1.9\%}_{-1.2-1.5\%}$
$tjHH(\cdot 10^{-3})$	$(5.057^{+2.0\%}_{-3.2\%})$	$5.606^{+4.4+3.9\%}_{-2.3-4.2\%}$	$(23.20^{+0.0\%}_{-0.8\%})$	$29.77^{+4.8+2.8\%}_{-2.8-3.2\%}$	$(28.79^{+0.0\%}_{-1.2\%})$	$37.27^{+4.7+2.6\%}_{-2.7-3.0\%}$

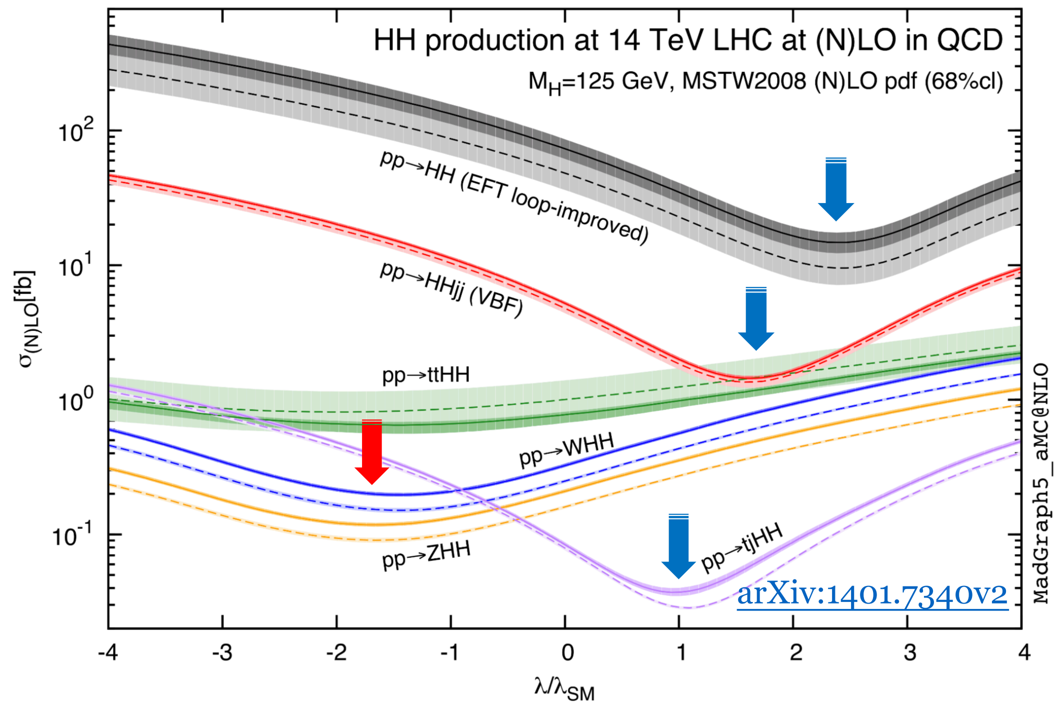
ggF
VBF
VHH

- Cross sections are enhanced by the constructive interference.

VHH

- Contribution on sensitivities over K_λ at positive side is expected.





- VHH channel has the unique feature to decompose ZZHH/WWHH(K_{ZZ}/K_{WW}) vertices according to the V-leptonic decay.

- Four orthogonal search channels depending on lepton multiplicity: MET, SL, DL, FH.*

* MET, Single-Lepton, Double-Lepton, Full-Hadronic

Documentation

★ CADI: HIG-22-006

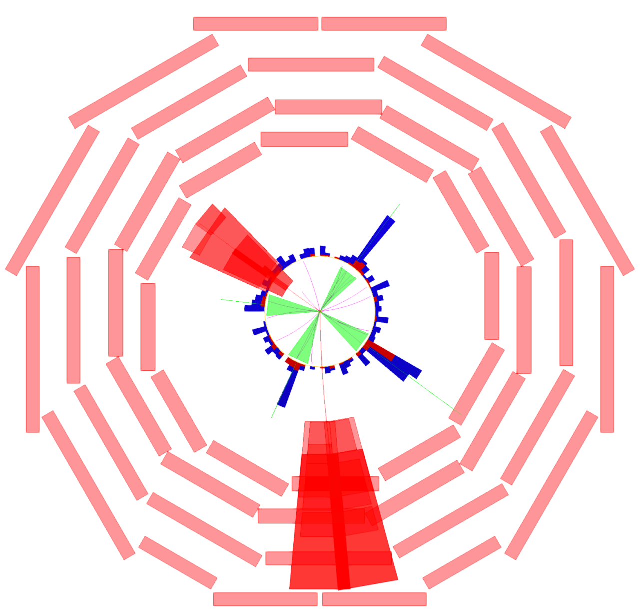
- ★ CMS Pub Talk
- ★ Pre-Approval Talk

★ Analysis Twiki

- ★ Main Page
- ★ Review Twiki Page
  comments from Hbb, HH and HIG have been addressed

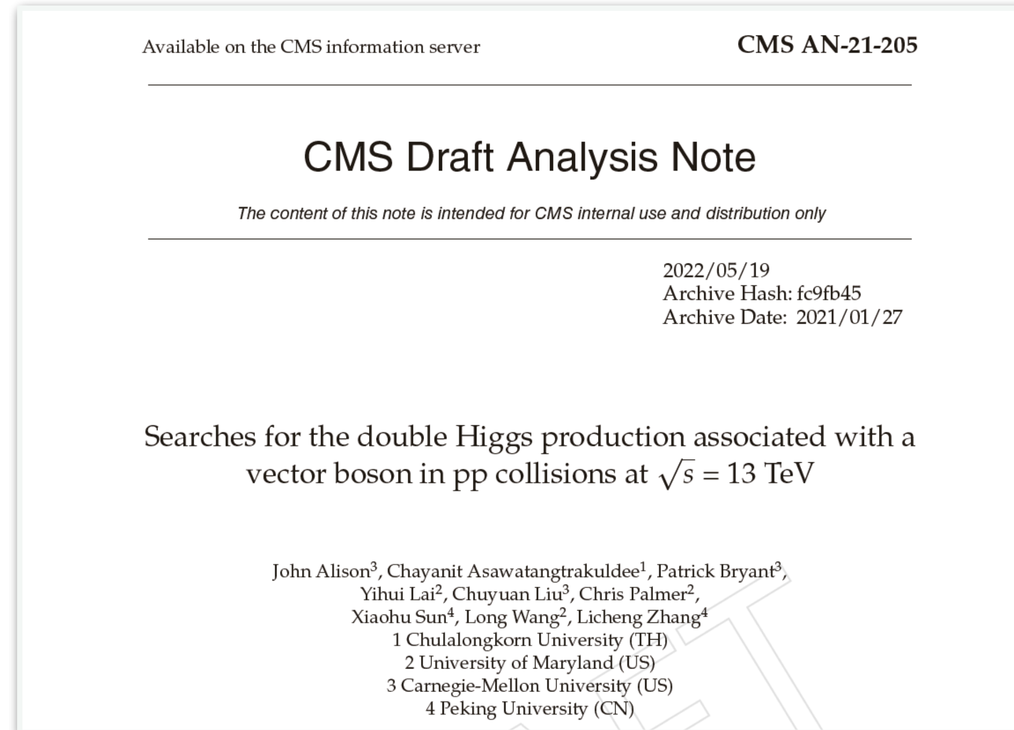
★ ARC:

- ★ James David Olsen (PRINCETON)
- ★ Giacomo Ortona (TORINO)
- ★ Michele Selvaggi (CERN)
- ★ Toni Sculac (SPLIT-UNIV)



Simulated VHH -> μμbbbb event

CMS AN-2021/205



Available on the CMS information server — CMS AN-21-205

CMS Draft Analysis Note

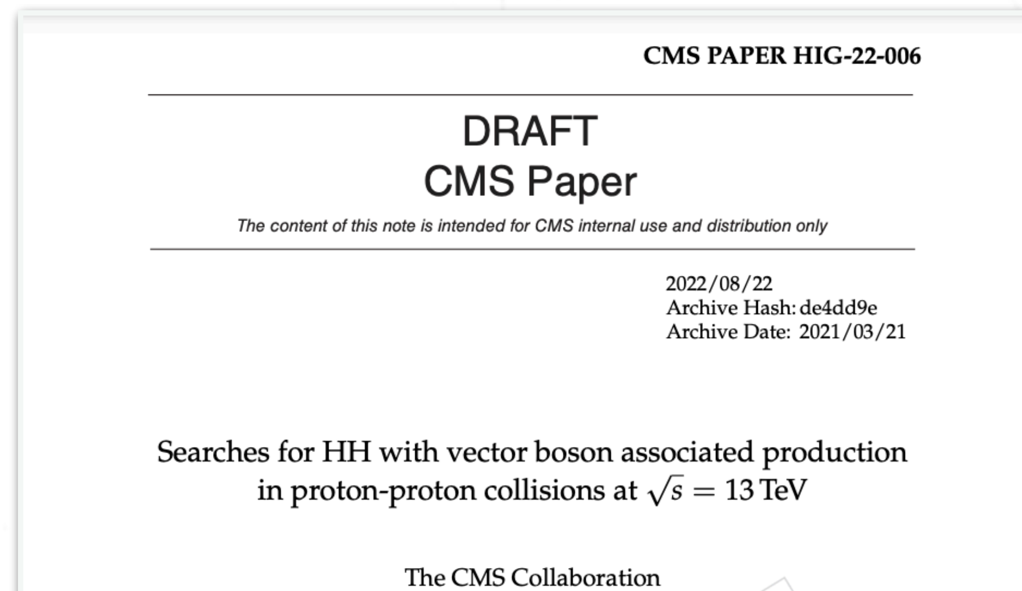
The content of this note is intended for CMS internal use and distribution only

2022/05/19
Archive Hash: fc9fb45
Archive Date: 2021/01/27

Searches for the double Higgs production associated with a vector boson in pp collisions at $\sqrt{s} = 13$ TeV

John Alison[3], Chayanit Asawatangtrakuldee[1], Patrick Bryant[3], Yihui Lai[2], Chuyuan Liu[3], Chris Palmer[2], Xiaohu Sun[4], Long Wang[2], Licheng Zhang[4]
1 Chulalongkorn University (TH)
2 University of Maryland (US)
3 Carnegie-Mellon University (US)
4 Peking University (CN)

Paper Draft



CMS PAPER HIG-22-006

DRAFT
CMS Paper

The content of this note is intended for CMS internal use and distribution only

2022/08/22
Archive Hash: de4dd9e
Archive Date: 2021/03/21

Searches for HH with vector boson associated production in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration

★ Data

- ★ SL/DL channel: SingleMuon/SingleElectron/DoubleMuon/DoubleElectron(2016/2017)
- ★ SL/DL channel: SingleMuon/DoubleMuon/EGamma(2018)
- ★ MET channel: MET
- ★ FH channel: BTagCSV(2016/2017), JetHT(2018)
- ★ Run: 2016(B,C,D,E,F,G,H); 2017(B,C,D,E,F); 2018(A,B,C,D);

Table 1: Golden JSON files used in the analysis and the corresponding integrated luminosities.

Year	File name	$\mathcal{L}$ (fb^{-1})
2016	Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt	36.3 fb^{-1}
2017	Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON.txt	41.5 fb^{-1}
2018	Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt	59.8 fb^{-1}

★ MC simulation

- ★ All the MC samples updated to UL
- ★ 8 VHH(ZHH/WHH) signal samples at LO (for VHH modeling)
- ★ DY+Jets, ttbar, ttbb, ttH, ttV, single-Top, Z(vv)+Jets samples at LO (for background modeling)

MC Corrections Applied

- Jet energy smearing according to JME prescriptions.
- B-tag efficiencies scale factors (WP scale factors) according to BTV prescriptions. (DeepJets and ParticleNets)
- Trigger efficiency scale factors as measured by VHH group analyzers.
- PileUP re-weighting.

★All corrections have the corresponding uncertainty associated.

★Other corrections will be introduced in specific channels.

All 3 years triggers have been studied

DL

Lepton	HLT Path
Trigger for Muon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8 (2018) HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8 OR HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_Mass8 (2017) HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL OR HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL OR HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ OR HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ (2016)
Trigger for Electron	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL (2018) HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL (2017) HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ (2016)

SL

Trigger	HLT_Ele27_WPTight_Gsf or HLT_IsoMu24 or HLT_IsoTkMu24 (2016) HLT_Ele32_WPTight_Gsf or HLT_IsoMu27 (2017) HLT_Ele32_WPTight_Gsf or HLT_IsoMu24 (2018)

MET

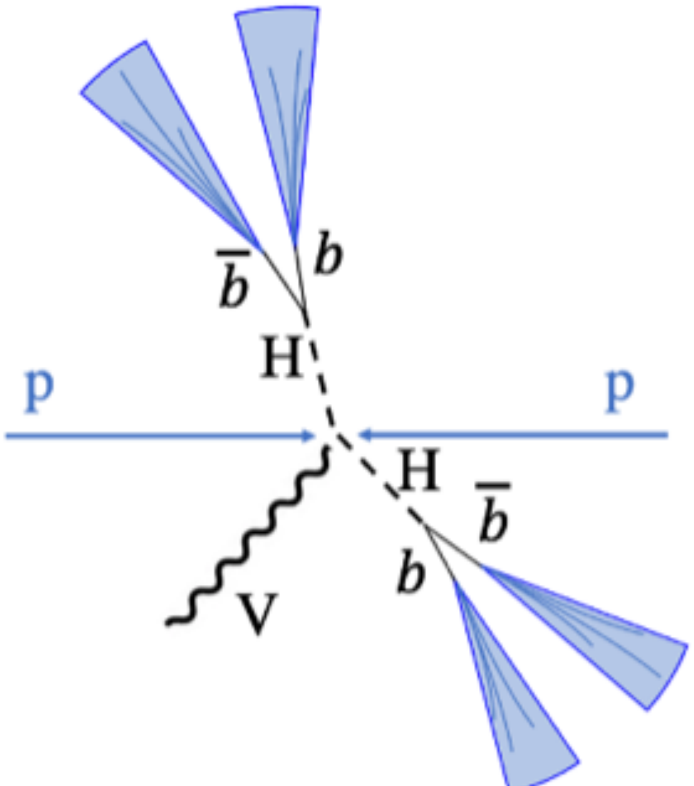
Trigger	HLT_PFMET120_PFMHT120_IDTight (2018)
	HLT_PFMET120_PFMHT120_IDTight OR HLT_PFMET120_PFMHT120_IDTight_PFHT60 (2017)
	HLT_PFMET110_PFMHT110_IDTight OR HLT_PFMET120_PFMHT120_IDTight OR HLT_PFMET170_NoiseCleaned OR HLT_PFMET170_BeamHaloCleaned OR HLT_PFMET170_HBHECleaned (2016)

FH

2016	HLT_QuadJet45_TripleBTagCSV_p087 HLT_DoubleJet90_Double30_TripleBTagCSV_p087 HLT_DoubleJetsC100_DoubleBTagCSV_p014_DoublePFJetsC100MaxDeta1p6
2017	HLT_PFHT300PT30_QuadPFJet_75_60_45_40_TriplePFBTagCSV_3p0 HLT_DoublePFJets100MaxDeta1p6_DoubleCaloBTagCSV_p33
2018	HLT_PFHT330PT30_QuadPFJet_75_60_45_40_TriplePFBTagDeepCSV_4p5 HLT_DoublePFJets116MaxDeta1p6_DoubleCaloBTagDeepCSV_p71

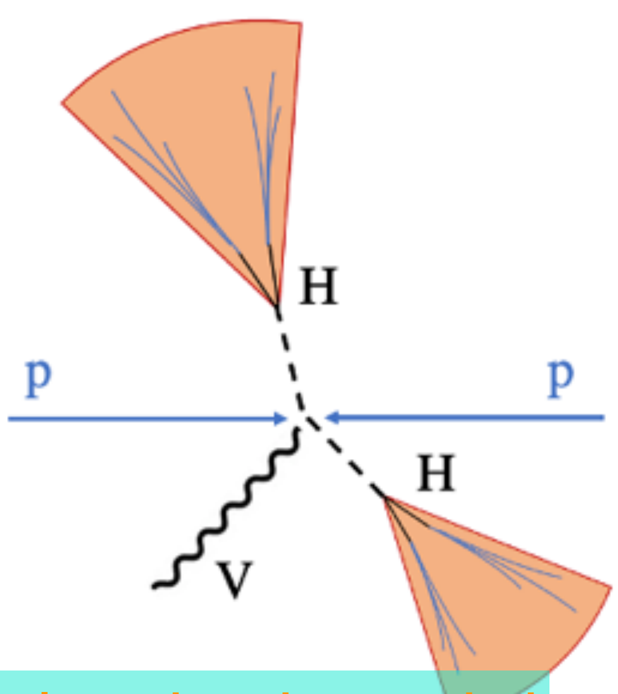
* Json files stored in: /eos/cms/store/group/phys_higgs/hbb/ntuples/VHH4b_Vleptonic_SF_UL

Analysis Topologies



Resolved Topology

Boosted Topology



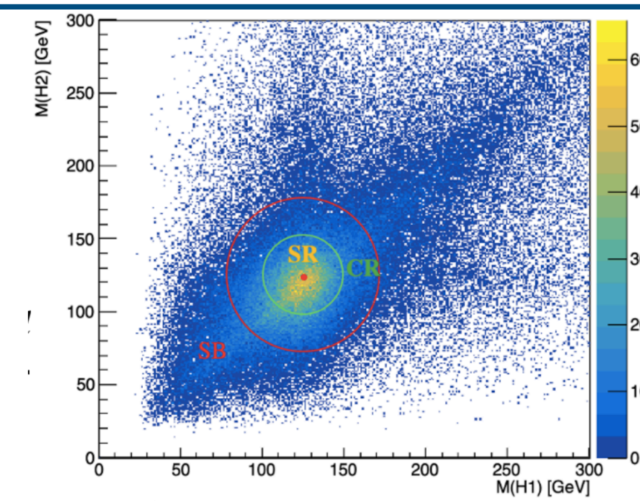
Both topologies have been studied

Leptonic Channels

$$R_{HH} = \sqrt{(M(H_1) - 125)^2 + (M(H_2) - 125)^2}$$

- Signal Region(**SR**) : R_{HH} < 25GeV
- Control Region(**CR**) : 25GeV< R_{HH} <50GeV
- SideBand(**SB**) : 50GeV < R_{HH} < 75GeV

① 2b-tagged events are re-weighted to mimic the background in 3/4b-tagged regions.(DL)

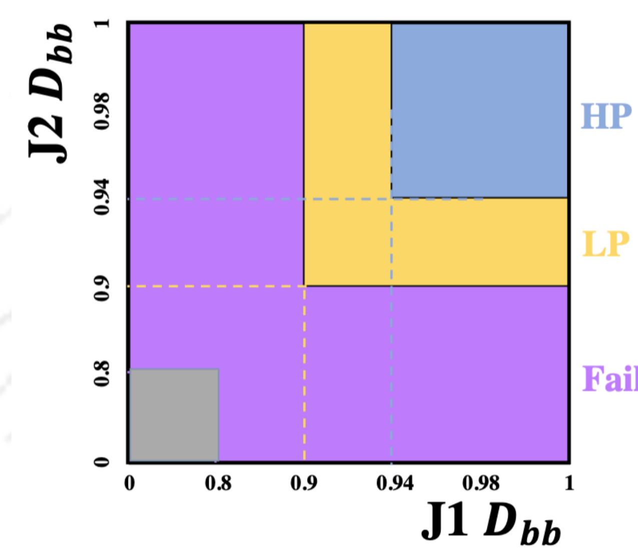


Hadronic Channels

- Signal Region $\sqrt{(\frac{m_1 - (125.0 \times 1.02)\text{GeV}}{0.1 m_1})^2 + (\frac{m_2 - (125.0 \times 0.98)\text{GeV}}{0.1 m_2})^2} < 1.9$
- Control Region $\sqrt{(m_1 - (125.0 \times 1.02)\text{GeV})^2 + (m_2 - (125.0 \times 0.98)\text{GeV})^2} < 30\text{GeV}$
- Sideband $\sqrt{(m_1 - (125.0 \times 1.02)\text{GeV})^2 + (m_2 - (125.0 \times 0.98)\text{GeV})^2} < 45\text{GeV}$

MET+SL

- High Purity (**HP**) : Dbb > 0.94
- Low Purity(**LP**) : 0.90<Dbb<0.94
- Validation Region(**Fail**) : 0.80<Dbb<0.90
- Follow the thresholds as in the VBFHH4b AN.
- Orthogonal variable R_{HH} is also used to define SR[0,50]GeV and SB[50,75]GeV



① HP+LP & SR are the research regions

② HP+LP & SR are the control regions

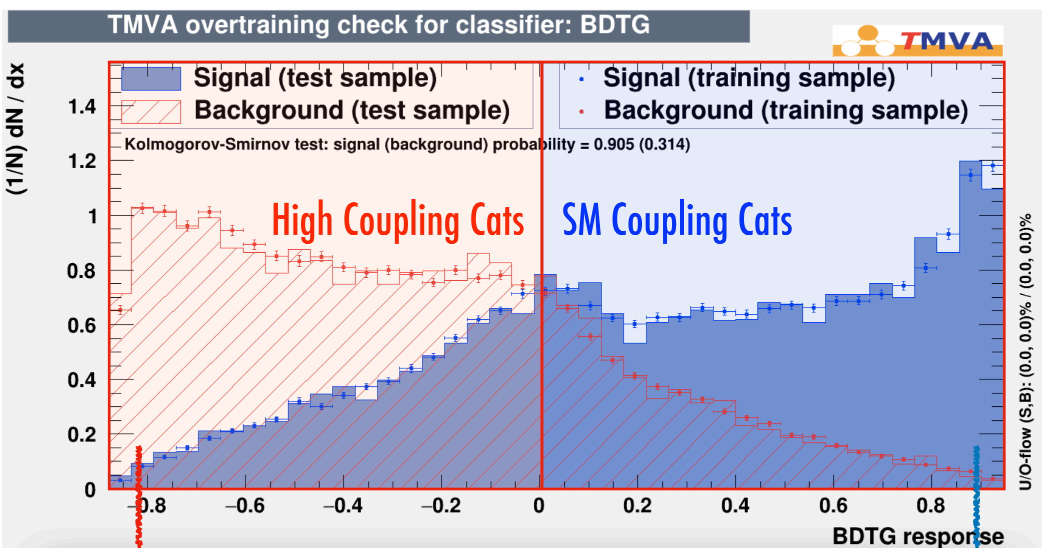
③ FR MC are re-weighted to mimic the background in research regions

ParticleNet for AK8 Jets

$$D_{bb} = \frac{ParticleNetMD\_score(X \to b\bar{b})}{ParticleNetMD\_score(X \to b\bar{b}) + ParticleNetMD\_score(QCD)}$$

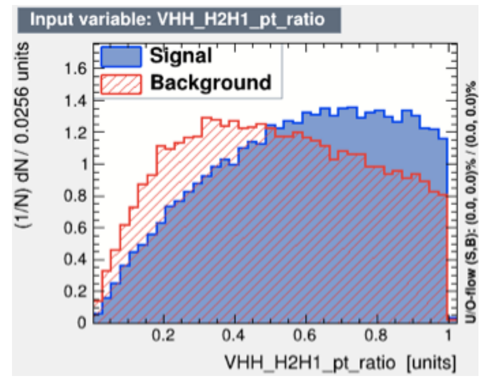
HLT/Object/Event Selections

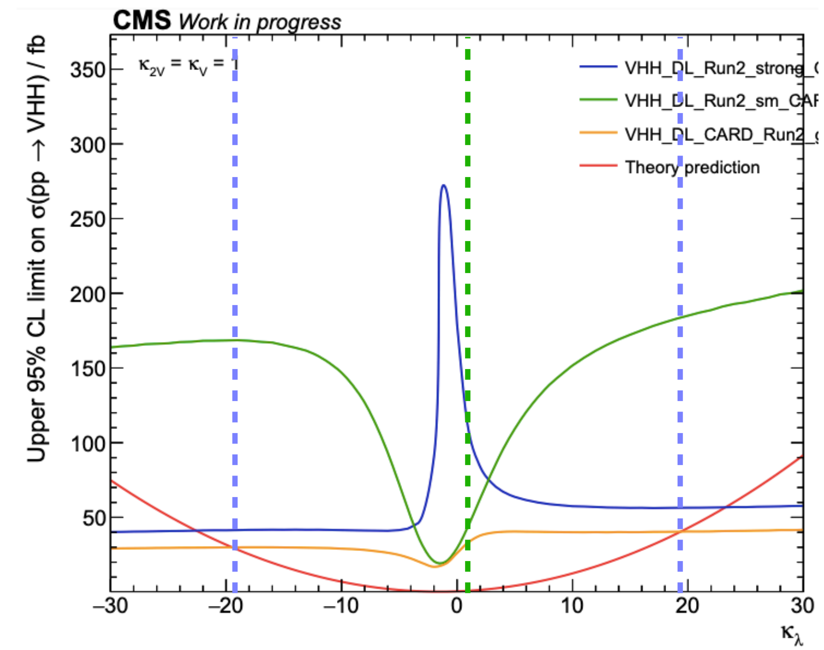
Kl Categorization



Kl Categorization

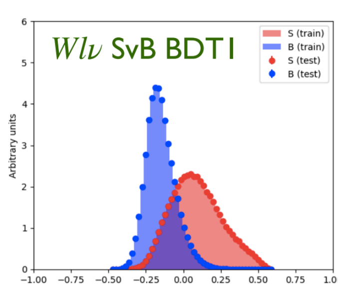
Bring extra sensitivity over κ_λ

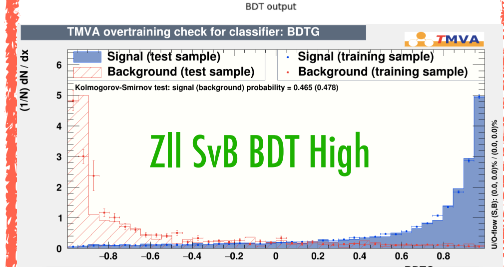




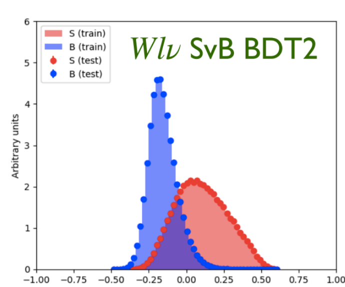
- Samples used for training is Kl = 20 vs Kl = 0
- 3 year MC are combined for training
- Variables and BDT models are optimized in all channels

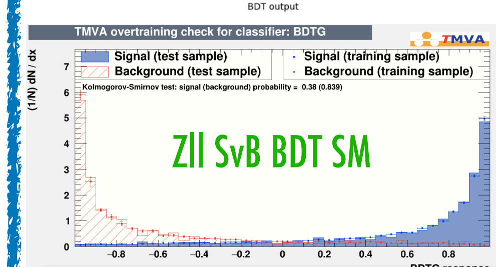
SvB Classifiers





SvB Classifiers



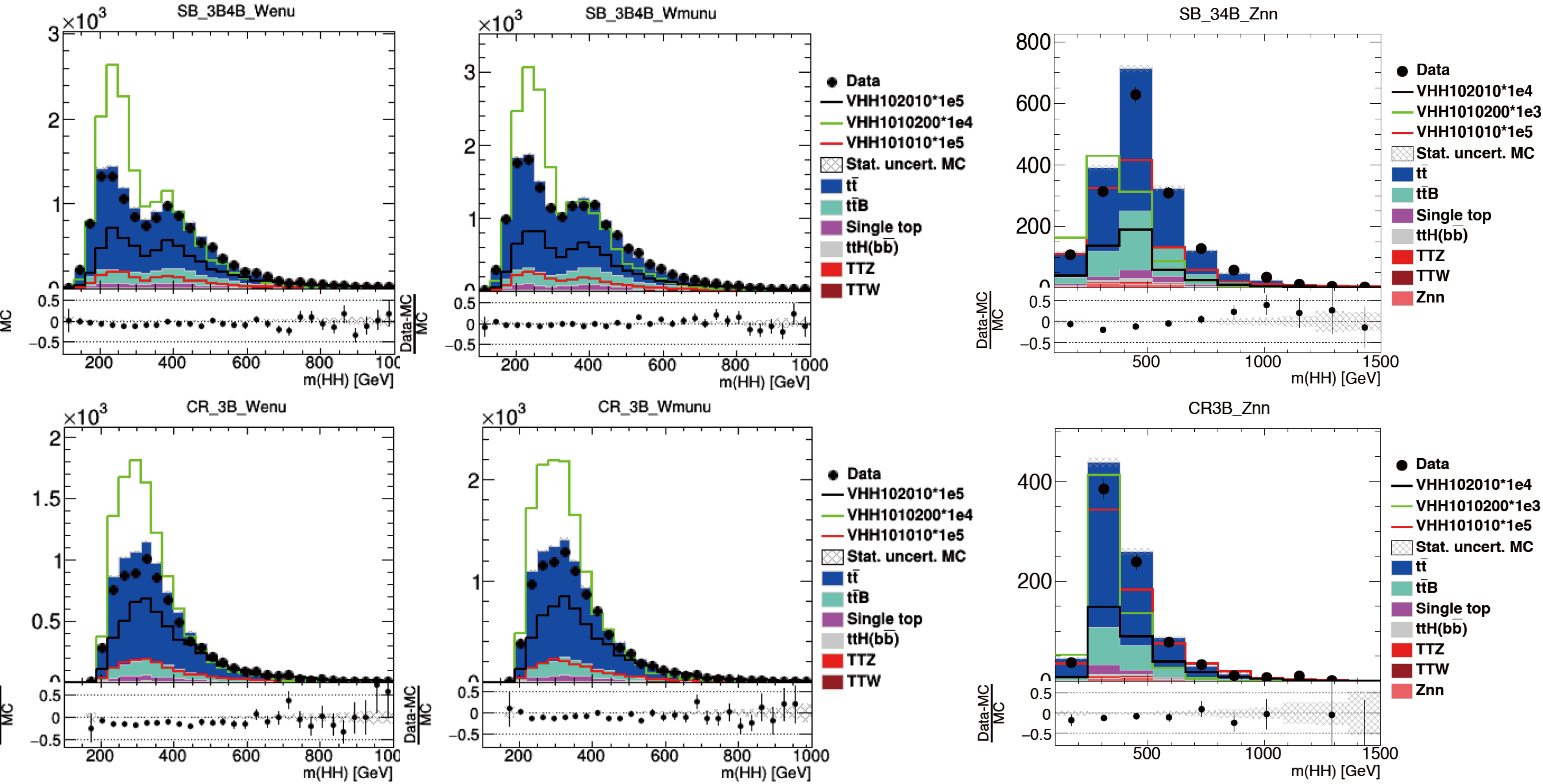


SvB Classifiers

- In V-Leptonic channel
  - 3 channels X 2 Kl Cats = 6 SvB BDTs
- In V-Hadronic channel
  - An ResNet based SvB Classifier is trained
  - Signal labeled by κ_λ Cats.

SvB Classifier scores will be used as the observables for template fit

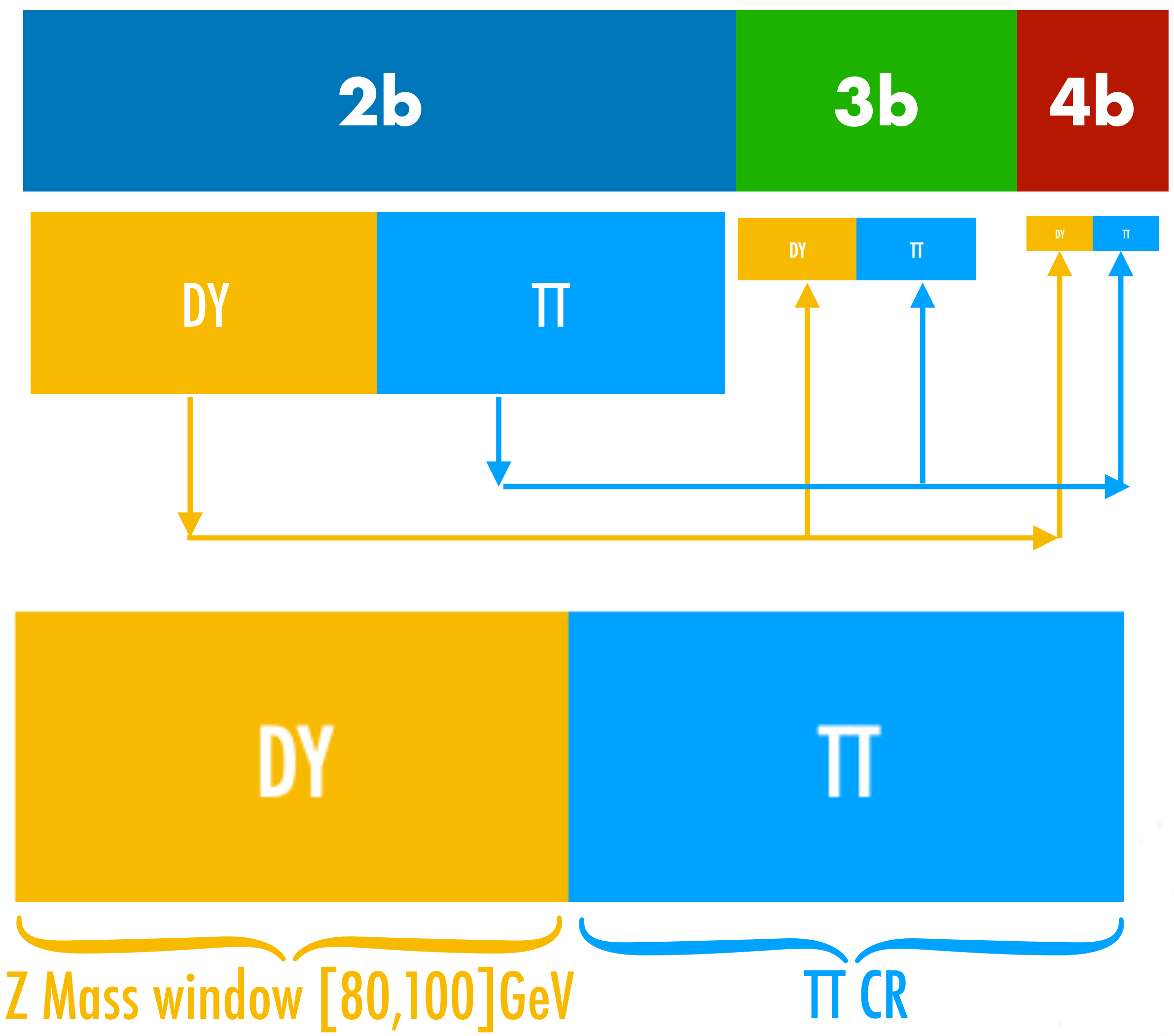
- MET, SL channel
  - Dominant backgrounds are TT, TTBB (Stitched).
  - Re-weighted over top pT. SFs according to the study by top PAG. Uncertainty included.

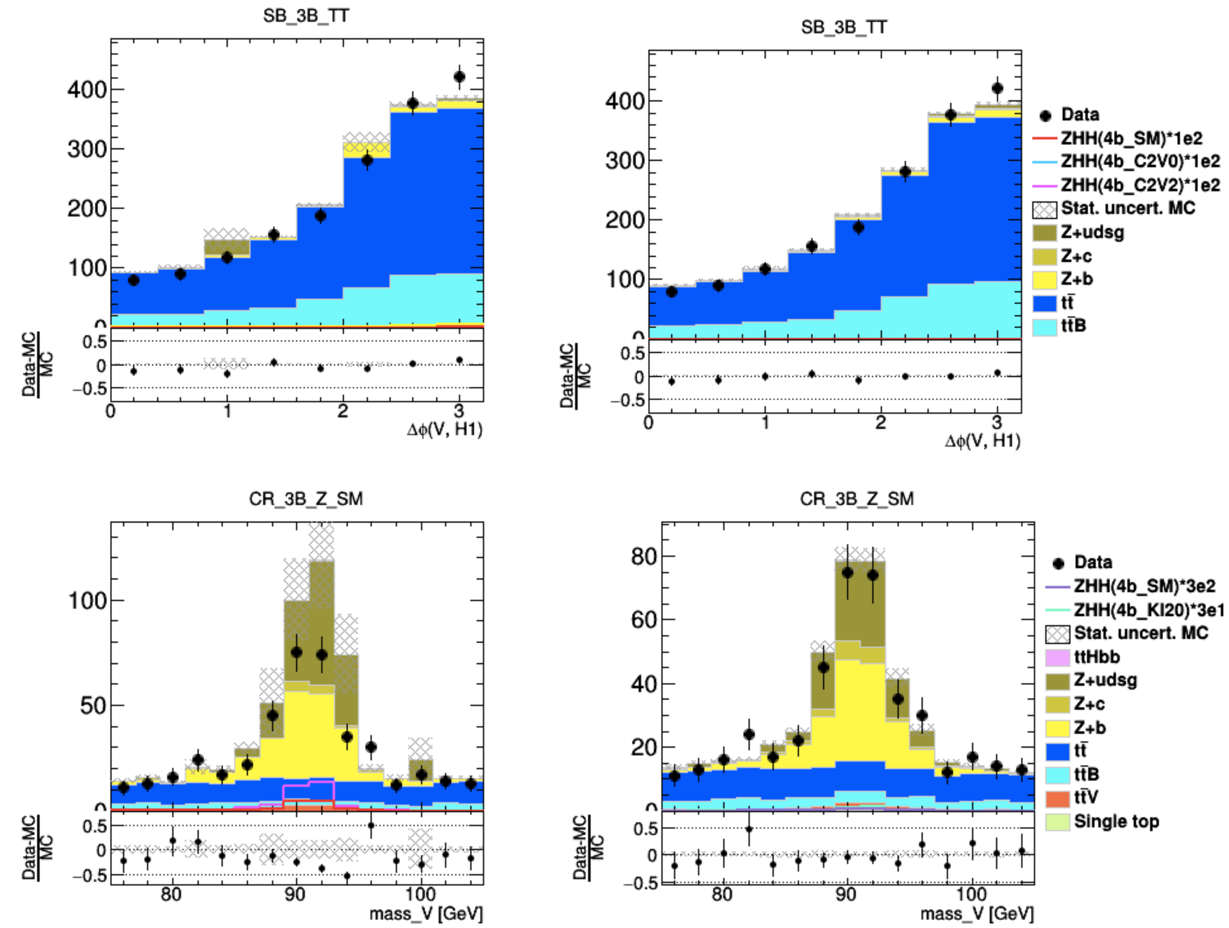


- Validation of All 3 years (link) done!

Double-Lepton Channel

- Main background are TT, DY+Jets
  - Re-weight BDTs are trained to include the information about the differences between 2b-tagged events and 3/4 b-tagged events.
  - 2 main BKGs and 2 b-jet multiplicities introduce 4 RwT. BDTs to realize the re-weighting.
  - SB events are used for training, CR for validation and finally apply on SR events.
  - Input variables are same as the SvB BDTs.





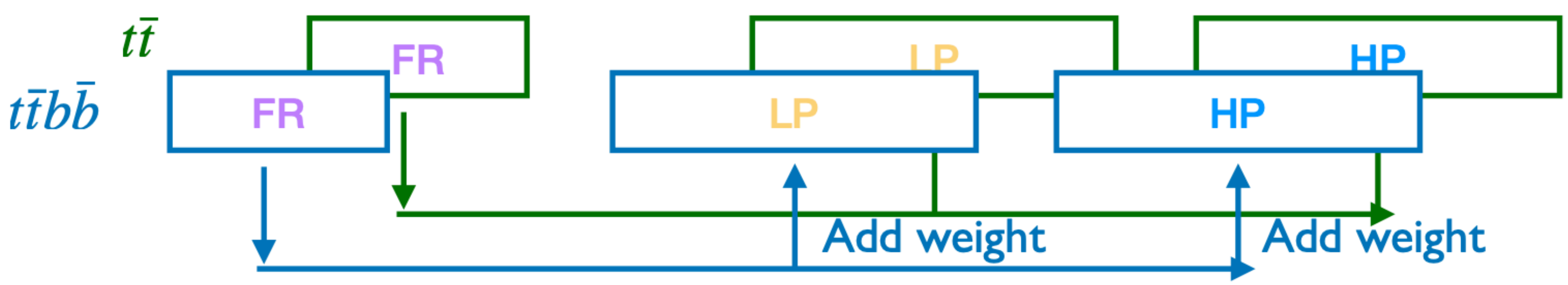
MC → Re-weighted MC

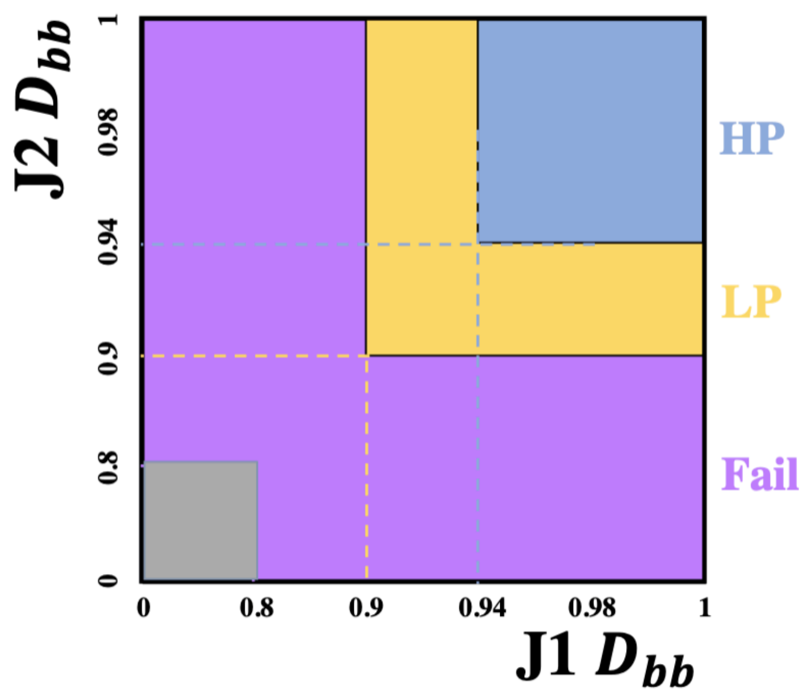
Smaller Stat. Uncertainties ; Reliable background model;

- Inside Z mass window, a fraction fit is applied in every regions to achieve better Data/MC agreement.
- DY/TT process will free float in the final fit.

Background modeling(Boosted)

- Main background are TT(TTBB)
- Same strategy from previous slides about DL channel
- Re-weight events in **FR** to mimic **LP**, **HP**
- 2 sets of weights





- Research region also requires rHH < 50GeV and SB is [50,75]GeV

pT(V)	pT(H1)	pT(H2)	pT(HH)
m(H1)	m(H2)	m(HH)	Phi(V)
Phi(H1)	Phi(H2)	Tau32(J1)	Tau21(J1)
Tau32(J2)	Tau21(J2)		

Variables for RwT BDTs and SvB Classifier in Boosted

- Topology priority
  - By comparing the limit scan results
- Conclusion
  - prioritize the Boosted topology

Boosted topology of the SL, MET channels are included

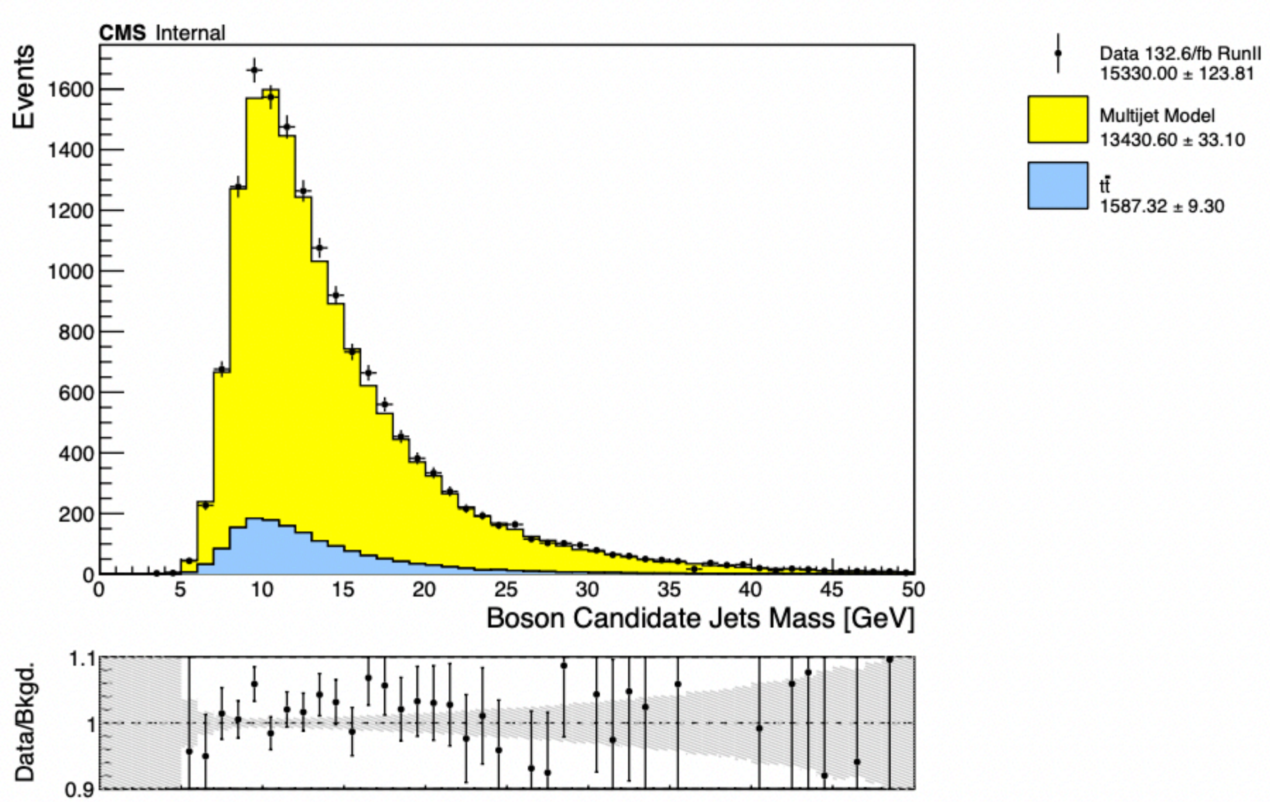
- FH channel
  - Main backgrounds are QCD Multijets and TT.
  - Multijets process is estimated with re-weighted 3-btagged data events.(2 orders fit)
    - Jet Combinatoric Model(JCM): A weight only based on jet multiplicity (pseudo-tag rate fitted in the data minus TT with JCM)

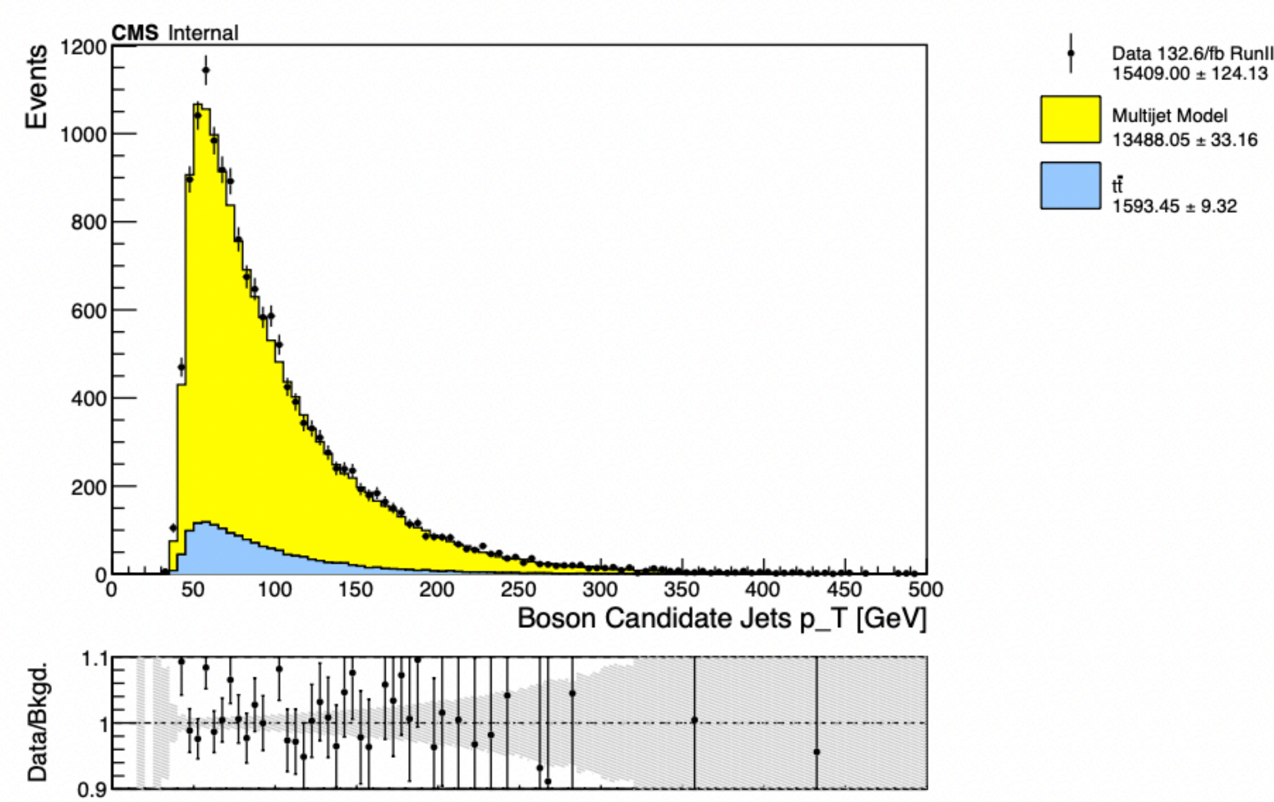
$$w_{\mathrm{JCM}} = t\sum_{i=1}^{n}\binom{n}{i}\times f^{i}(1-f)^{n-i}\times\left[1+(i \mod 2)\times e/n^{d}\right]$$

    - FvT Classifier: A weight mostly based on kinematic, derived by a neural network which has the same architecture as the SvB Classifier.

$$\mathrm{P(M)} = \mathrm{P(M_{4b})} + \mathrm{P(M_{3b})} = \mathrm{P(D_{4b})} - \mathrm{P(t\bar{t}_{4b})} + \mathrm{P(D_{3b})} - \mathrm{P(t\bar{t}_{3b})} \qquad r_{\mathrm{FvT}} = \frac{\mathrm{P(M_{4b})}}{\mathrm{P(D_{3b})}}$$

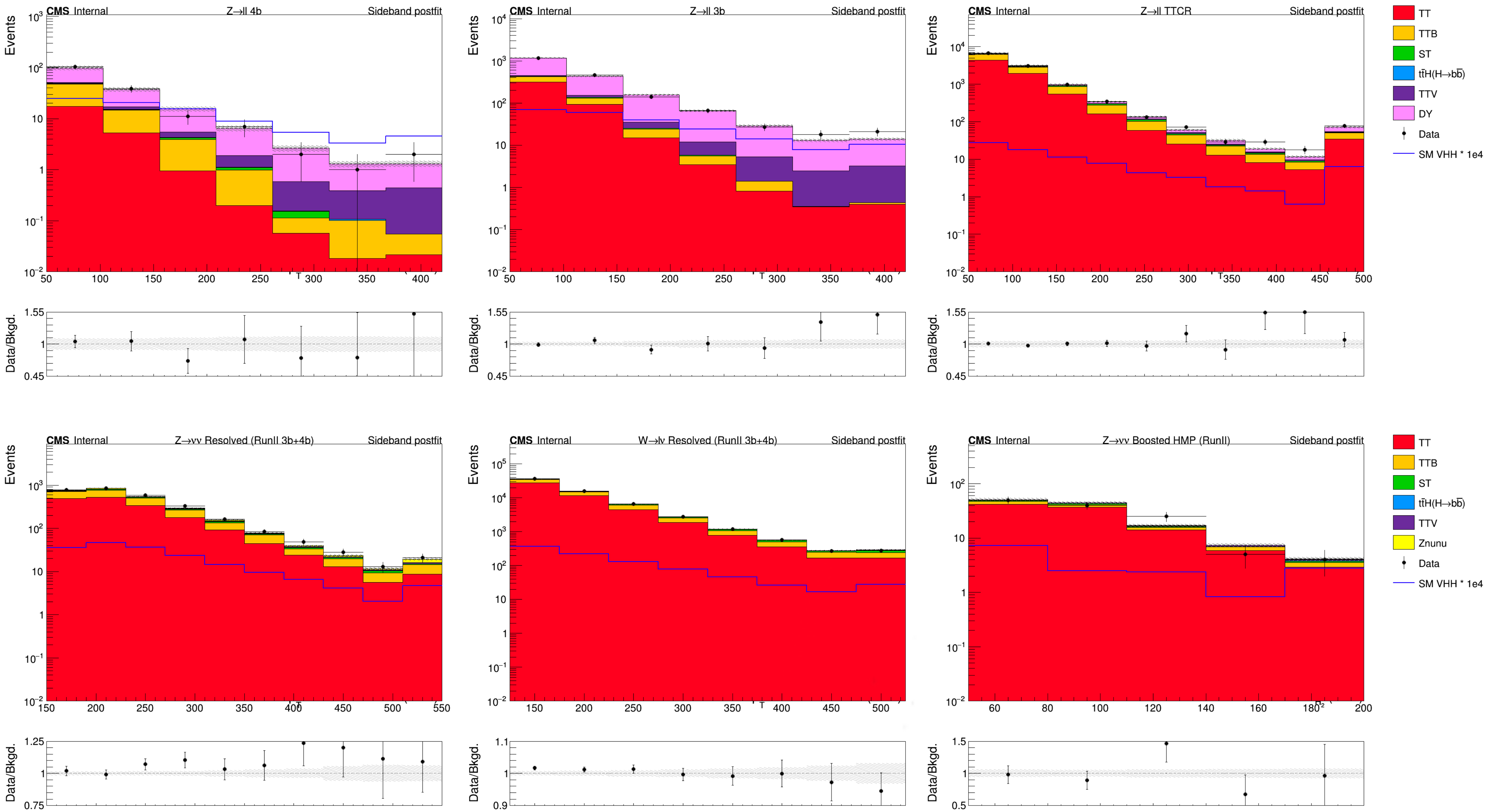
$$r_{\mathrm{FvT}} \times w_{\mathrm{JCM}}$$





* Background systematic uncertainty is extracted from a mixed sample based closure test.

Post-Fit plots in SB and TTCR

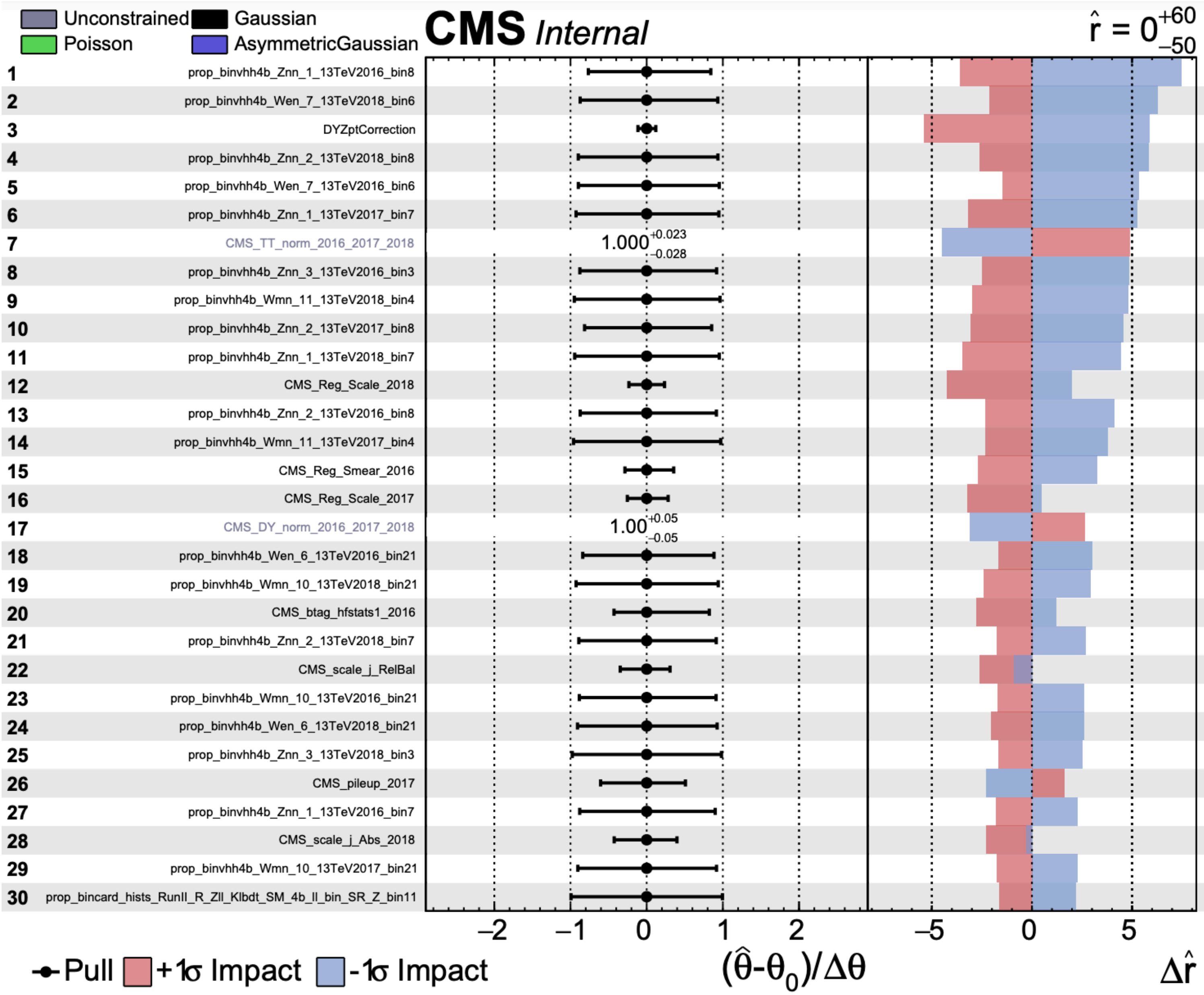


Z mass (0,80) U (100, +∞)

SB and TT CR are fitted simultaneously with Analysis Regions to suppress the Dominant Backgrounds

Systematic uncertainties

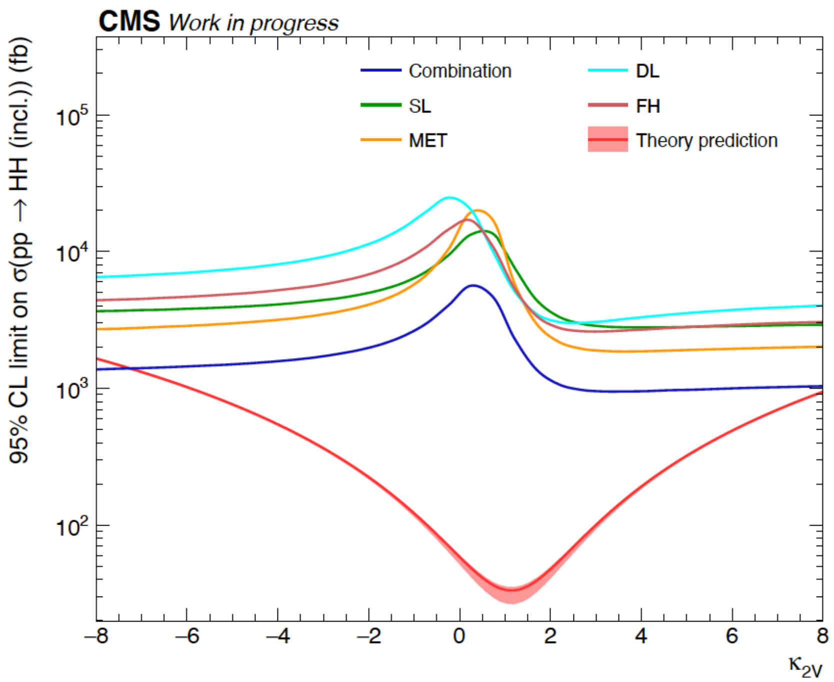
	$We\nu$	$W\mu\nu$	$Z\nu\nu$	Zll	FH
Experimental uncertainties					
autoMCStats	Y	Y	Y	Y	Y
BR_hbb	Y	Y	Y	Y	Y
Luminosity	Y	Y	Y	Y	Y
CMS_TT_norm	Y	Y	Y	Y	N
CMS_TTB_norm	Y	Y	Y	Y	N
CMS_DY_norm	N	N	N	Y	N
CMS_PNet	Y	Y	Y	N	N
CMS_btag	Y	Y	Y	Y	Y
CMS_eff_lepton	Y	Y	N	Y	N
CMS_eff_MET	N	N	Y	N	N
CMS_pileup	Y	Y	Y	Y	W
CMS_MSD_JMR	Y	Y	Y	N	N
CMS_MSD_JMS	Y	Y	Y	N	N
CMS_res_j	Y	Y	Y	Y	Y
CMS_scale_j	Y	Y	Y	Y	Y
CMS_unclusteredEnergy	Y	Y	Y	N	N
CMS_eff_j_PUJET_id	N	N	N	N	Y
CMS_bbbb_Multijet	N	N	N	N	Y
Re-weight DY	N	N	N	Y	N
Re-weight TT	Y	Y	Y	Y	N
Re-weight TTB	Y	Y	Y	Y	N
Theoretical uncertainties					
QCDscale_VHH	Y	Y	Y	Y	Y
pdf_Higgs_VHH	Y	Y	Y	Y	Y
ZHH_NNLO	Y	Y	Y	Y	~~W~~ Y

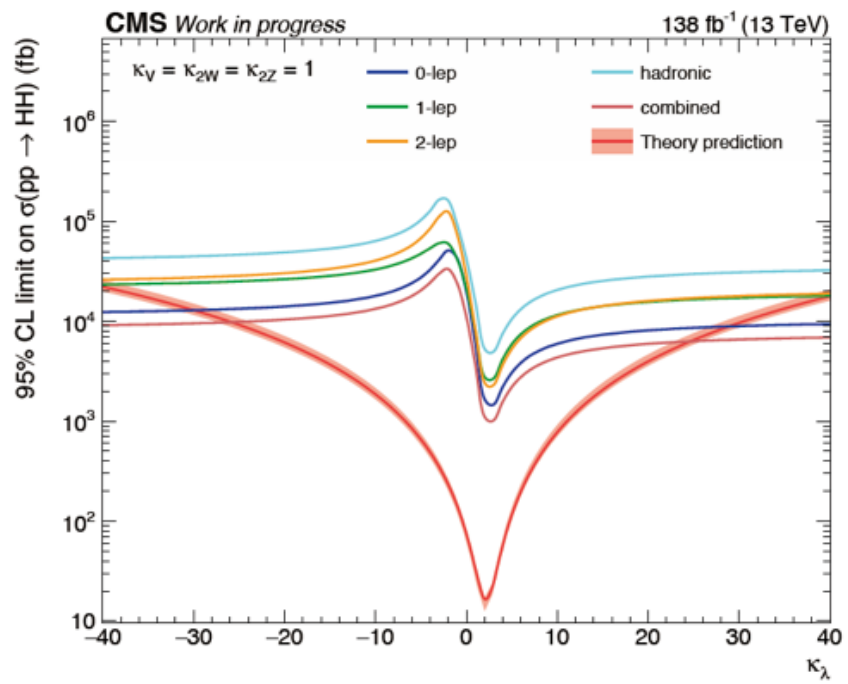


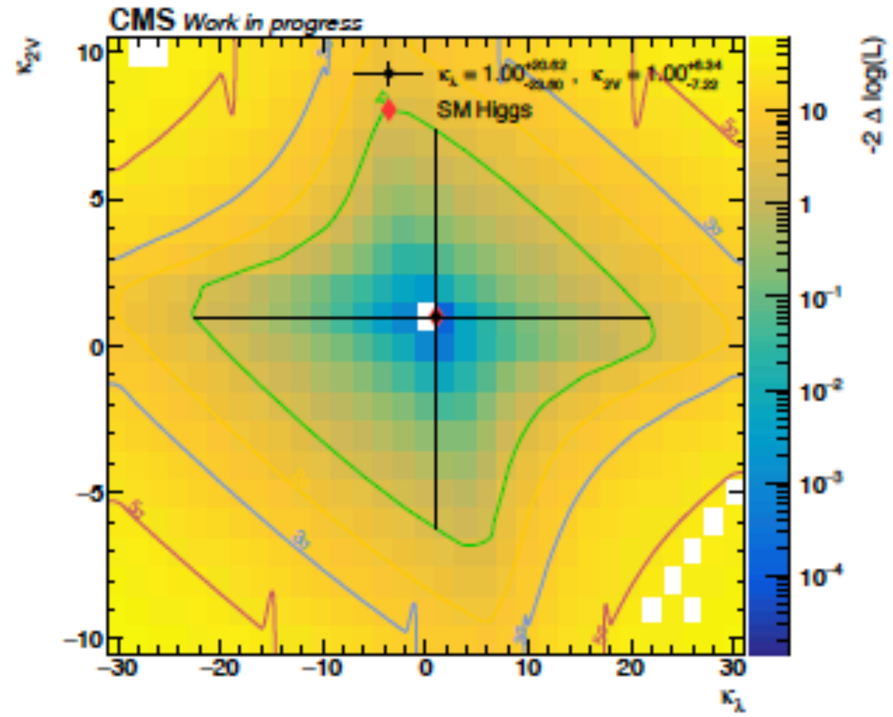
- Top 30 nuisance parameters ranked in the impact plot

All systematic are included in the data-cards

Result - Upper Limit Scan



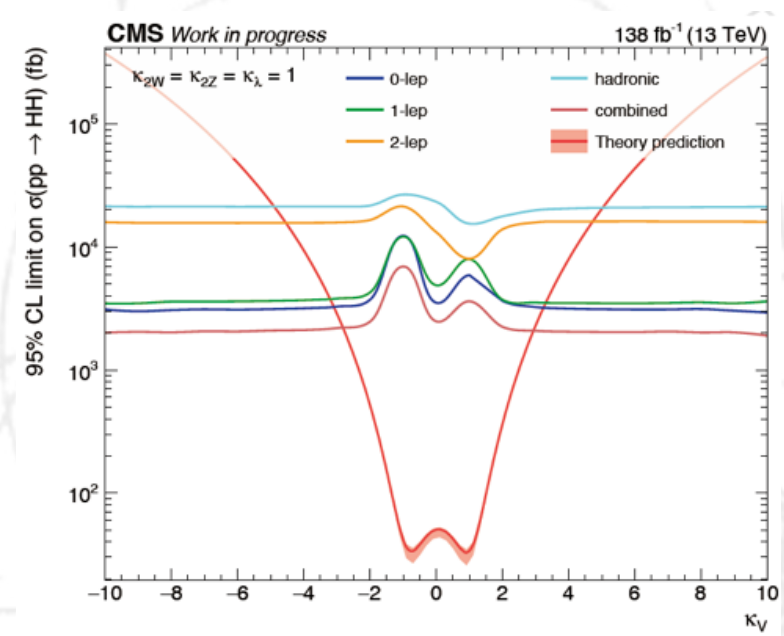




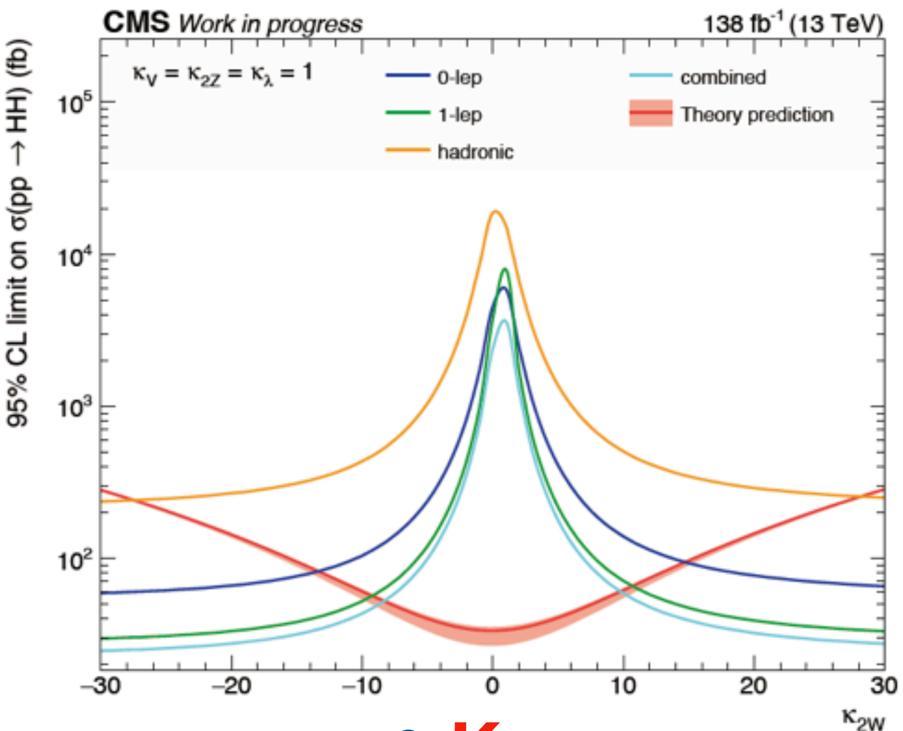
- K_{VV}
- K_λ
- K_{VV} vs K_λ

- Expected upper limits in different VHH sub-channels
- Powerful sensitivity by combining multiple channels
- Table: 95% exclusion [lower, upper] limits on Kv, K_{vv} and K_λ in the combined results.

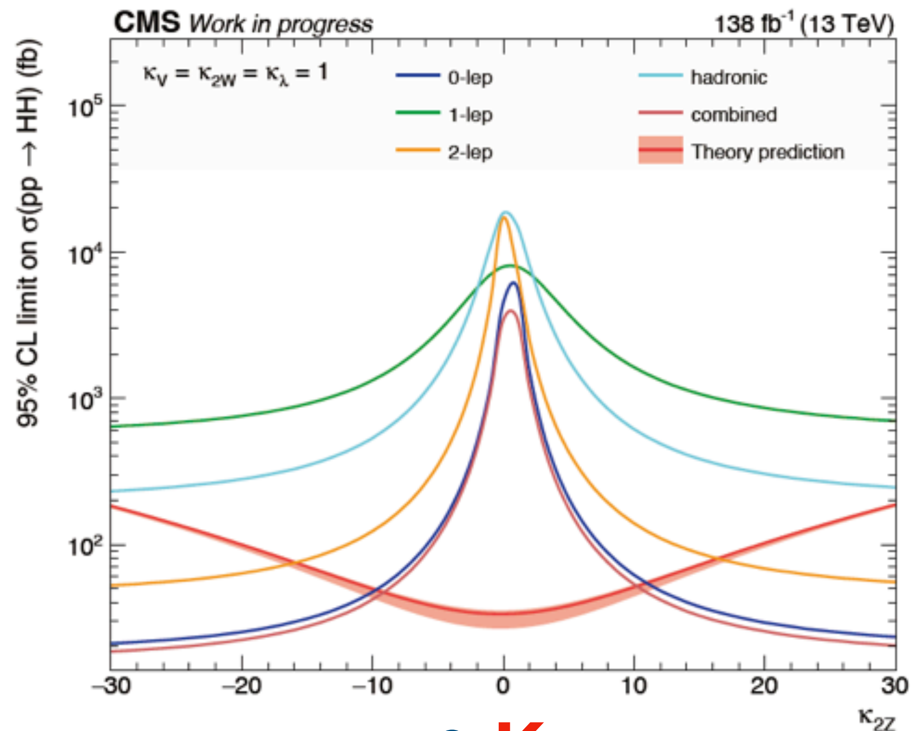
Coupling	K_V	K_W	K_λ
Combination	[-2.82, 2.93]	[-7.24 ,8.35]	[-25.85 , 24.64]



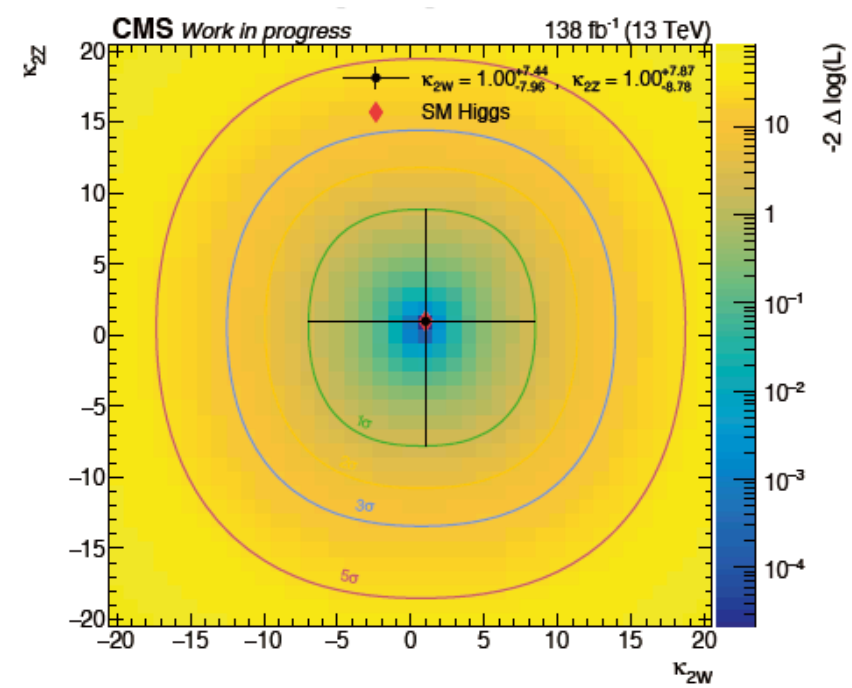
VHH Analysis — New Result - Upper Limit Scan



- K_{WW}



- K_{ZZ}

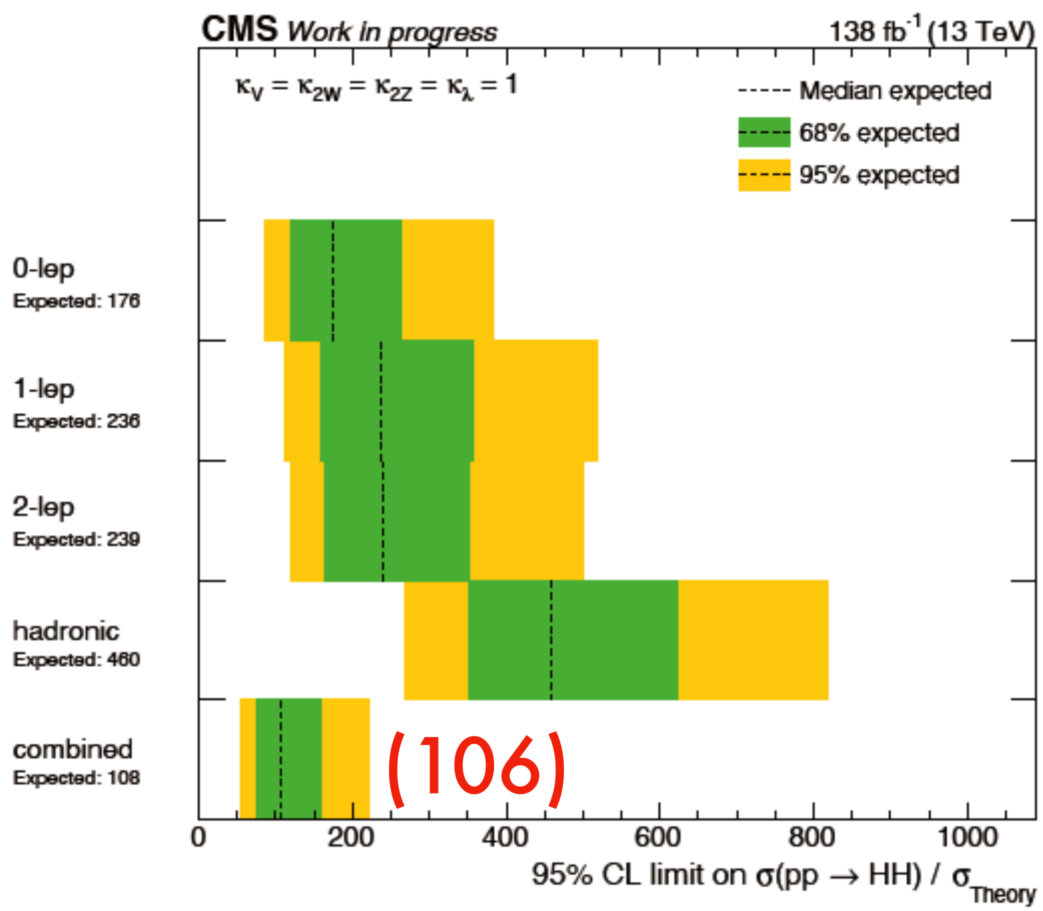


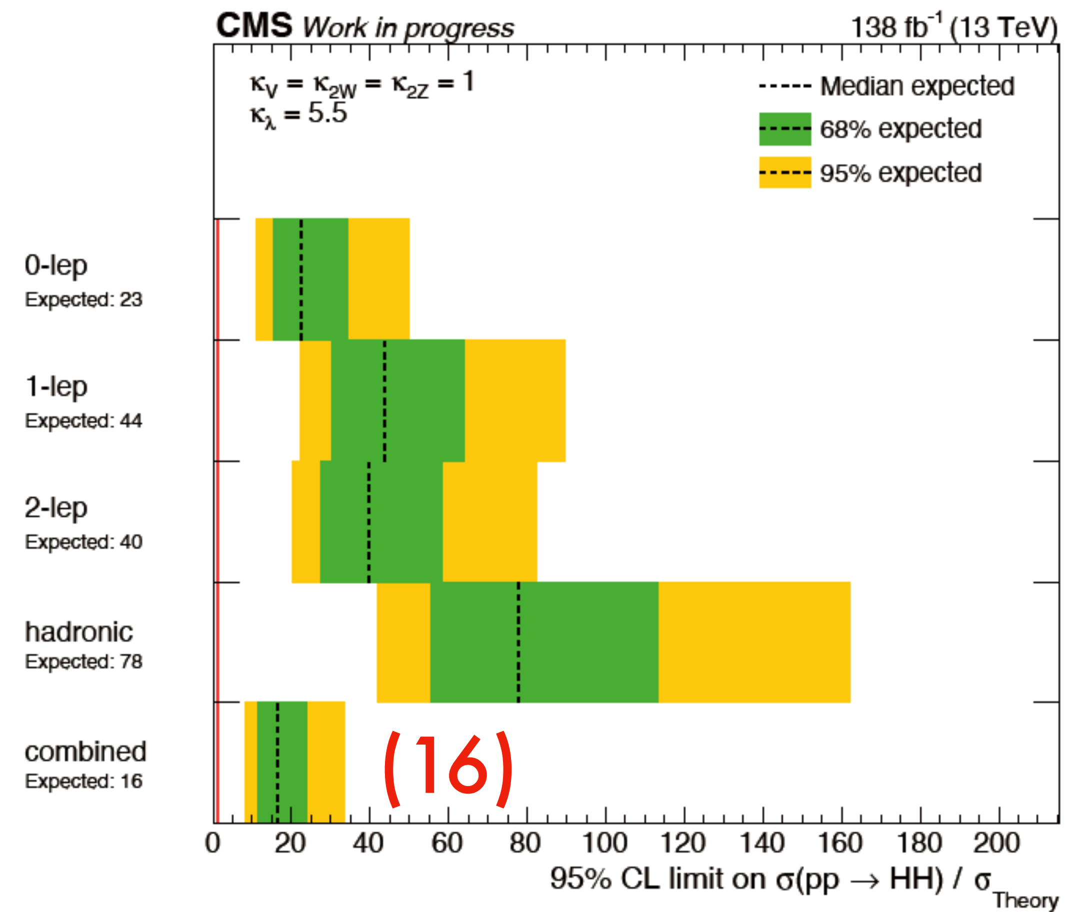
- K_{ZZ} vs K_{WW}

- Expected upper limits in different VHH sub-channels
- The first time probing K_{ZZ} and K_{WW} !

- Table: 95% exclusion [lower, upper] limits on Kv, K_{vv}(K_{ww} K_{zz}) and K_{λ} in the combined results.

Coupling	K_v	K_W	K_{λ}	K_{ZZ}	K_{WW}
Combination	[-2.82, 2.93]	[-7.24 ,8.35]	[-25.85 , 24.64]	[-9.16 , 10.23]	[-8.34 , 9.79]

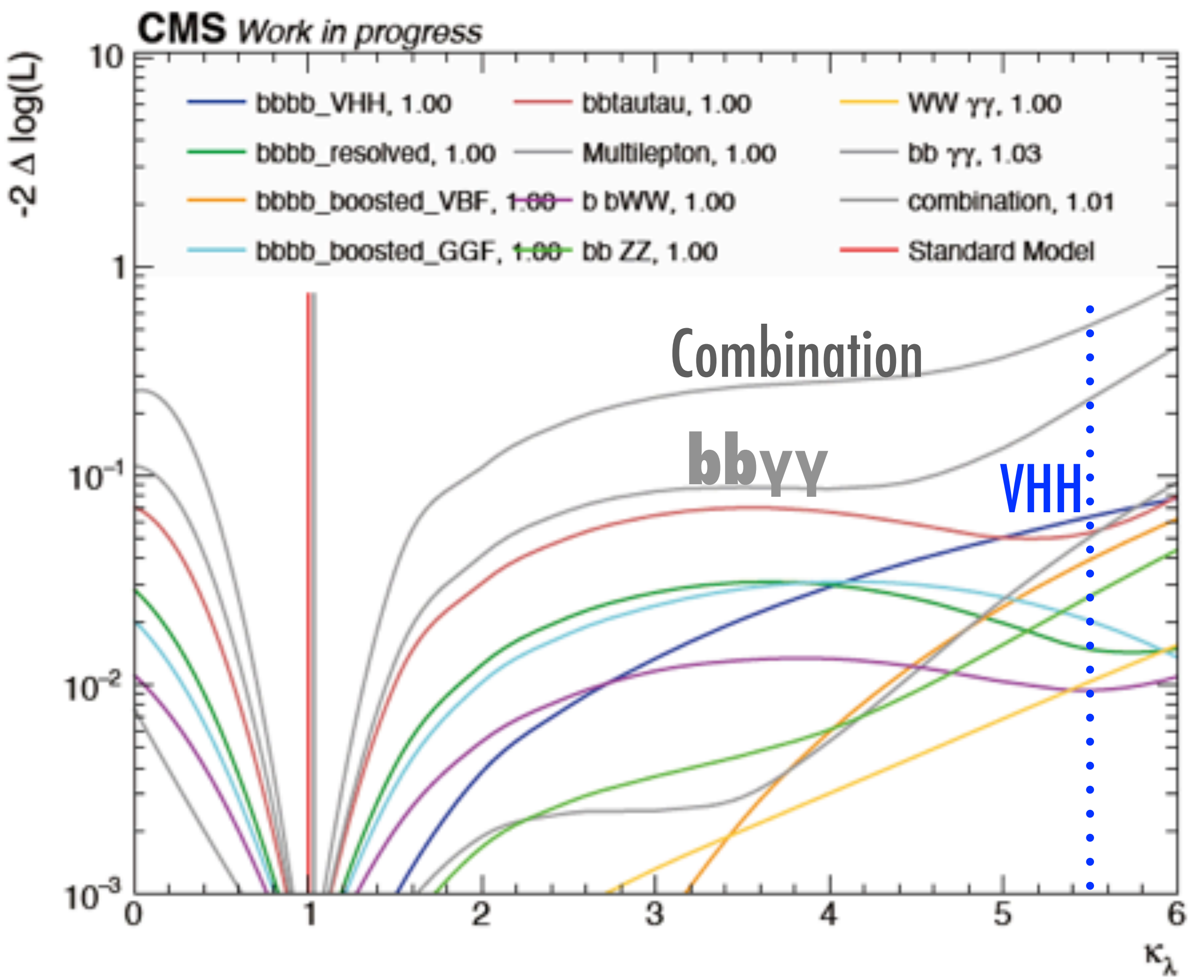


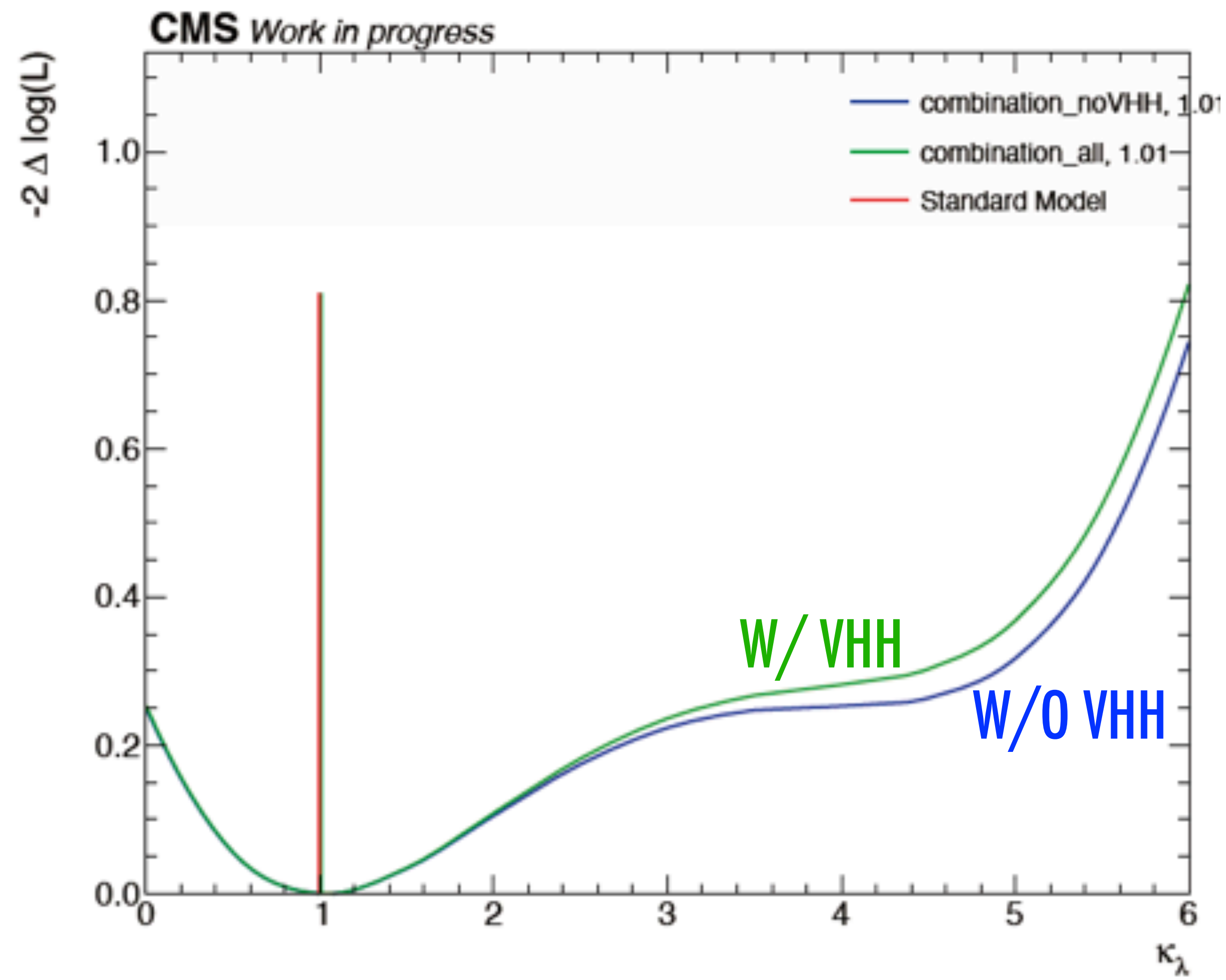


- 95% CL expected Limit on VHH production At **K_λ** = 1.0

- 95% CL expected Limit on VHH production At **K_λ** = 5.5

- Expected signal strength at different **K_λ** strength points
- Reasonable sensitivity toward **K_λ**
- Powerful sensitivity at **K_λ** = 5.5 (at the positive side) considering the expectations based on SM sections.





- Likelihood scan for **κ_λ** comparing to other HH analysis

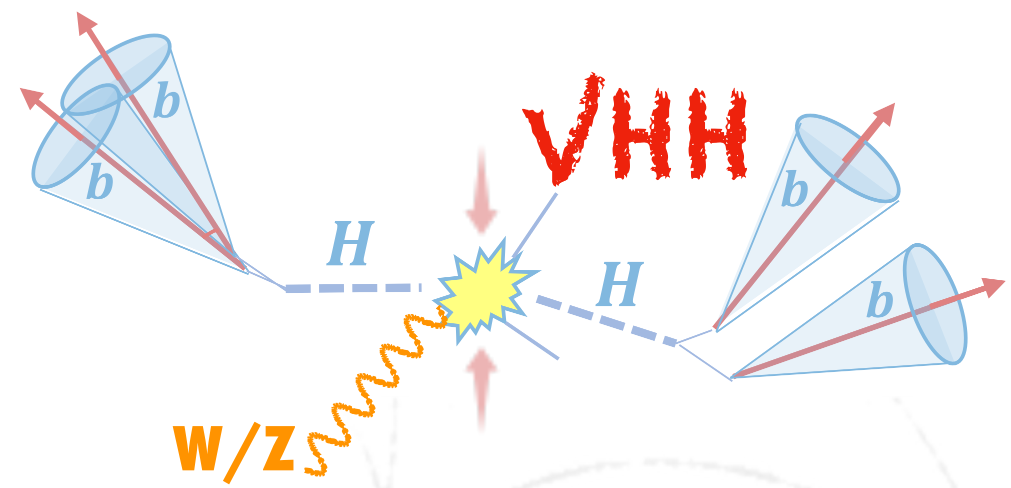
- Likelihood scan comparison between VHH and other analysis
- VHH provides good augment to the **κ_λ** sensitivity above SM point due to non-destructive feature. (Second, only after bbγγ around **κ_λ** = 5.5)
- The likelihood improves around 10% start from **κ_λ** = 3 and up to more than 15%.
- Data-card taken from latest version of HH Run 2 card repository

Summary

- First search for VHH production in the HH to bbbb final state in CMS
  - Complementary to ggF and VBF analyses.
  - Probe the κ_{ZZ} and κ_{WW} coupling.
- Analysis strategy optimized for maximal sensitivity
  - SM κ_λ and Strong κ_λ categories with BDT discriminant.
  - Kinematic subcategories for maximal sensitivity.
  - Prioritize the boosted topology for better combined upper limit.
  - SvB classifiers trained in each channels for maximal sensitivity.
  - SB/TTCR are fitted simultaneously to suppress the uncertainty of the dominant backgrounds.
- Background estimation
  - BDT re-weighting method for accurate multidimensional modeling in stats limited channels and topologies.
  - Dedicate data-driven method used for multi-jet background estimation.
  - Closure tests/Validations done in control region/side band/validation regions in all channels.
- Expected Limit: 106 times the SM cross section prediction
- Couplings: κ_λ [-25.85 , 24.64], κ_{VV} [-7.24 ,8.35], κ_V [-2.82, 2.93]
- First Probe correlation: κ_{ZZ} [-9.16, 10.23], κ_{WW} [-8.34 , 9.79]
- Improve the HH combination results especially in positive κ_λ side.

Search for Higgs boson pair production in $b\bar{b}b\bar{b}$ final state in association with a vector boson with the CMS detector

HIG-22-006



John Alison[2], Chayanit Asawatangtrakuldee[1], Agni Bethani[3], Yihui Lai[3], Chuyuan Liu[2], Chris Palmer[3], Xiaohu Sun[4], Long Wang[3], Licheng Zhang[4]

1. Chulalongkorn University
2. Carnegie-Mellon University
3. University of Maryland
4. Peking University

23rd-27th Nov. 2022 Higgs Meeting

On behalf of PKU CMS group and VHH4b analysis team

backup

★ Signal Samples

- Linearly interpolate/extrapolate existing samples to get more couplings for limit scan
  - According to the talk, implemented in HHModel that used by HH analysis

- Use Moore-Penrose inverse to accommodate 8 signal samples

LO

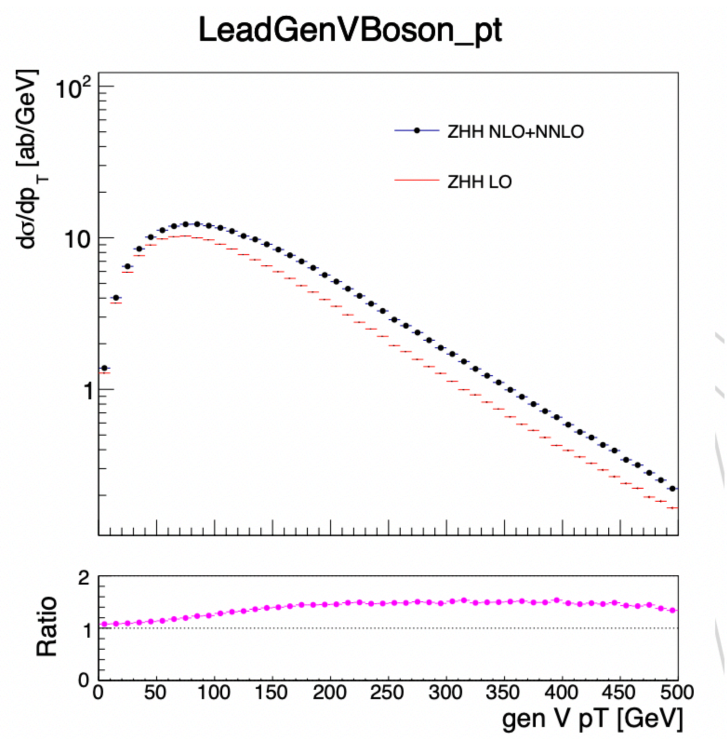
κ_V	κ_{2V}	κ_λ
0.5	1.0	1.0
1.0	0.0	1.0
1.0	1.0	0.0
1.0	1.0	1.0
1.0	1.0	2.0
1.0	2.0	1.0
1.5	1.0	1.0
1.0	1.0	20.0

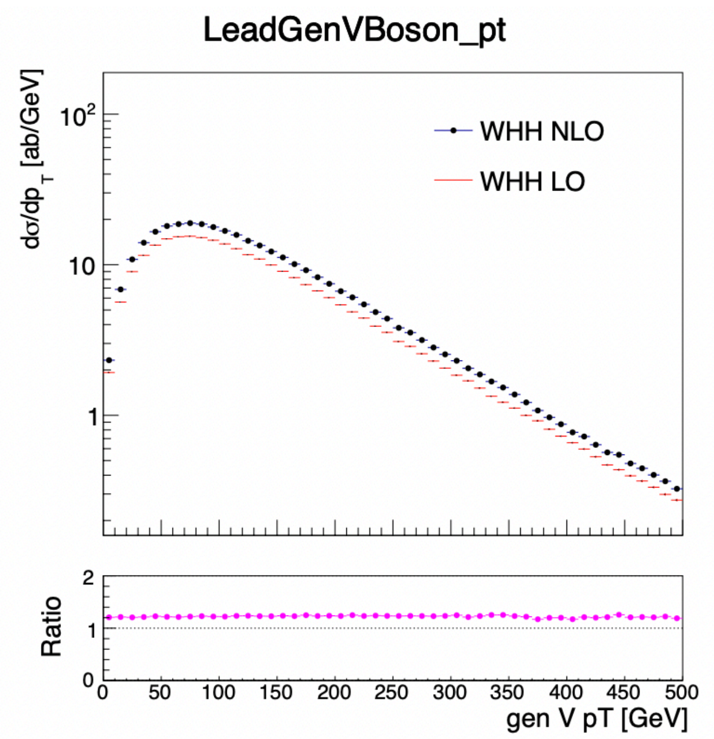
$$\sigma(\kappa_\lambda, \kappa_V, \kappa_{2V}) = c^T(\kappa_\lambda, \kappa_V, \kappa_{2V})C^{-1}\sigma$$

- ZHH signal re-weighted and scaled to NNLO
- WHH signal scaled to NLO

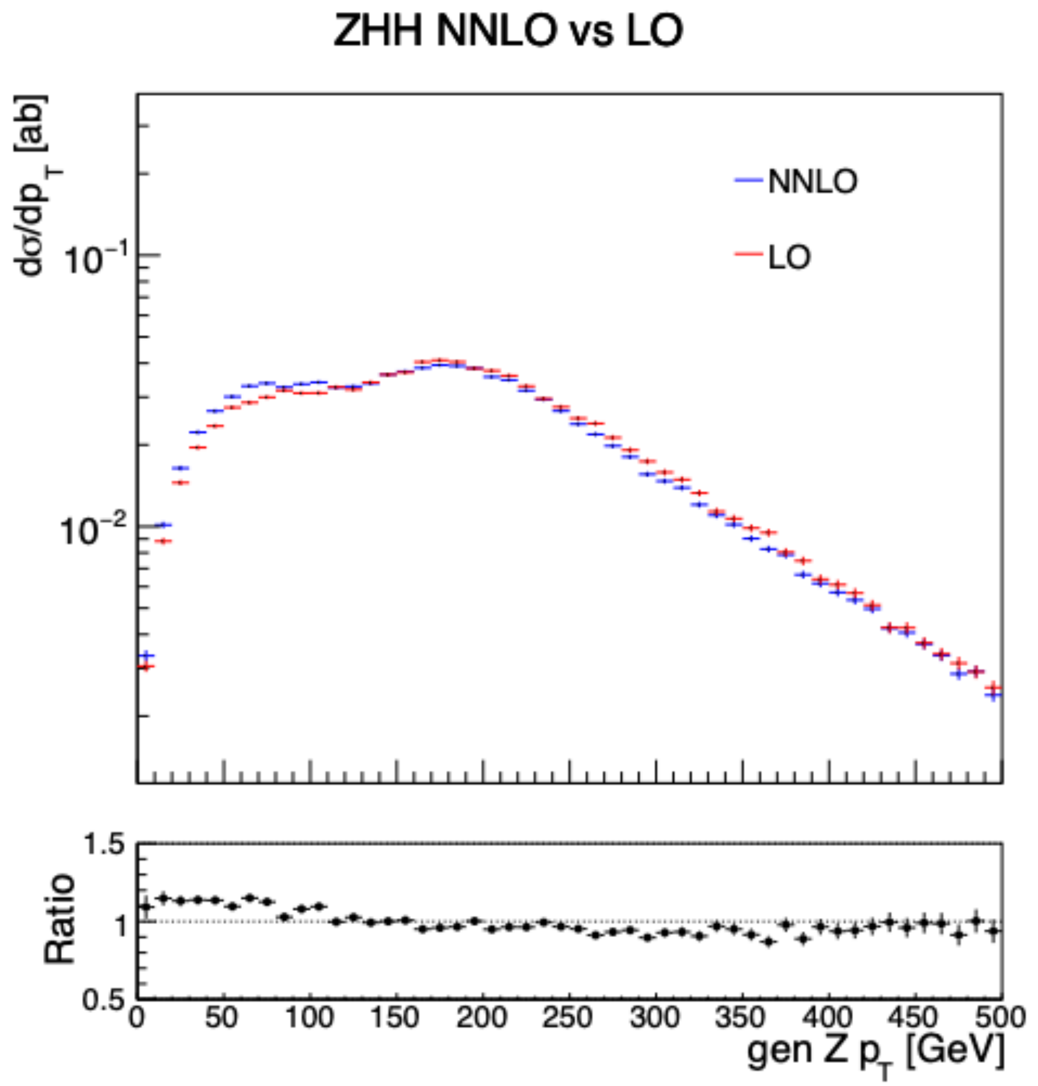
VHH NLO/NNLO correction

- ZHH signal re-weighted and scaled to NNLO
- WHH signal scaled to NLO





- The re-weighted LO compared with NNLO after full selection.



- ZHH signal has been further studied (link):

Look at other dimensions variables after LHE_Vpt re-weighting

- Residual difference over LHE_Vpt will be calculated as a composition of systematics uncertainty.

Background MC Processes

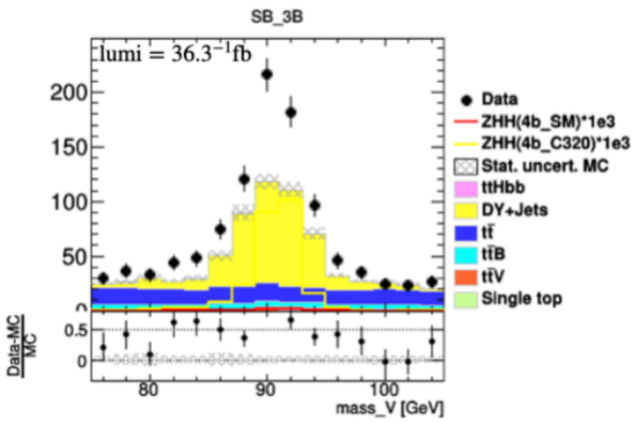
★ DY+Jets

- DY+Jets (LO) MC are re-weighted to NLO over $p_T(V)$ (LHE_Vpt), which makes the samples softer and improves the agreement with data.
- Then scaled to NNLO by K-Factor.

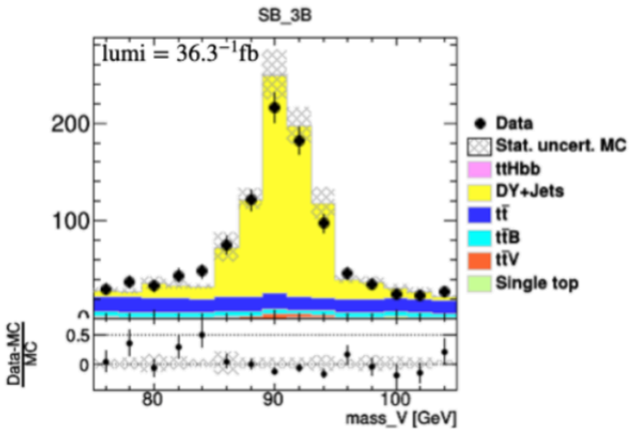
- Same as VHbb RunII Analysis

Table 8: NLO sample

Channel	nB	NLO/LO
DYJetsToLL	0	$1.650 \pm 0.002 - (1.707 \pm 0.020) \times 10^3 pT(\mathrm{V})$
DYJetsToLL	1	$1.534 \pm 0.010 - (1.458 \pm 0.080) \times 10^3 pT(\mathrm{V})$
DYJetsToLL	2	$1.519 \pm 0.019 - (1.916 \pm 0.140) \times 10^3 pT(\mathrm{V})$



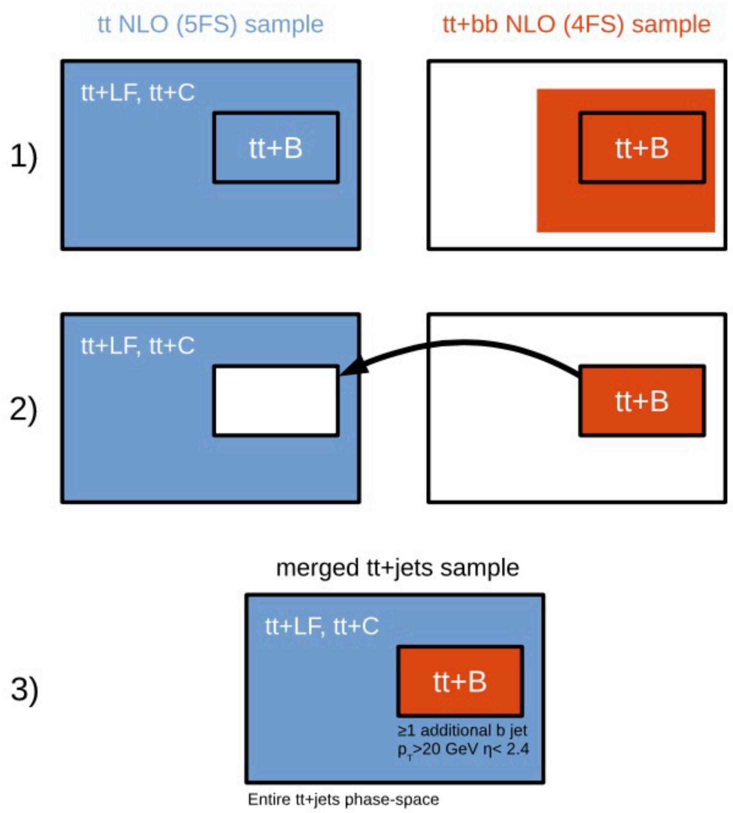
Before



After

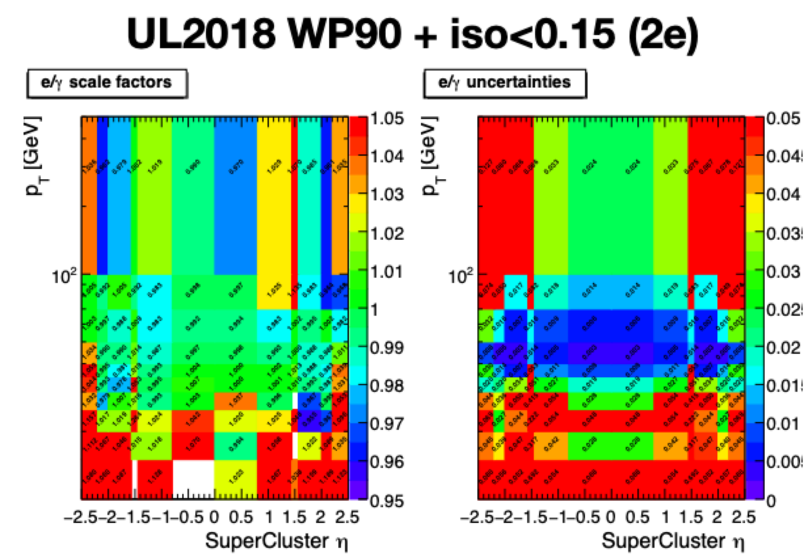
★ TTbar

- Follow the strategies in tH/ttH(bb) Analysis: (in Lep-Channels)
  - $t\bar{t}b\bar{b}$ in Powheg NLO $t\bar{t}$ 5FS sample is from parton shower which will bring large uncertainties.
  - Powheg NLO 4FS sample has better performance in modeling $t\bar{t} + b\bar{b}$ kinematics.
- We need to stitch together these two samples for better background modeling.
  - In each $t\bar{t}$ event, define 'additional b-jet' as a particle level b jet with pT>20GeV and |eta|<2.4 and not from top decay.
  - Replace $tt + B$ events in tt(5FS) with ttbb(4FS)

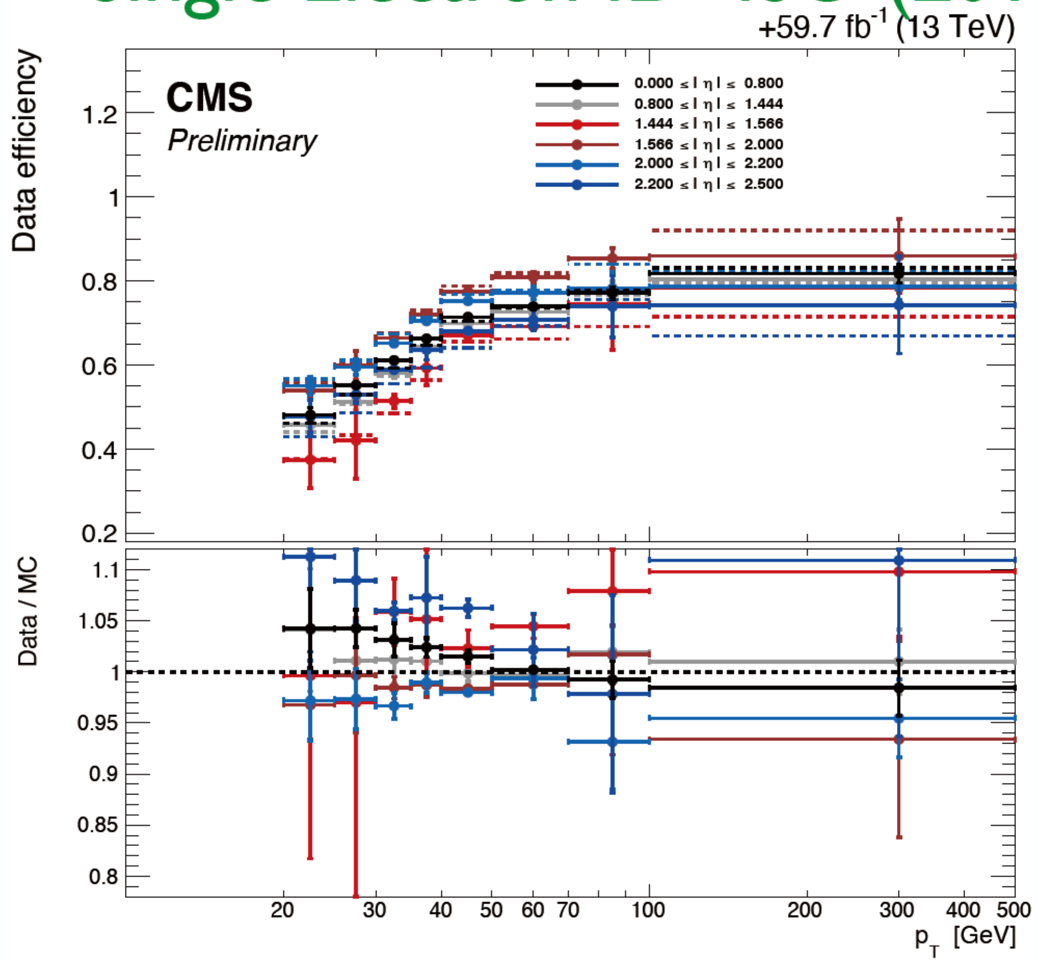


Leptonic Channel

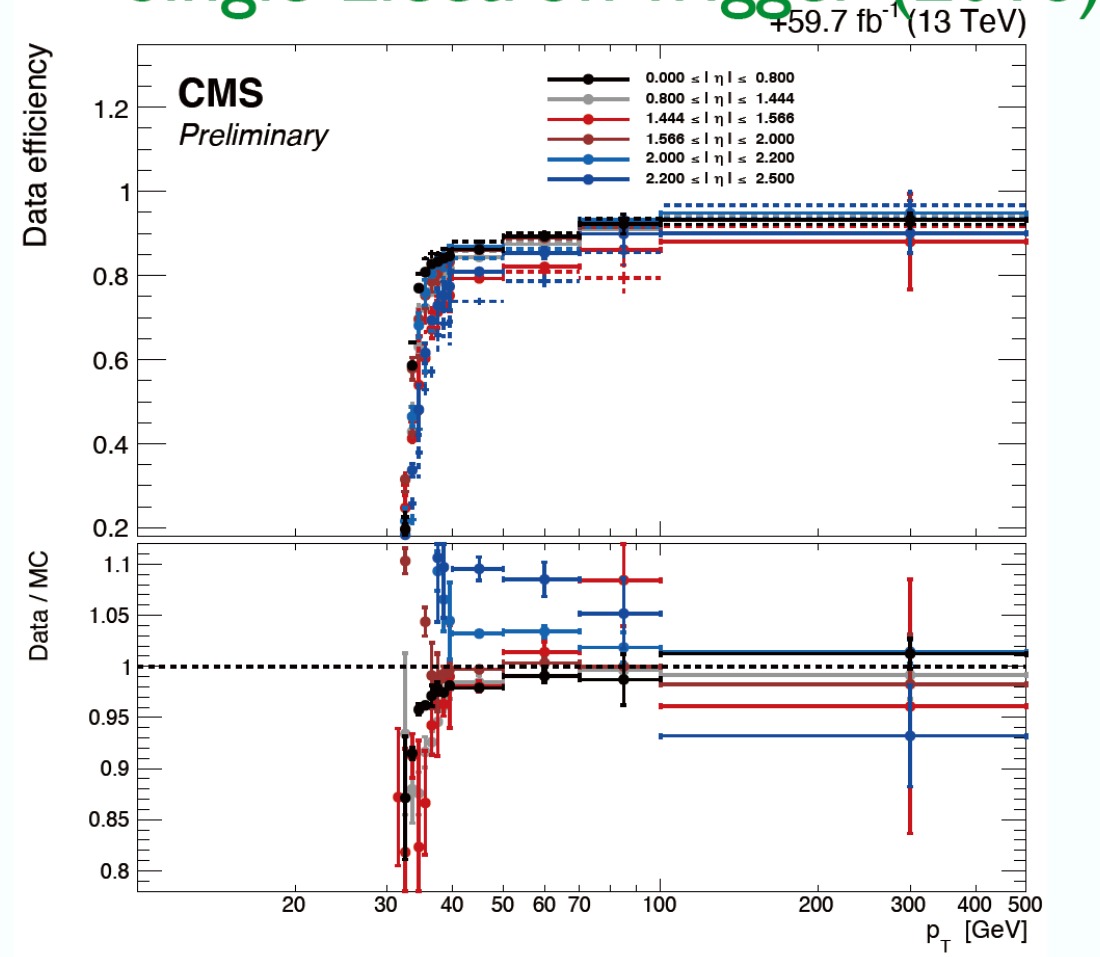
- All the lepton efficiencies are measured, Json file stored*
  - <u>Electron SF</u>: Reco X ID_ISO X Trigger



Single Electron ID+ISO (2018)

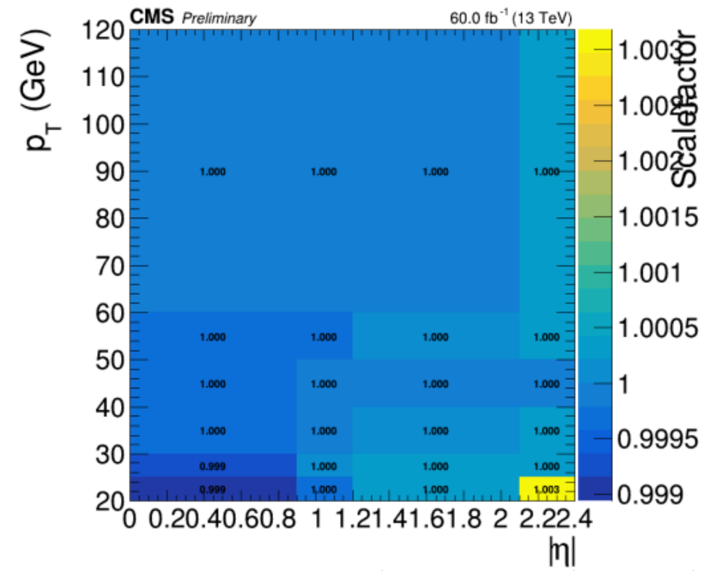


Single Electron Trigger (2018)



* Json files stored in: /eos/cms/store/group/phys_higgs/hbb/ntuples/VHH4b_Vleptonic_SF_UL

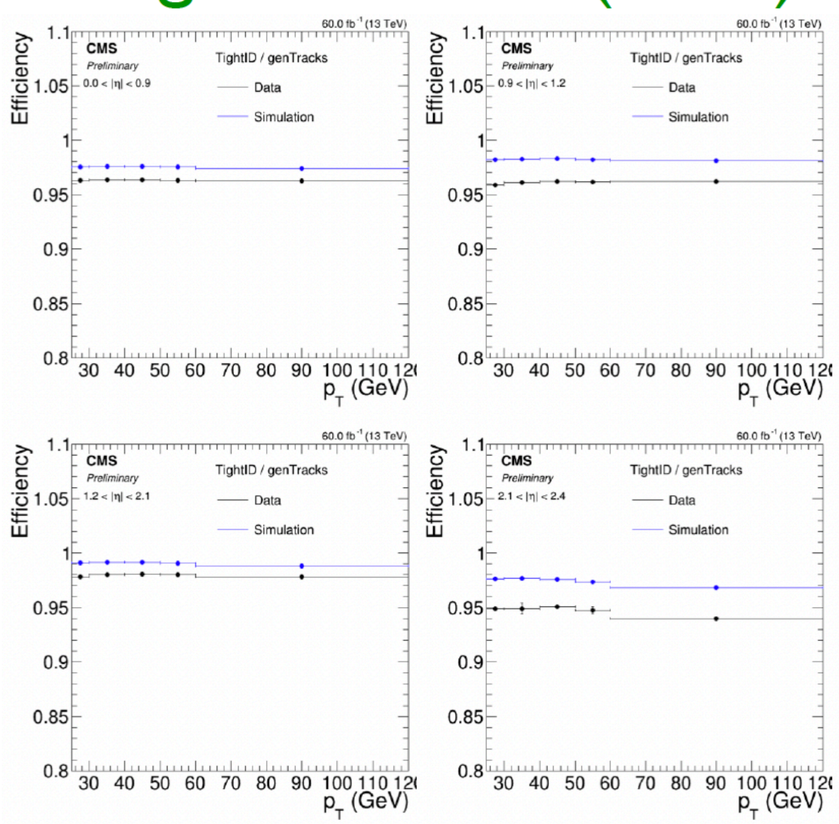
Leptonic Channel



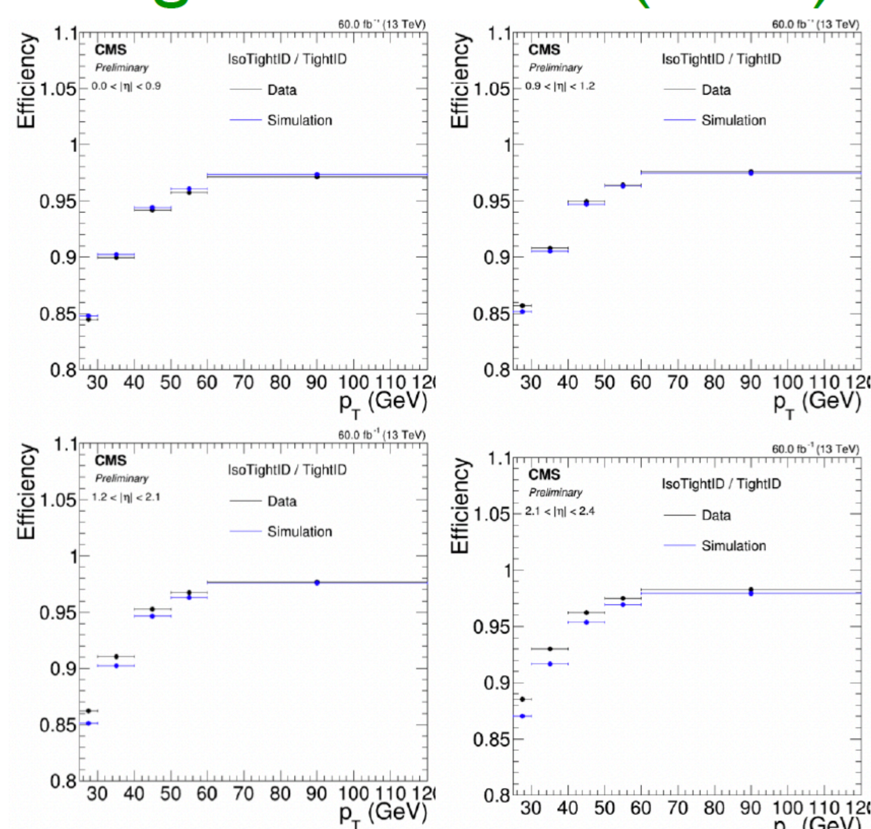
- All the lepton efficiencies are measured, Json file stored*
    - Muon SF: ID X ISO X Trigger

- LooseISO/LooseID (2018)

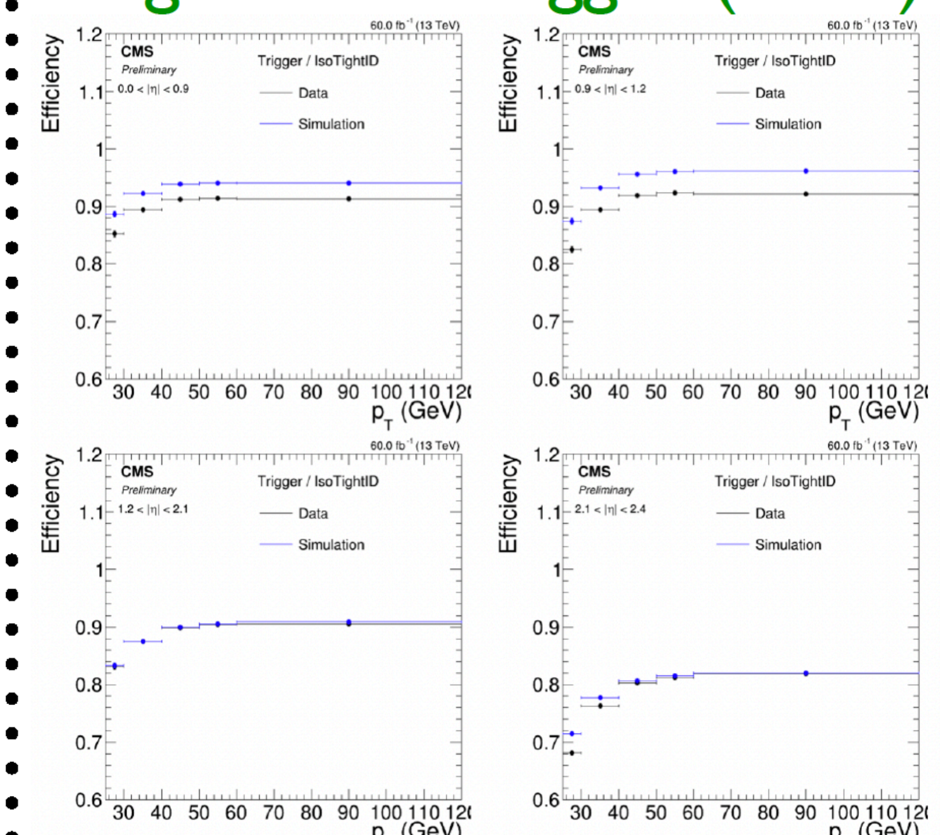
Single Muon ID (2018)



Single Muon ISO (2018)



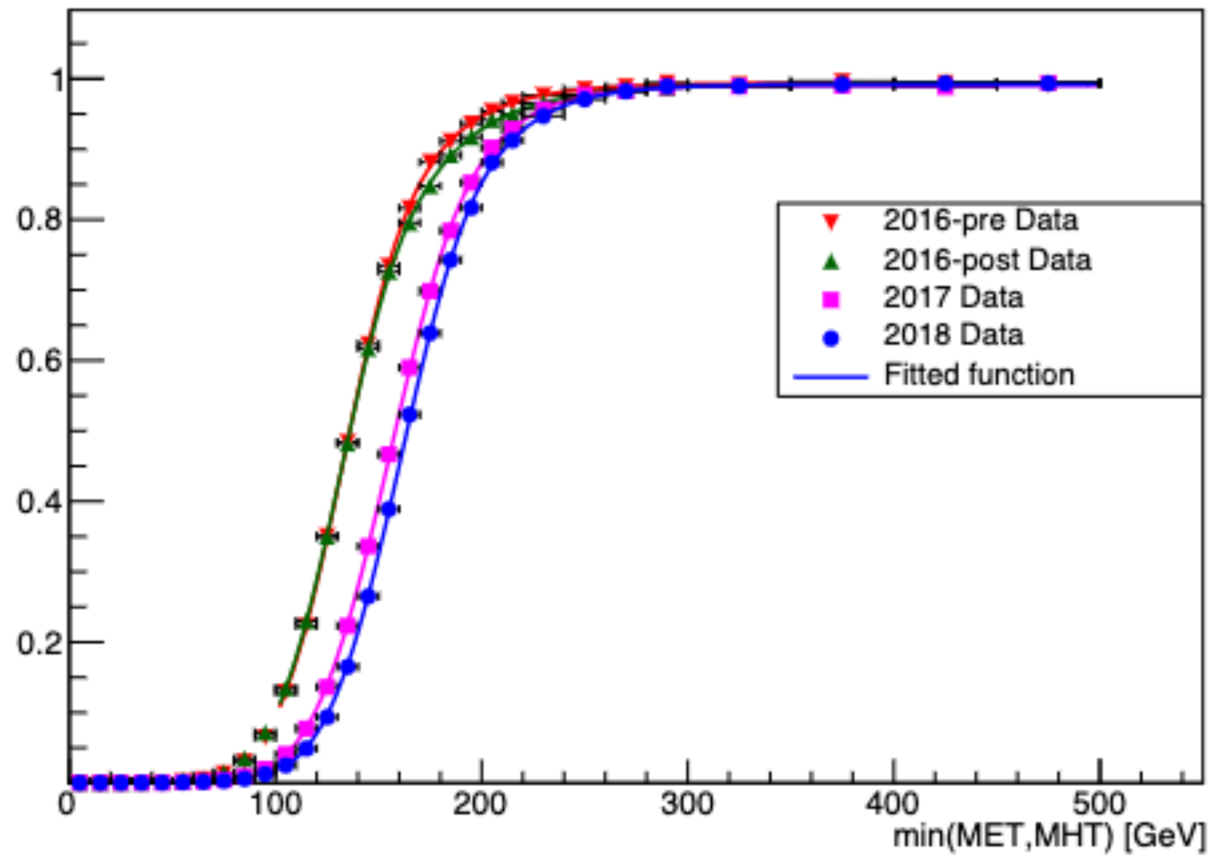
Single Muon Trigger (2018)

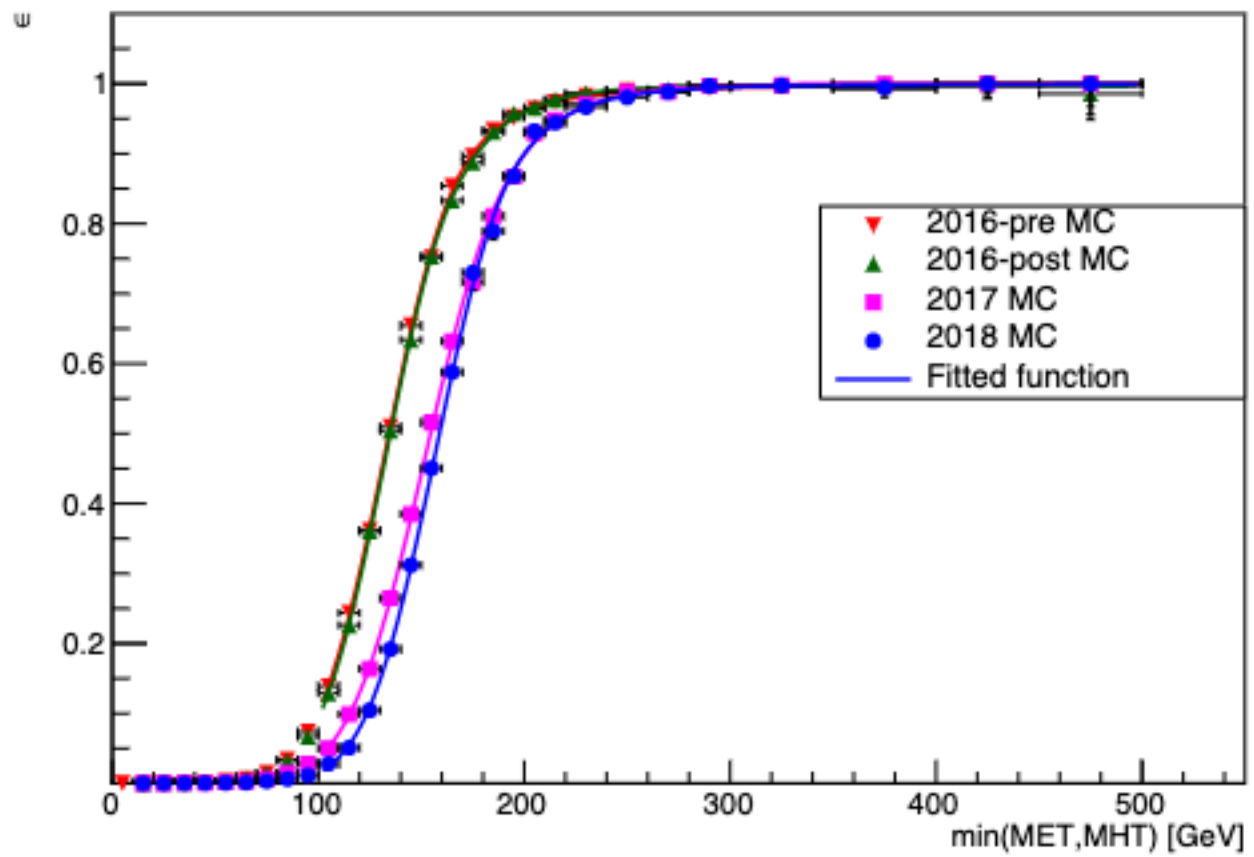


* Json files stored in: /eos/cms/store/group/phys_higgs/hbb/ntuples/VHH4b_Vleptonic_SF_UL

Leptonic Channel

- MET Trigger SF shown in the figure
- Documented in **AN - 21 - 209**, V-Leptonic SFs are good for any Vleptonic tagging UL analysis as long as our selections are used.
- Already compatible with VHbb and VHcc selections.



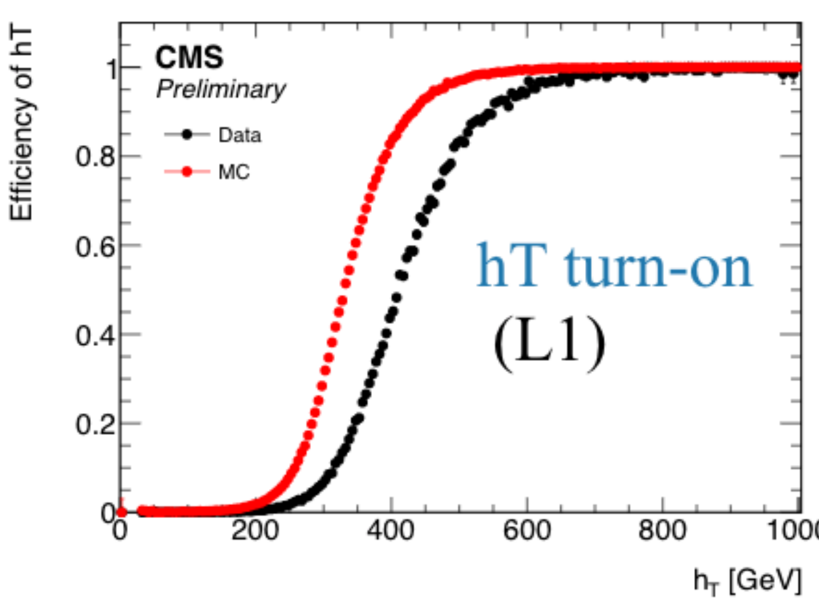


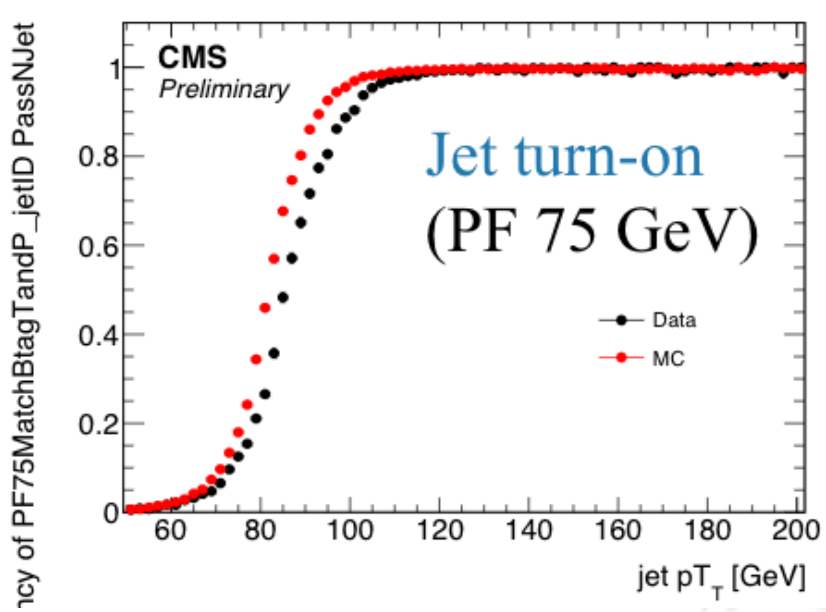
Efficiency measurements are done for all 4 channels

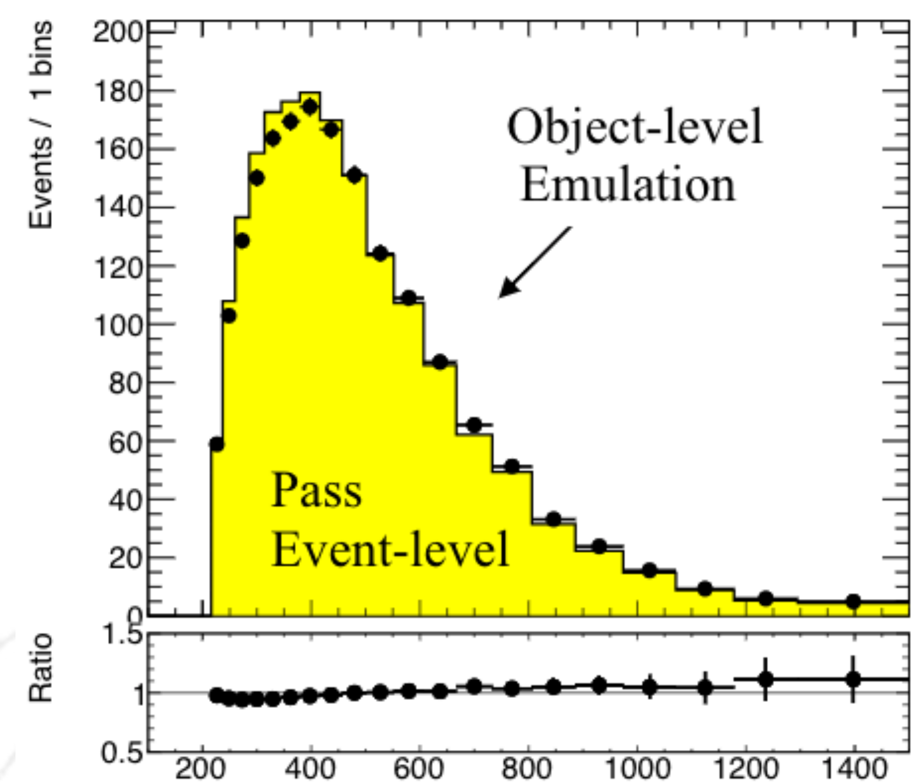
Hadronic Channel

★ Measure object level(and HT) turn-ons in ttbar eμ in both data and MC.

- Emulate event-level decision by emulating individual objects and HT turn-ons.
- Assume object turn-ons independent.
- Verify in MC closure test (right panel).
- Use turn-ons measured data as "calibrated" trigger decisions.







* Full trigger study results reported in the link.

Lepton Selection

- **MET**
  - Met Filters
  - MET_pt > 150GeV (250GeV for Boosted)
  - No jet (pt >30GeV and eta <10) with dPhi(Jet,MET) < $0.4 \times exp(\frac{200 - MET}{50}) + 0.07$ (suppress QCD only on Higgs candidate jets)

- **Single-Lepton**
- Muon: (1 Muon for W)
  - |Muon_eta| < 2.4
  - Muon_pt > 25GeV
  - Muon_pfRelIso04_all < max_rel_iso [0.06]
- WpT(lv) > 125GeV, dPhi(lep,MET) < 2
- Electron: (1 Electron for W)
  - |Electron_eta| < 2.5
  - Electron_pt > 32(17/18); Electron_pt > 28(16)
  - Electron_pfRelIso03_all < max_rel_iso [0.06]
  - Electon_mvaFall17V2Iso_WP90>0

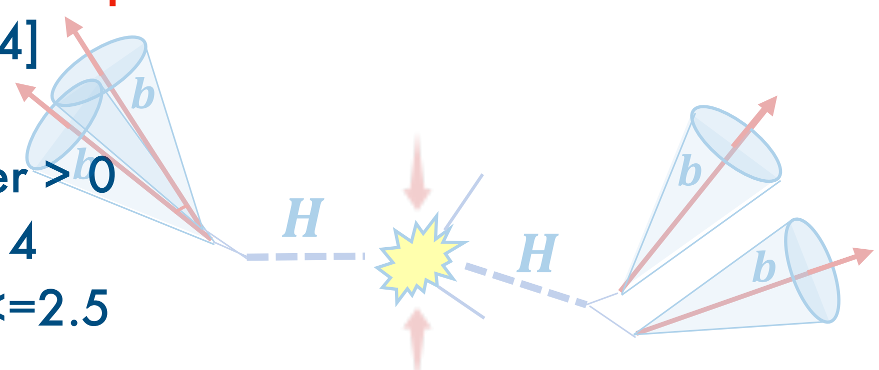
- **Double-Lepton**
- Muons: (2 Muons for Z)
  - |Muon_eta| < 2.4
  - Muon_pt > 20GeV
  - Muon_pfRelIso04_all < max_rel_iso [0.25]
- Electrons: (2 Electrons for Z)
  - |Electron_eta| < 2.5
  - Electron1_pt > 23GeV; Electron2_pt > 14GeV
  - Electron_pfRelIso03_all < max_rel_iso [0.15]
  - Electon_mvaFall17V2Iso_WP90>0
- Z mass window [80,100] GeV in DL channel
  - [80,100] GeV: Z mass region (research region)
  - Outside: TT Control region

Electron selections optimized in all 3 V-Leptonic channel

Jets Selection

V-Leptonic channels

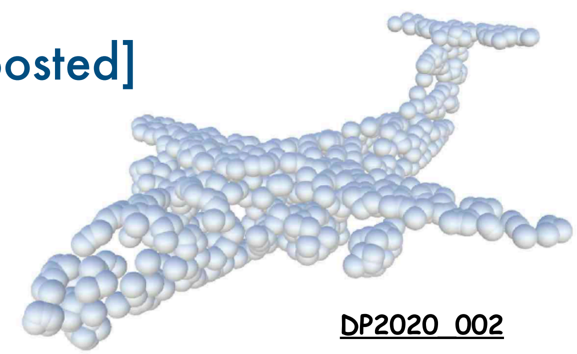
- GoodJets: [AK4]
  - Jet_Pt>50
  - Jet_lepFilter > 0
  - Jet_jetId > 4
  - |Jet_eta|<=2.5



- 4 leading DeepJet GoodJets [Resolved]
  - Using DeepJet b-tag [WP = Medium] (Loose in MET)
  - ptCut for all 4 jets
    - MET: pT>35GeV
    - 1-lep: pT>25GeV(j1-j3), pT>15GeV(j4)
    - 2-lep: pT>20GeV

DeepJet for AK4 jets and ParticleNet for AK8 jets

- 2 leading Dbb AK8 jets [Boosted]
  - |Jet_eta|<=2.5
  - Jet_Pt > 200GeV
  - Softdrop mass>50GeV
  - mHH > 300GeV

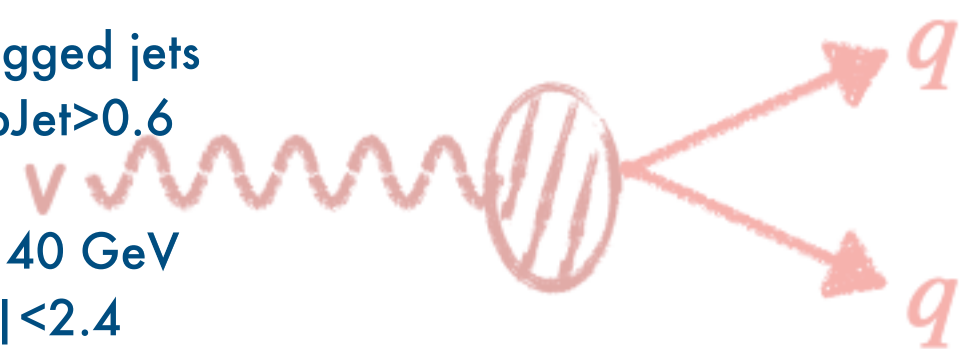


DP2020_002

$$D_{bb} = \frac{ParticleNetMD\_score(X \to b\bar{b})}{ParticleNetMD\_score(X \to b\bar{b}) + ParticleNetMD\_score(QCD)}$$

V-Hadronic channel

- >= 4 b-tagged jets
  - DeepJet>0.6
- >= 6 jets
  - pT > 40 GeV
  - |eta|<2.4
  - PUID medium WP



DeepJet for AK4 jets

- Higgs Candidate Jets: (4 leading DeepJet score jets)
  - Form 2 di-jets satisfying 45GeV < m_jj < 190GeV

$$360/m_{4j} - 0.5 < \Delta R(Leading\ S_T\ dijets) < max(1.5, 650/m_{4j} + 0.5)$$
$$235/m_{4j} < \Delta R(Sub\ leading\ S_T\ dijets) < max(1.5, 650/m_{4j} + 0.7)$$

- Vector Boson Candidate Jets:
  - Pairing all candidates and find 65GeV < m_{jj} < 105GeV

Jet selections are optimized in all 4 channels

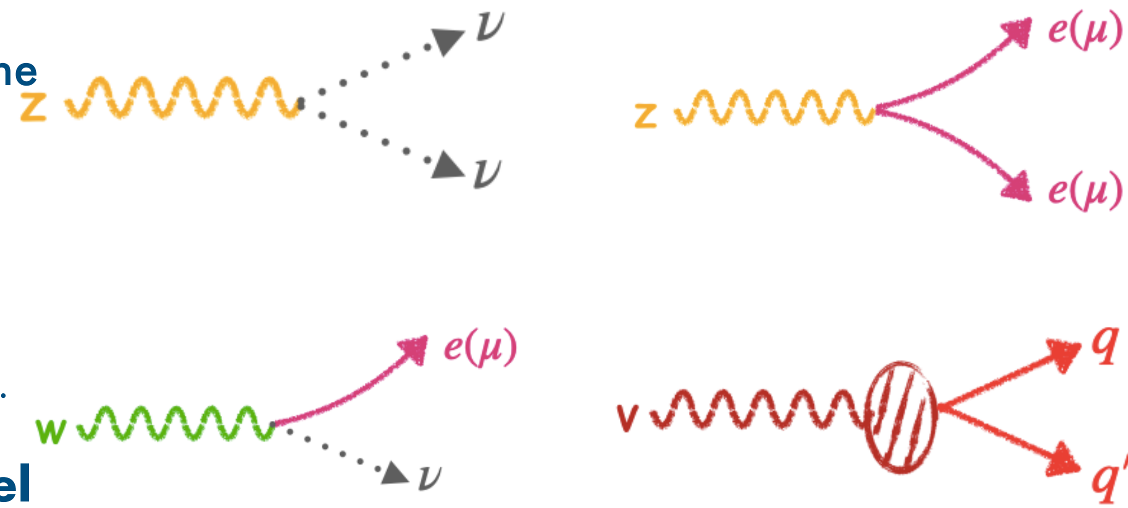
★ Reconstruction of H-bosons and V-bosons

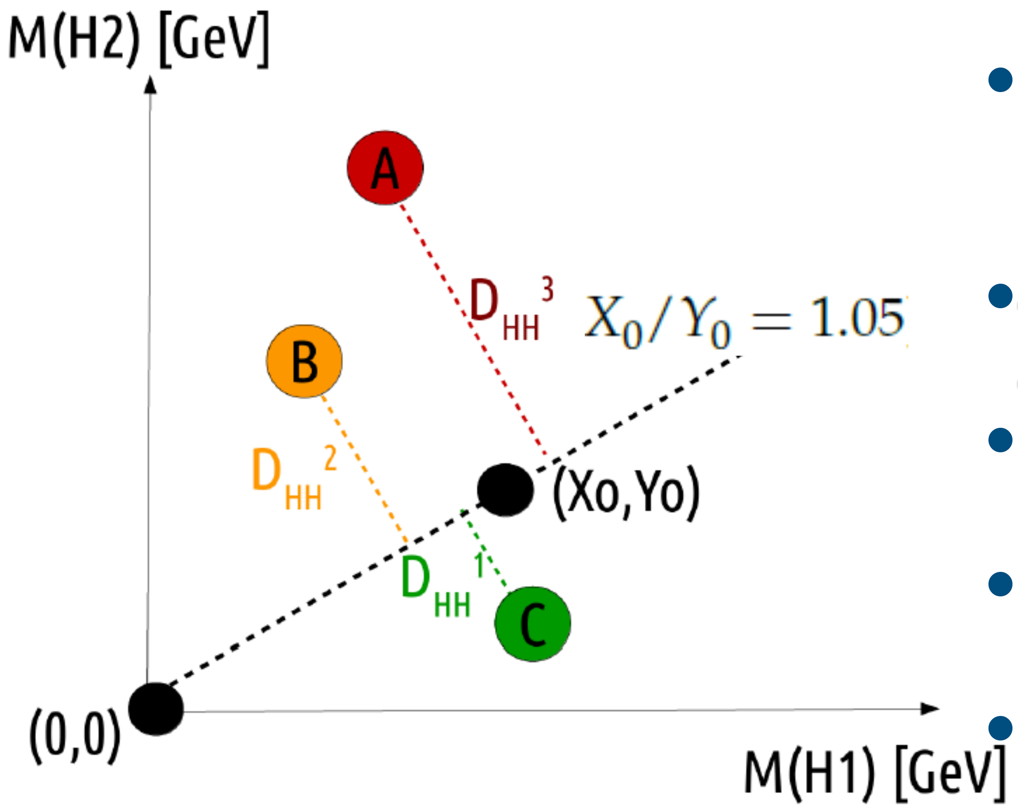
- With all the combination of the **Higgs** Candidate Jets, we choose the one that can minimize the variable **D_{HH}**.

$$D_{HH} = \frac{|m_{H1} - c \times m_{H2}|}{\sqrt{1 + c^2}}$$

where H_1 and H_2 are the leading and sub-leading pT(HT in FH) Higgs candidates, respectively.

- C = 1.05 for **leptonic channels** and C = 1.04 for **hadronic channel**





- For those **V-boson** candidate leptons(MET) and jets that pass the object selections.
- In DL, FH channel, V-boson candidates inside a V mass window,
- In SL channel, V-boson candidates with pT>125GeV, dPhi(lep, MET)<2,
- In boosted topo, dR(Jet,V)>0.8 in MET and dR(Jet,lep)>0.8 in SL.
- Then the candidate with leading pT will be chosen.

Analysis Strategies

Variables	DL	SL	MET	FH
pT(H2)/pT(H1)	✓	✓	✓	✓
mass(HH)	✓	✓	✓	✓
dR(H1,H2)	✓	✓	✓	✓
pT(H1)	✓	✓	✓	✓
pT(H2)		✓	✓	✓
pT(V)	✓	✓	✓	✓
pT(HH)		✓	✓	✓
mass(H1)		✓	✓	✓
mass(H2)		✓	✓	✓
E(H1)		✓	✓	✓
E(H2)		✓	✓	✓
E(HH)		✓	✓	✓
eta(HH)		✓	✓	✓
deta(H1, H2)		✓	✓	✓
dPhi(H1, H2)		✓	✓	✓
dPhi(V, H2)	✓	✓	✓	✓
pT(L1)	✓			
dEta(L1,L2)	✓			
dPhi(L1,L2)	✓			
pT(L1)/mass(V)	✓			
pT(L2)/pT(L1)	✓			
dR(H2b1,H2b2)	✓			
dR(H1b1,H1b2)	✓			

Variables used in the Kl BDTs

Variables	DL	SL	MET
mass(V)	✓		
mass(H1)	✓	✓	✓
HT(lljjjj)/(MET)	✓		✓
dPhi(V, H1)	✓		
dR(H1,H2)	✓		
pT(j No.4 btag)	✓		
deta(H1, H2)	✓		
pT(V)/pT(HH)	✓		
dEta(L1,L2)	✓		
mass(HH)	✓	✓	✓
E(H1)	✓		
pT(j No.3 btag)	✓		
pT(L1)/mass(V)	✓		
dPhi(V, HH)	✓		
Eta(H1)			✓

Variables used in the SvB BDTs

Variables	DL	SL	MET
pT(H1)	✓	✓	✓
pT(V)	✓	✓	✓
pT(HH)	✓	✓	✓
mass(H2)		✓	✓
B-tag(H1j1)		✓	✓
B-tag(H1j2)		✓	✓
B-tag(H2j1)		✓	✓
B-tag(H2j2)		✓	✓
pT(H2)		✓	✓
Phi(H1)		✓	✓
Phi(H2)		✓	✓
Phi(V)		✓	✓
Year		✓	
nJets_pt25			✓
Eta(H2)			✓

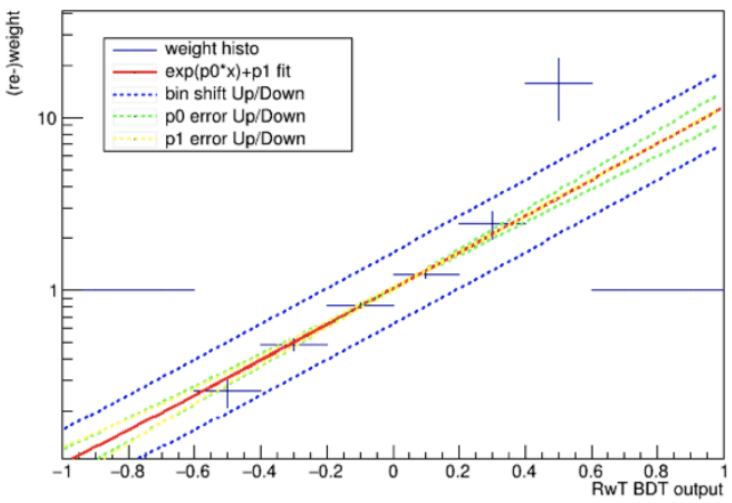
Variables used in the SvB BDTs

Variables	DL
pT, Eta, Phi, M of all jets and Di-jets	✓
pT, Eta, Phi, M of all Di-jets	✓
pT, Eta, M of all Quad-jets	✓
Year	✓
xW	✓
xbW	✓

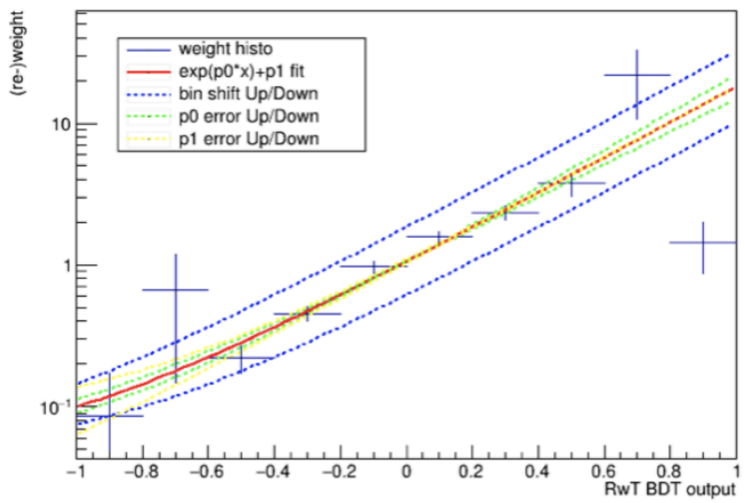
Variables used in the SvB NN

Coupling Cats BDT and SvB Classifiers are optimized in all channels

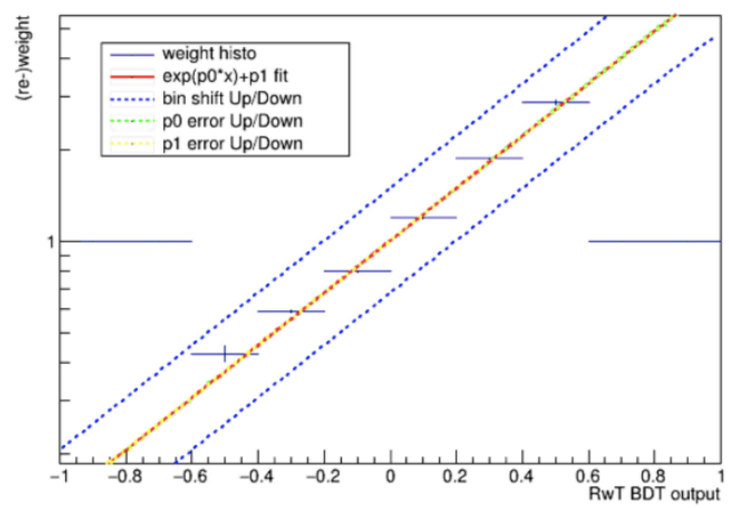
Background modeling



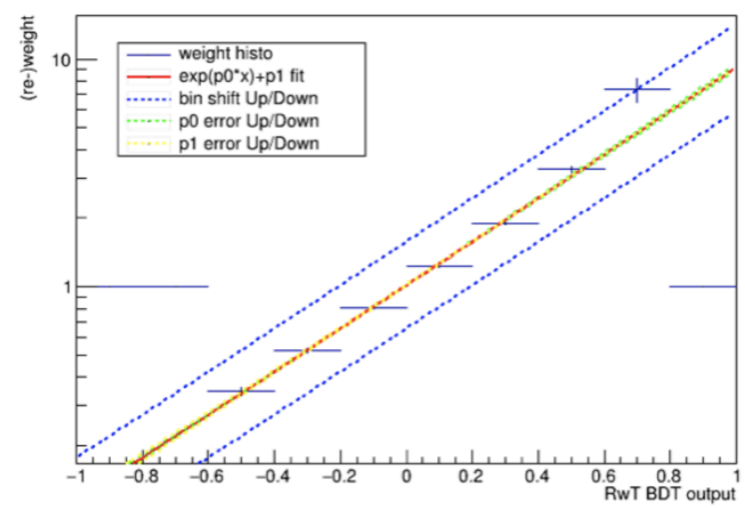
TT_2TO3_3b



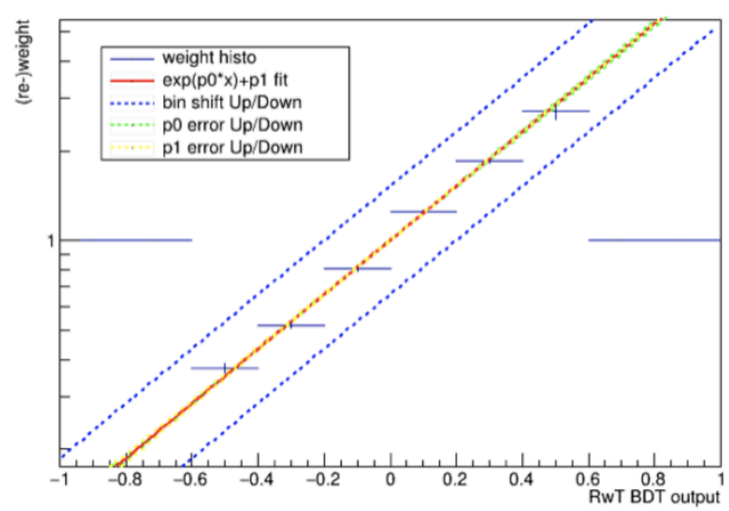
TT_2TO4_4b

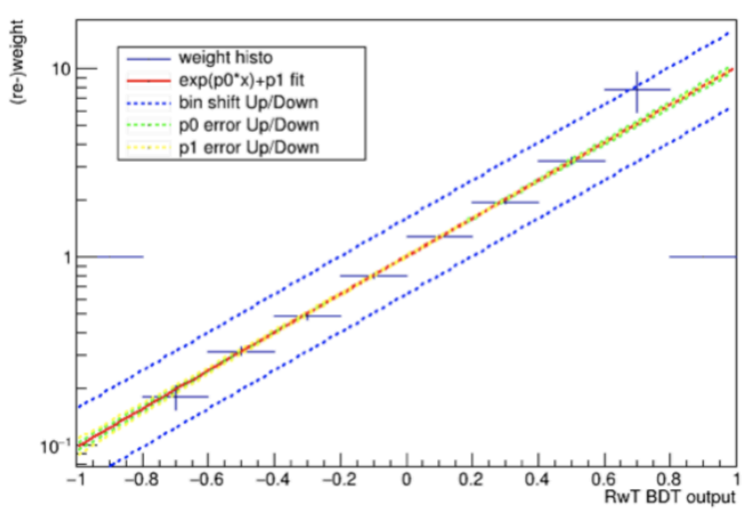


TTBB_2TO3_3b

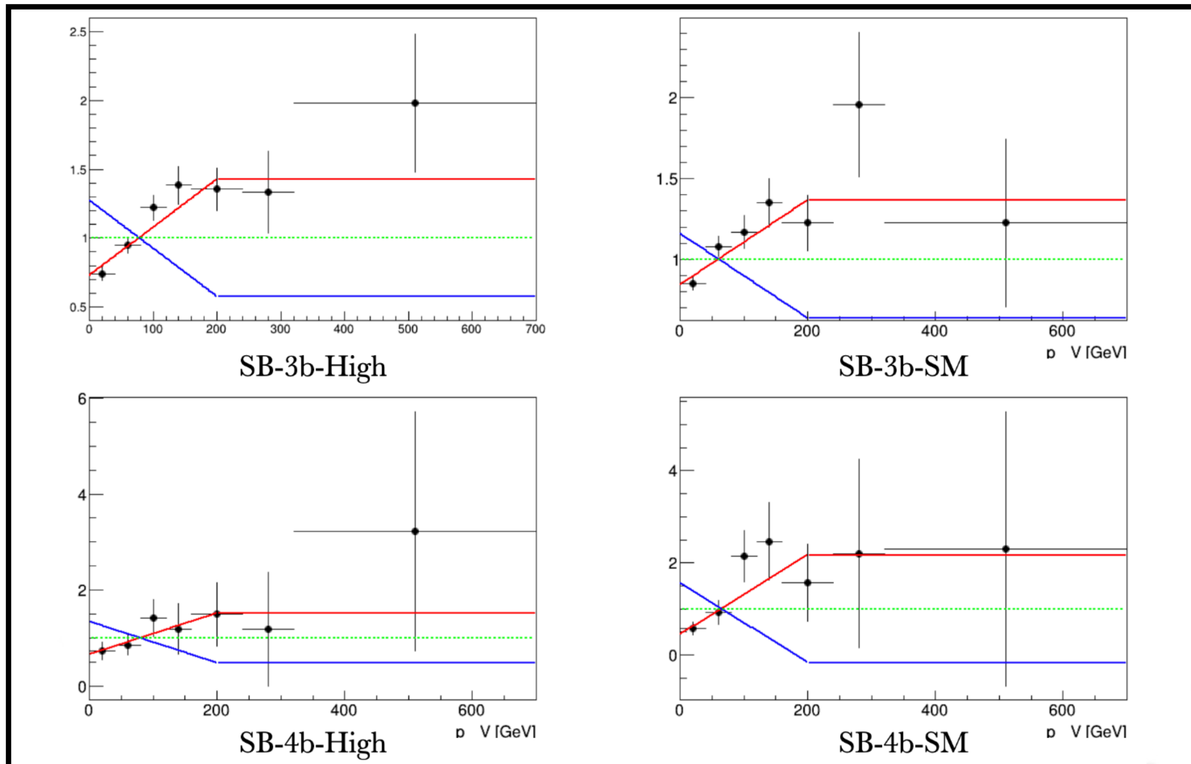


TTBB_2TO4_4b





- The (re-)weight over different processes are fitted with exponential functions.
- We calculate the uncertainty by shift the histogram left and right for one bin width.(the blue dashed lines)
- Re-weighting has been validated in control regions.



- The residual difference between data and re-weighted MC observed in SB/CR over Z_pt are covered by introducing a modeling uncertainty on DY+Jets sample.

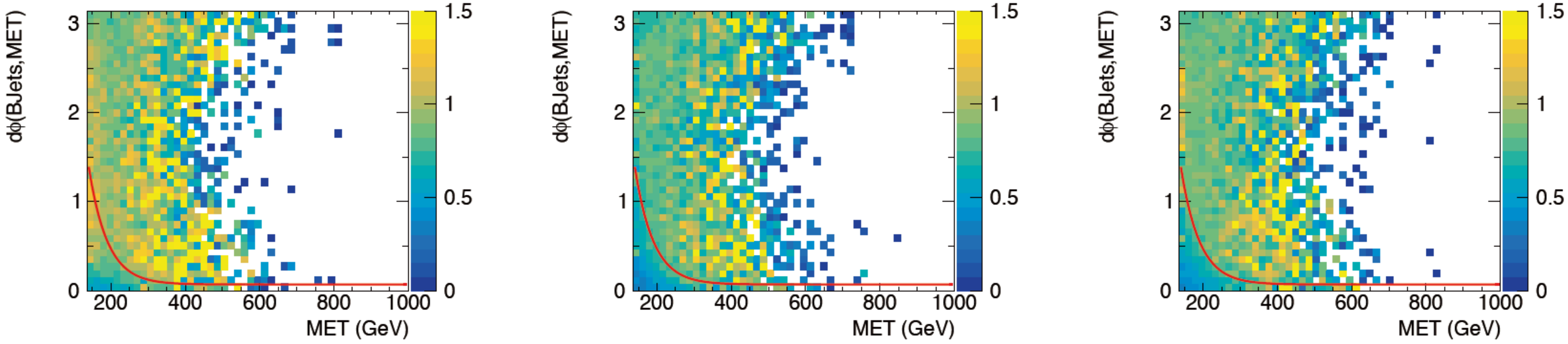


Figure 74: $abs(d\phi(Jet, p_T^{miss}))$ vs. p_T^{miss} for $\frac{MC}{Data}$. Plots from left to right are 2016,2017,2018. Selected events are in sideband, 3B or 4B. Excessive MultiJets events cluster in the bottom left corner. The red line represents $abs(d\phi(Jet, p_T^{miss})) = 0.4e^{\frac{200-p_T^{miss}}{50}} + 0.07$ and events below this line are removed.

VHH Analysis backup

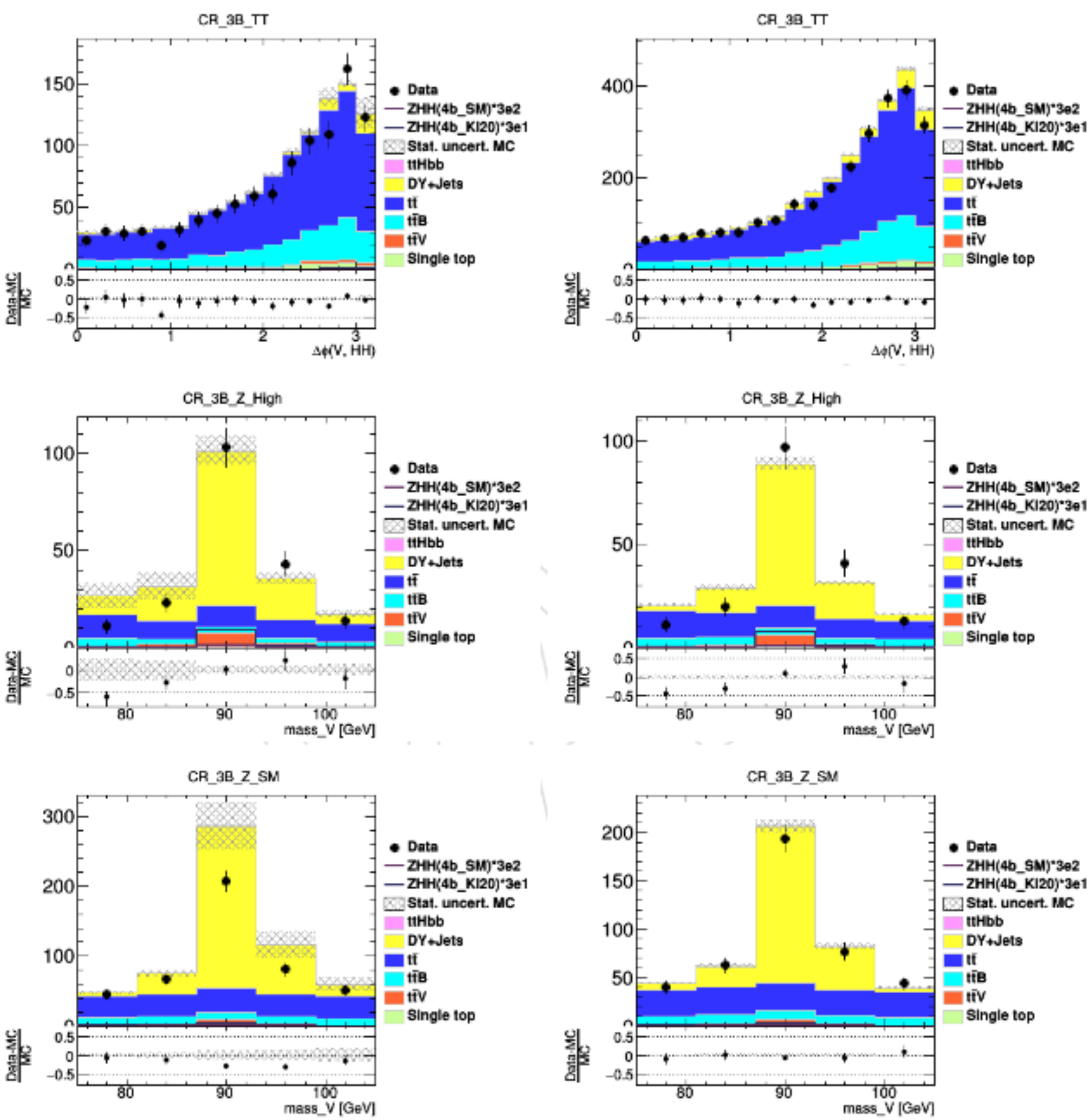


Figure 27: Data vs. MC background validation plots from 2018 in different regions. (Left: Original MC estimation; Right: Dominant background estimated using re-weighted 2-btagged MC events.)

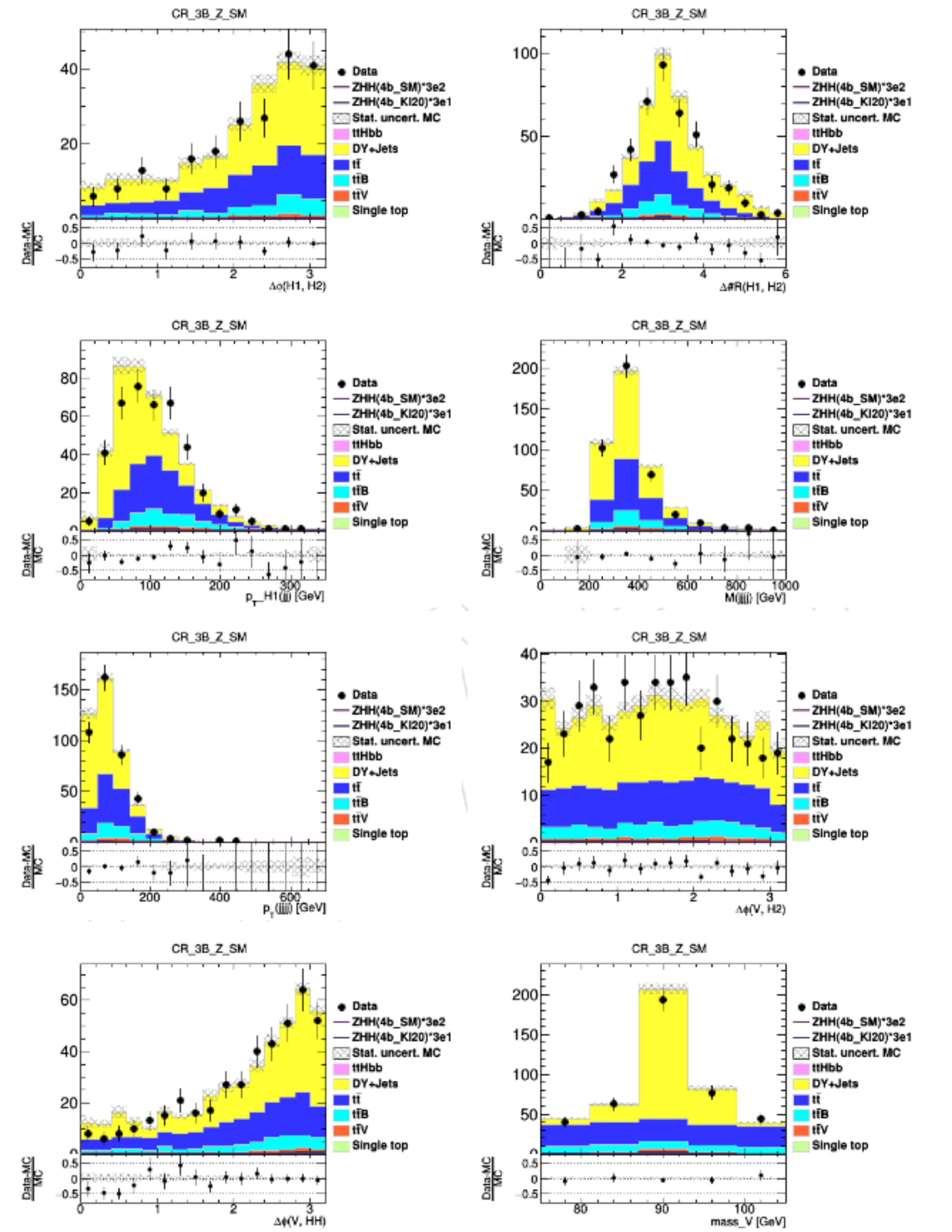
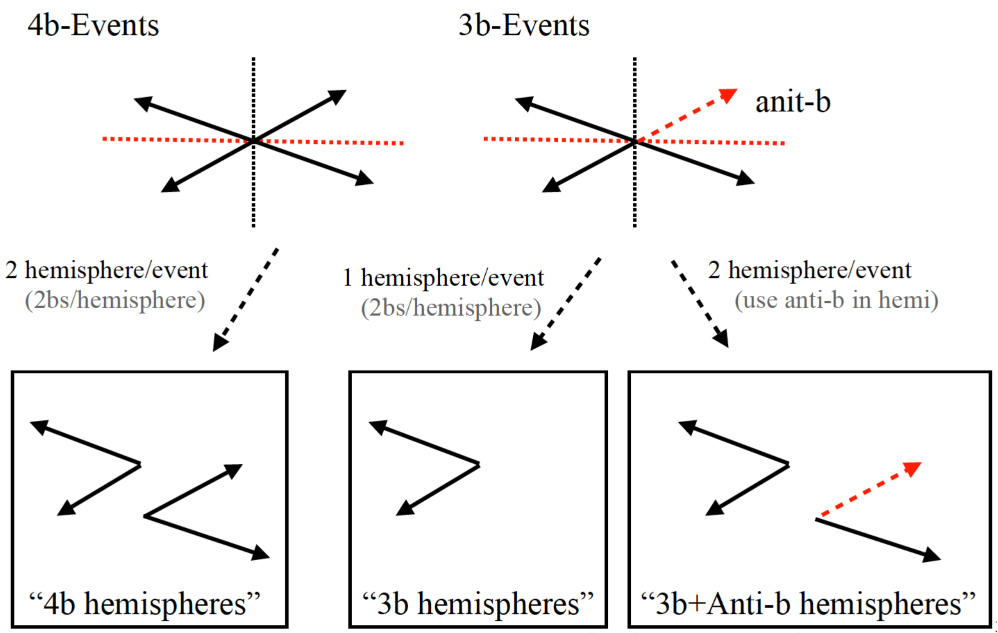
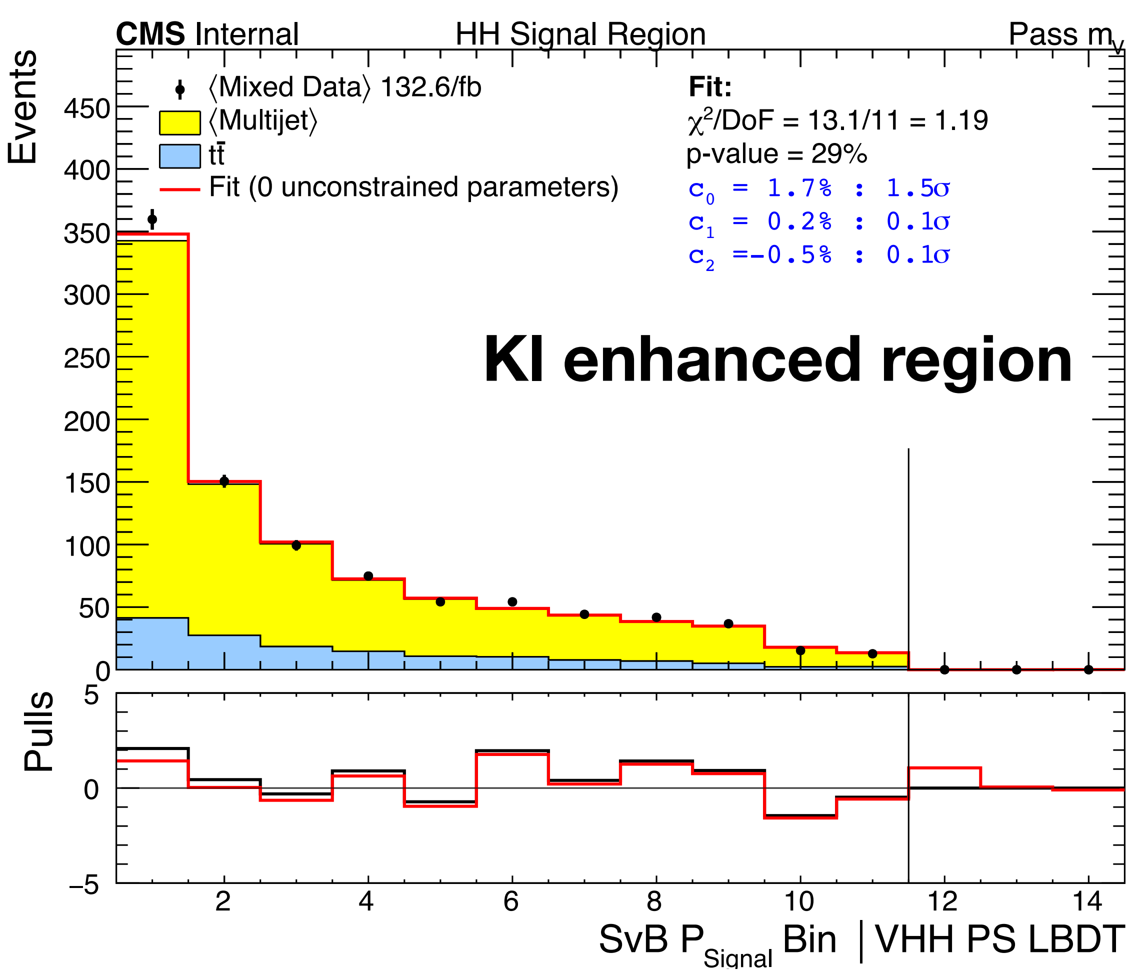


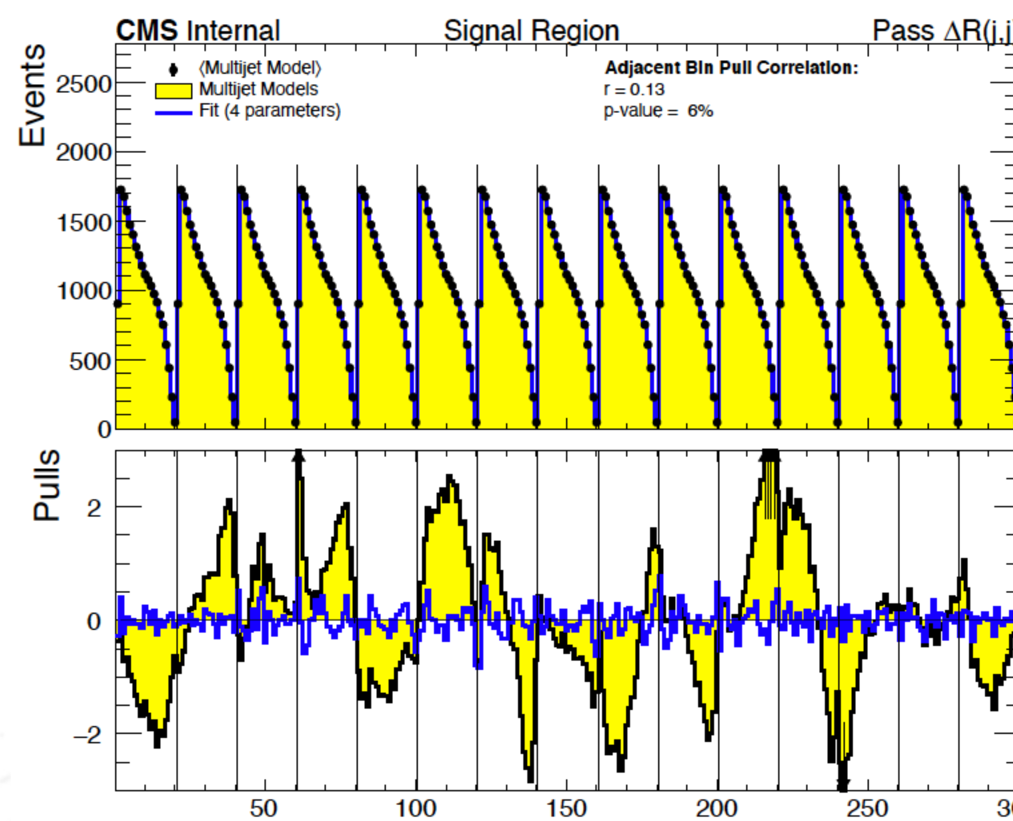
Figure 29: Data vs. MC background validation plots from 2018 in Z mass window and SM κ_λ region.



Six Options:

Library \ Starting Events	4b	3b
4b-hemispheres	*Too much signal contamination*	*Will use for closure*
3b-hemis	*Too much signal contamination*	*Option*
3bA-hemis	*Too much signal contamination*	*Option (Not investigated)*





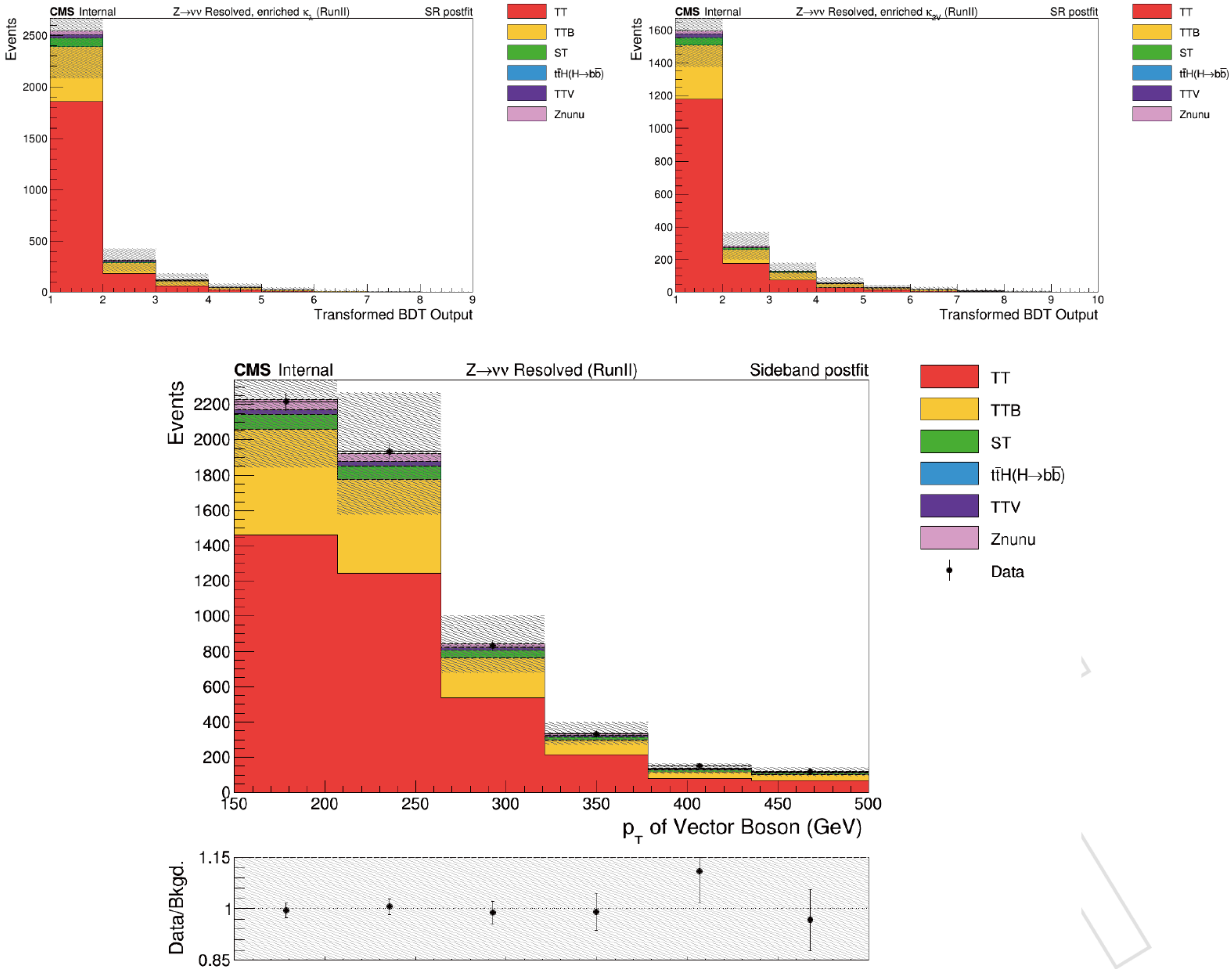


Figure 121: Postfit plots for $Z\nu\nu$ channel. The bottom row are the sideband plots, the upper row are the blinded signal region plots.

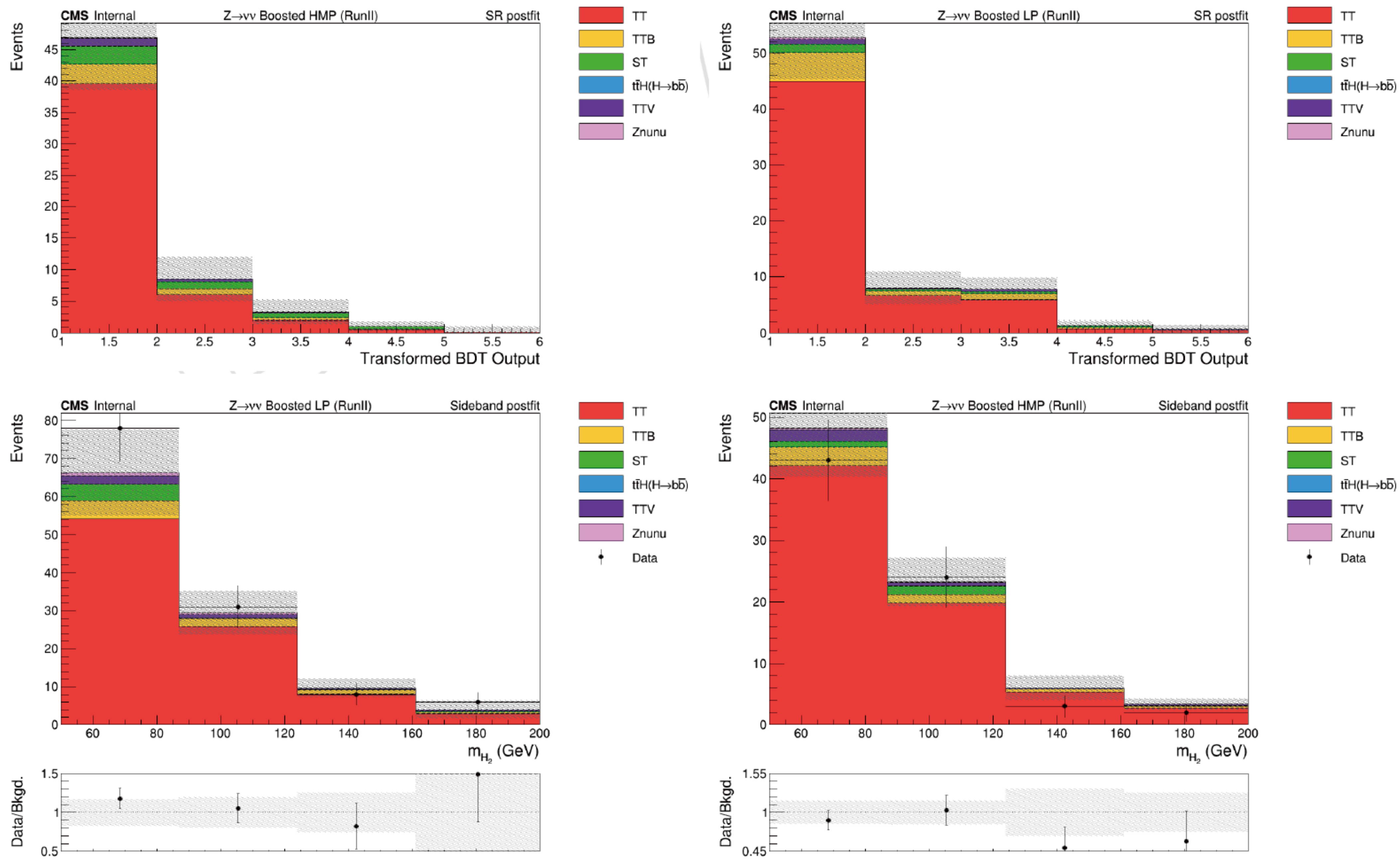


Figure 122: Postfit plots for $Z\nu\nu$ channel. The bottom row are the sideband plots, the upper row are the blinded signal region plots.

VHH Analysis backup

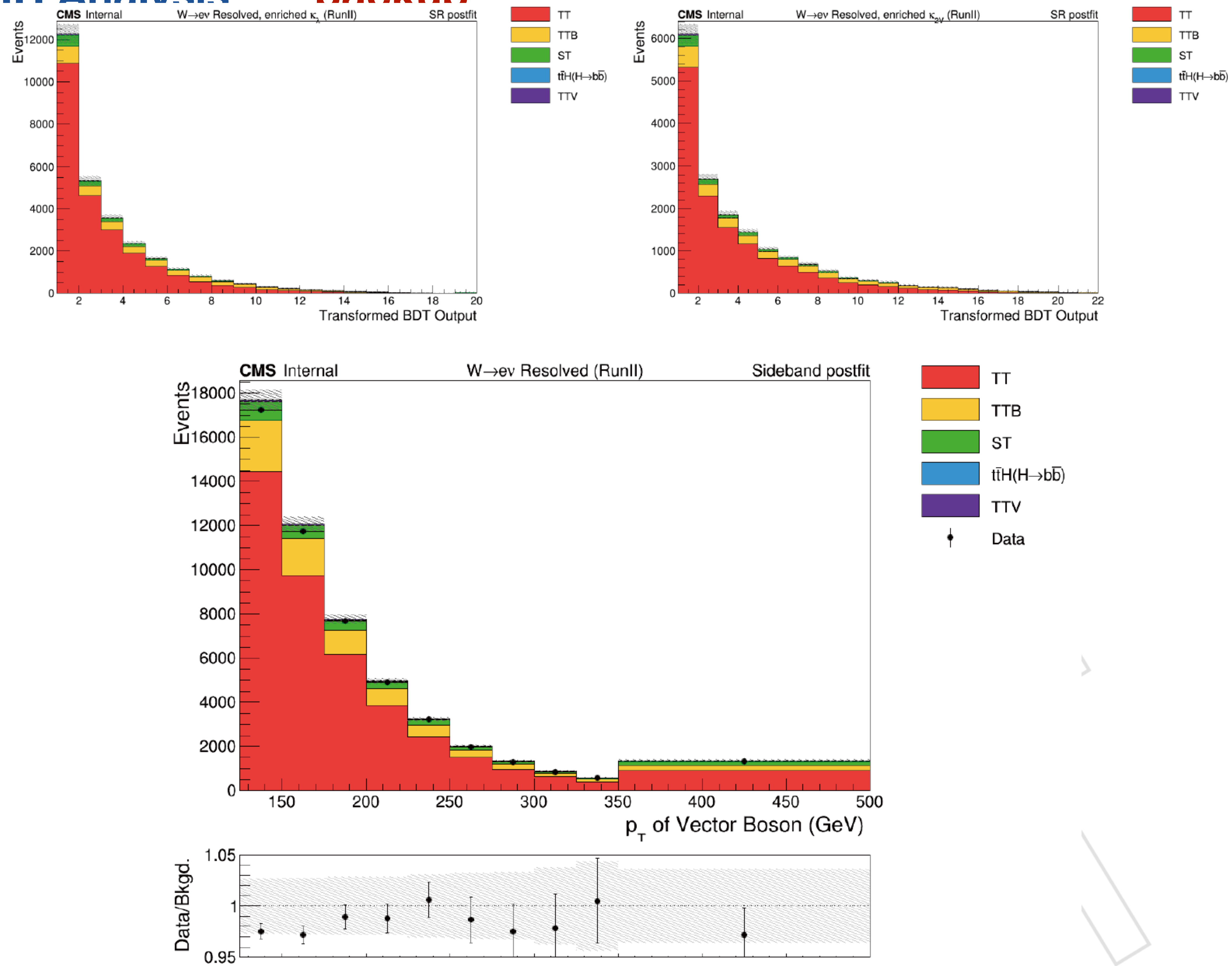


Figure 123: Postfit plots for *Wev* channel. The bottom row are the sideband plots, the upper row are the blinded signal region plots.

VHH Analysis backup

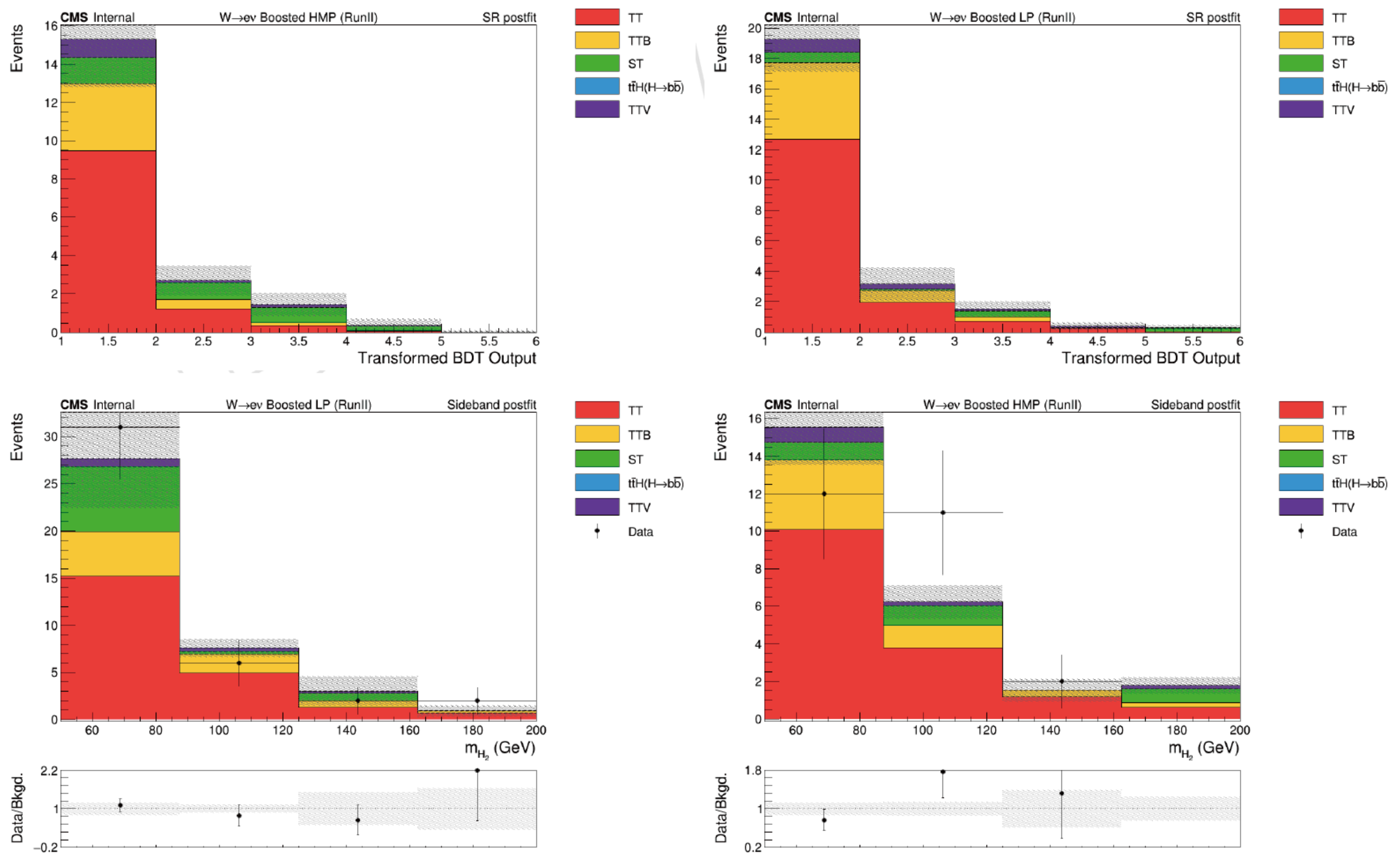


Figure 124: Postfit plots for *Wev* channel. The bottom row are the sideband plots, the upper row are the blinded signal region plots.

VHH Analysis backup

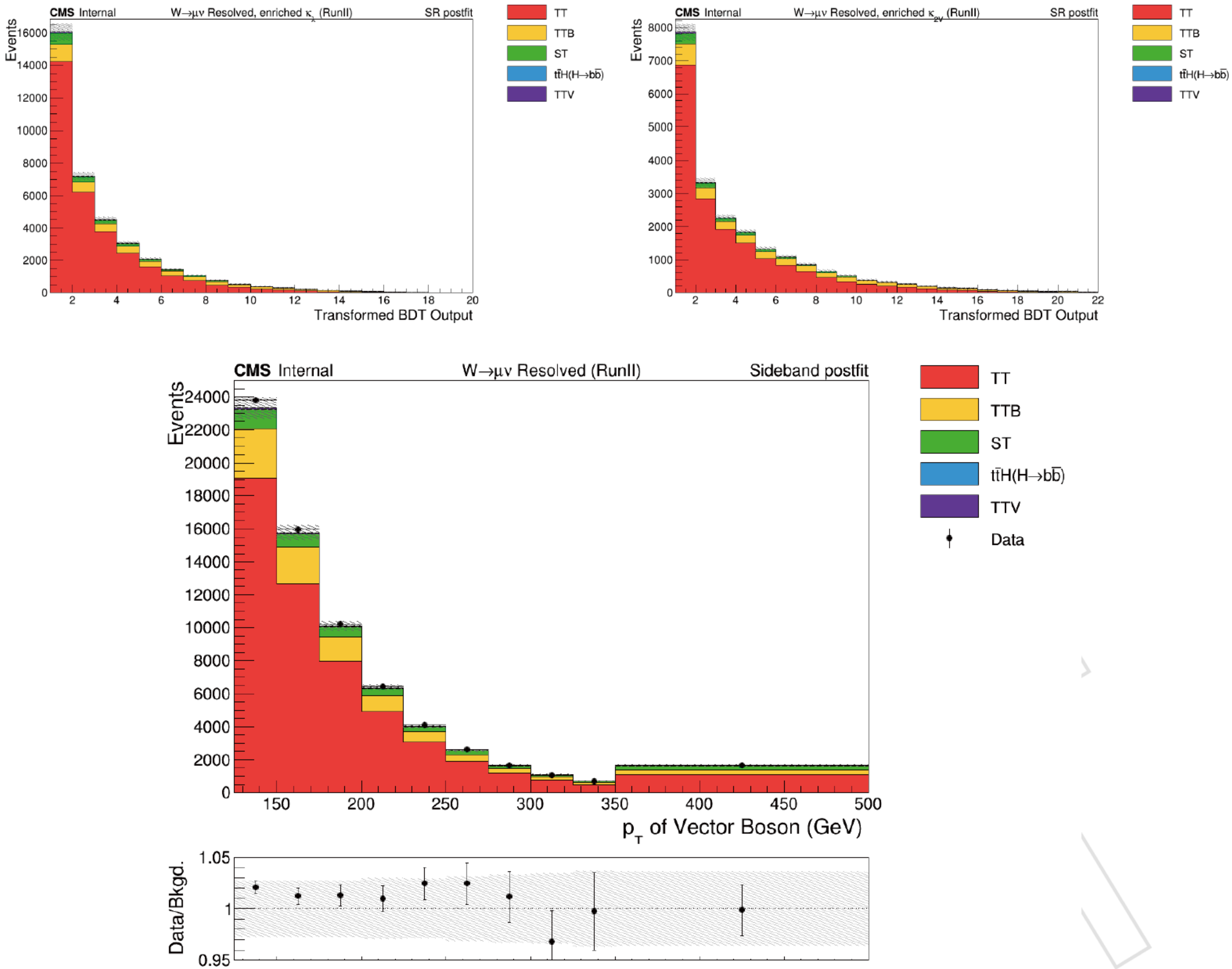


Figure 125: Postfit plots for $W\mu\nu$ channel. The bottom row are the sideband plots, the upper row are the blinded signal region plots.

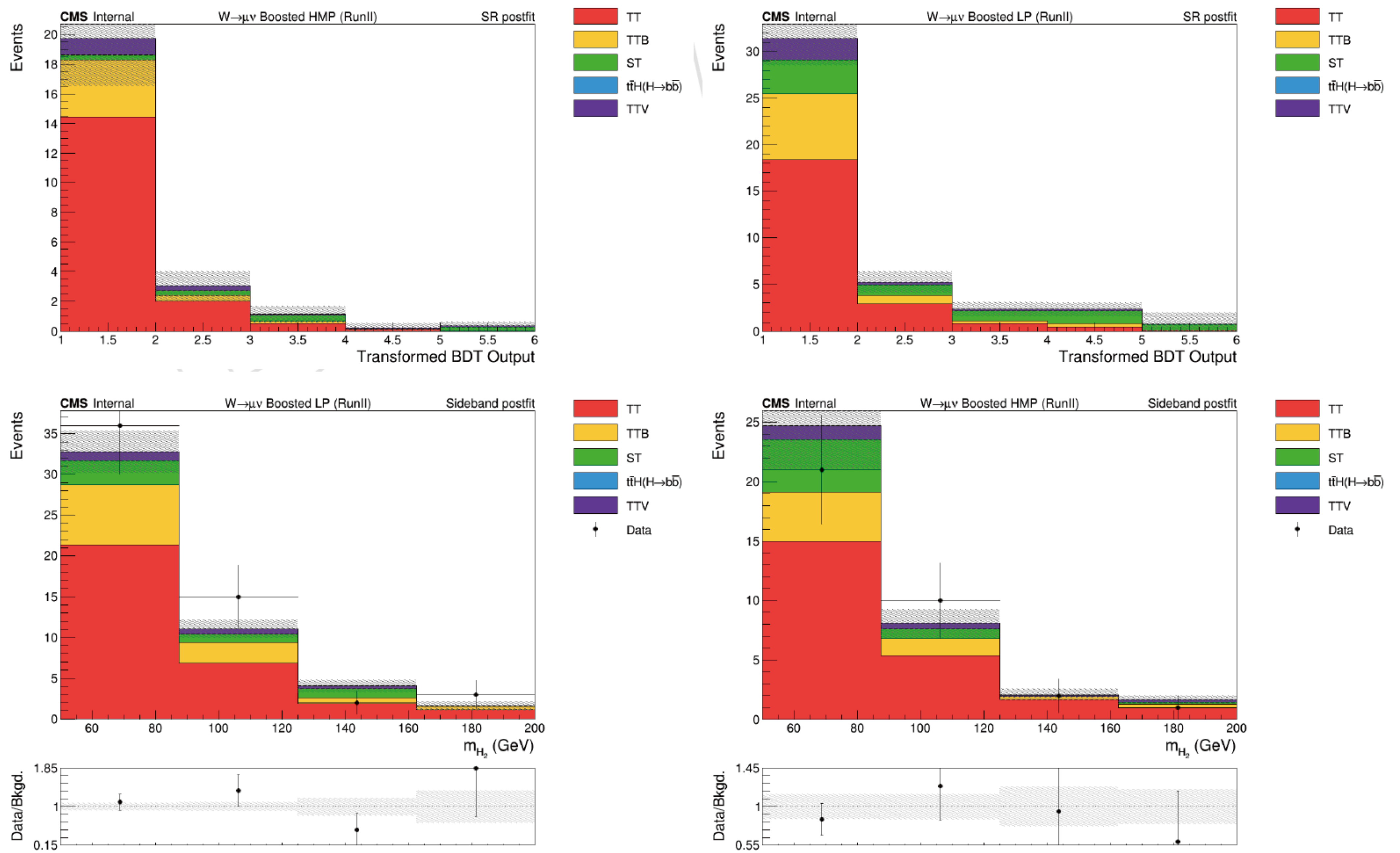


Figure 126: Postfit plots for $W\mu\nu$ channel. The bottom row are the sideband plots, the upper row are the blinded signal region plots.

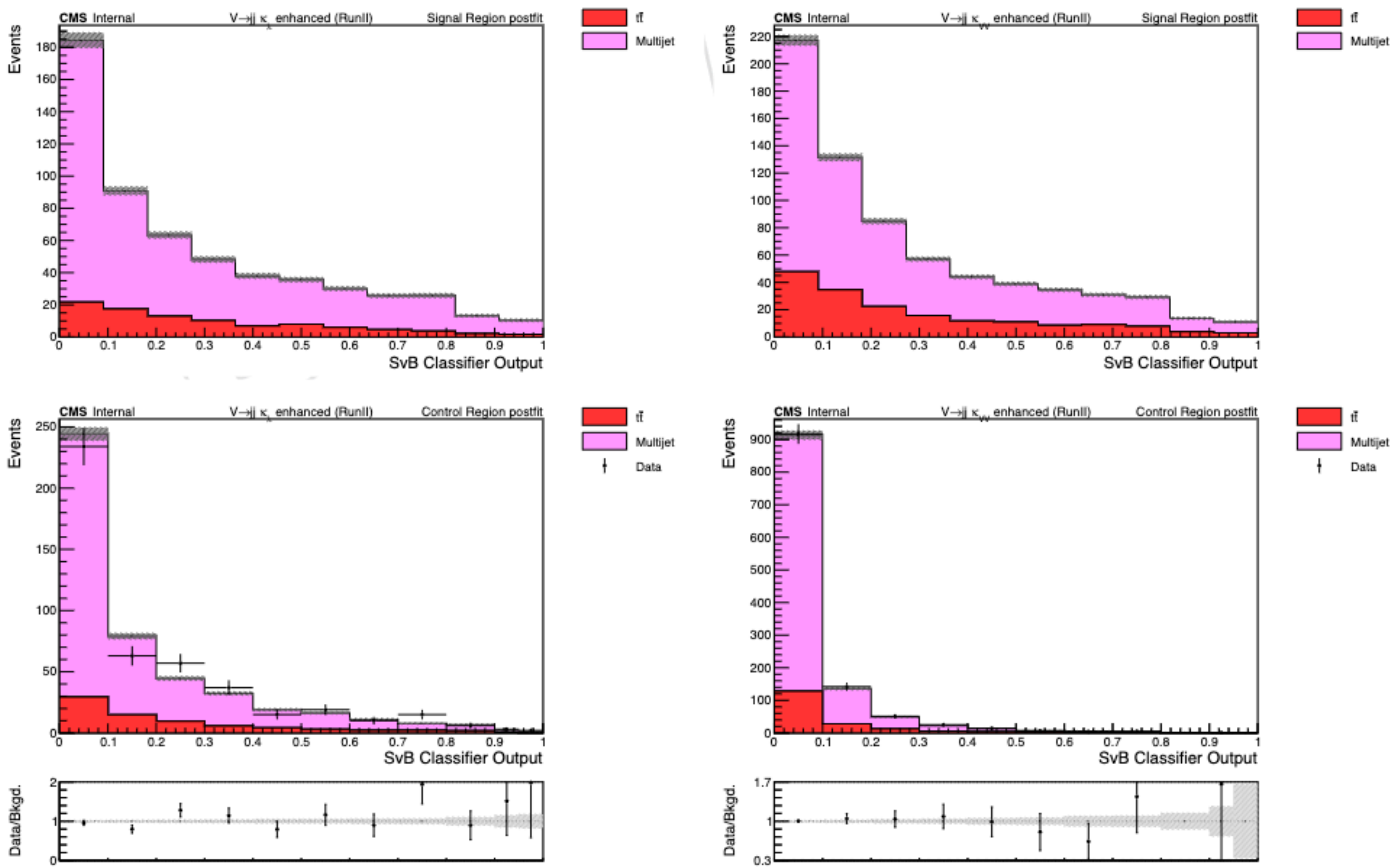


Figure 129: Postfit plots for Full-hadronic channel in signal region (blinded) and control region.

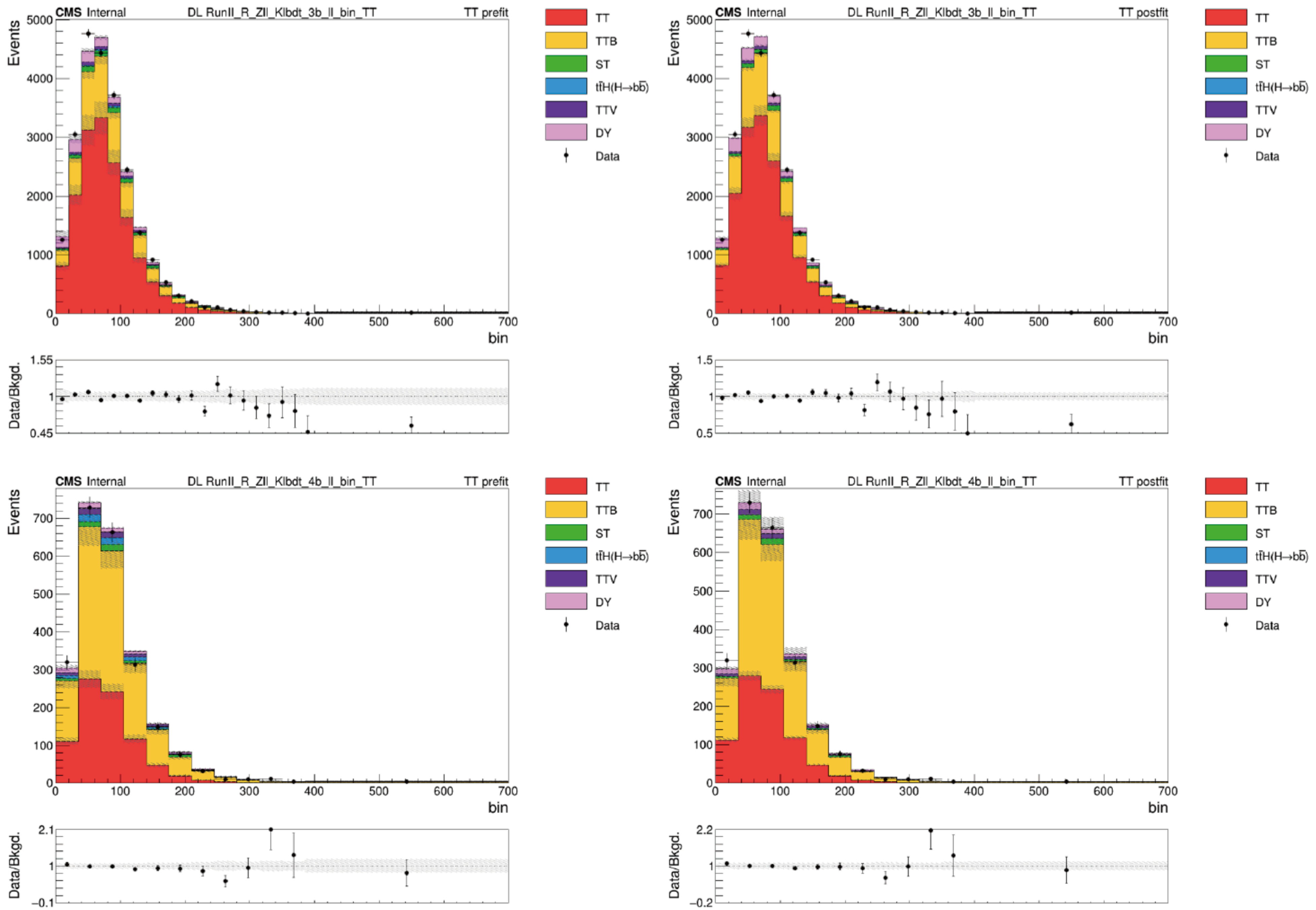
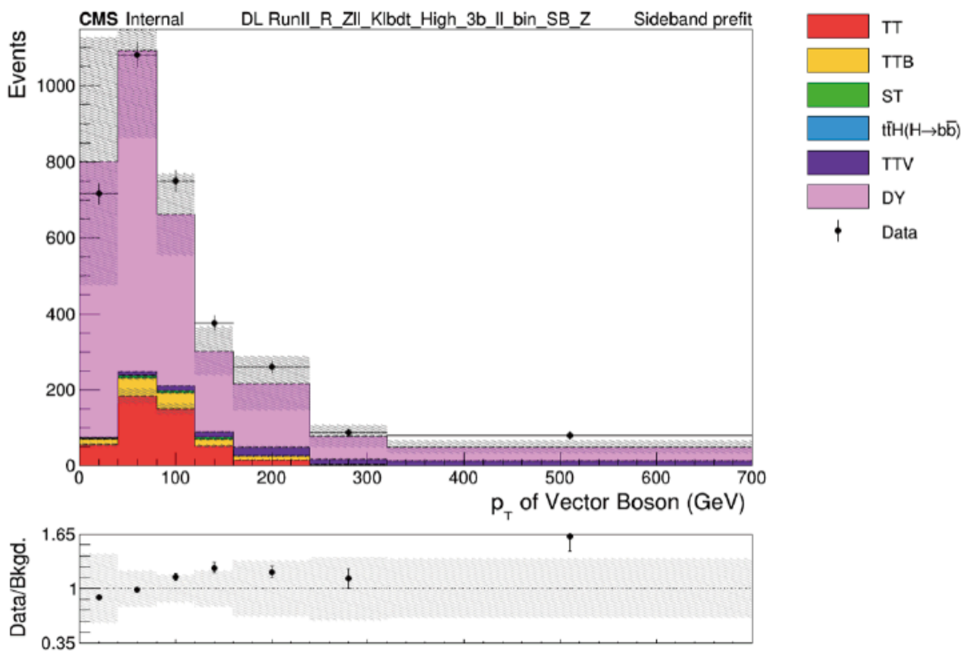
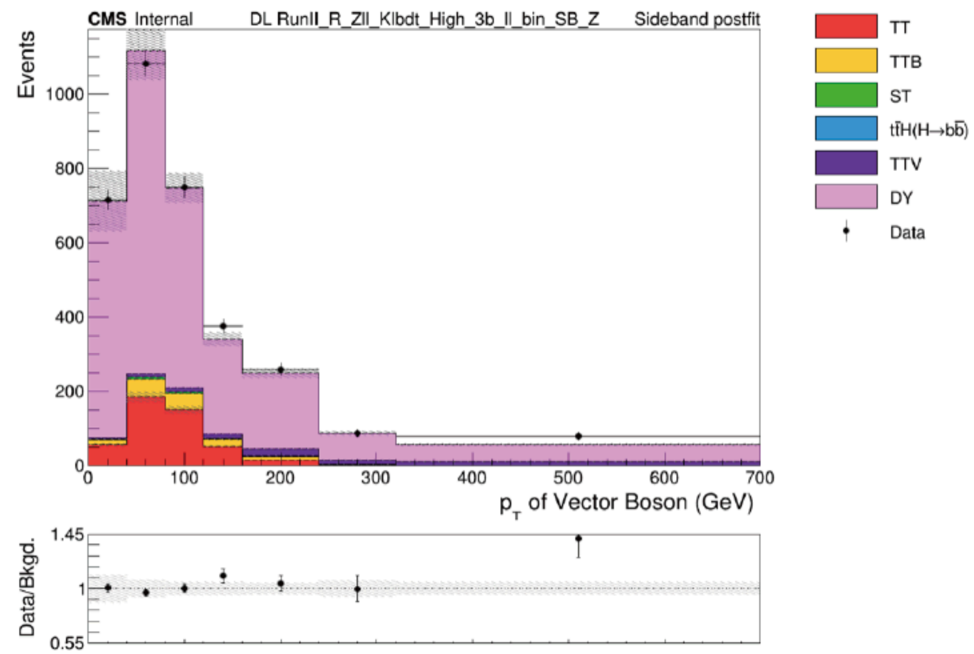
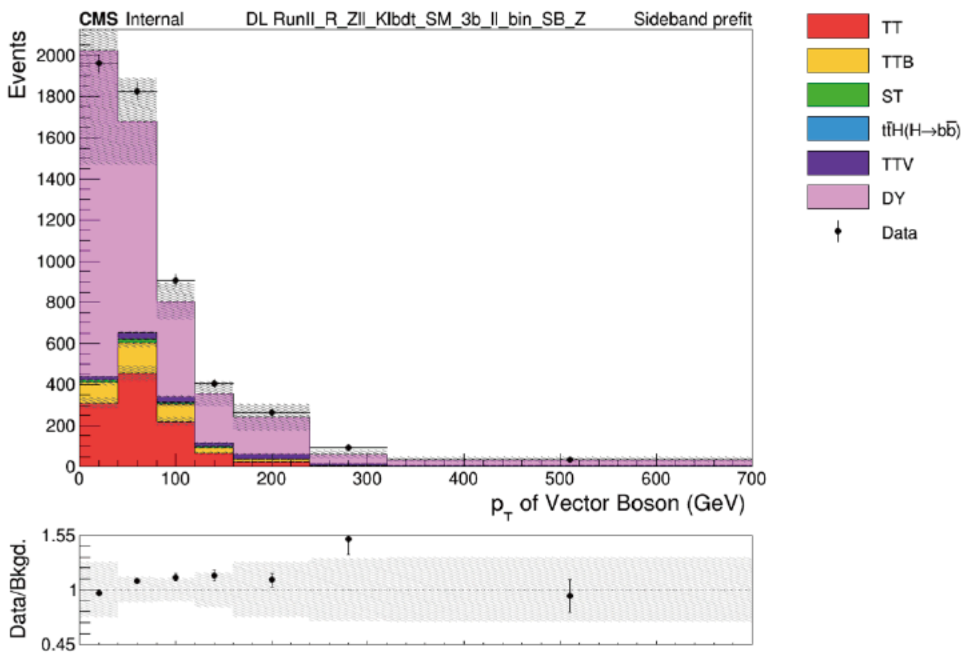
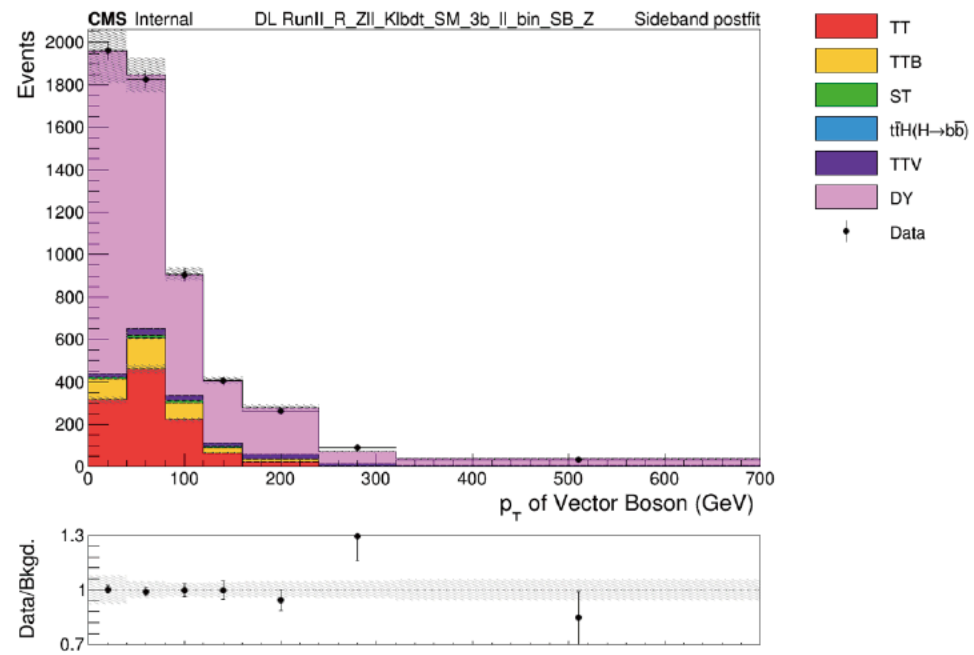


Figure 44: Pre-fit and Post-fit figures in TT control Region.


CMS Internal
DL RunII_R_Zll_Klbdt_High_3b_ll_bin_SB_Z
Sideband prefit
Events
p_T of Vector Boson (GeV)
Data/Bkgd.
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY
Data


CMS Internal
DL RunII_R_Zll_Klbdt_High_3b_ll_bin_SB_Z
Sideband postfit
Events
p_T of Vector Boson (GeV)
Data/Bkgd.
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY
Data


CMS Internal
DL RunII_R_Zll_Klbdt_SM_3b_ll_bin_SB_Z
Sideband prefit
Events
p_T of Vector Boson (GeV)
Data/Bkgd.
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY
Data


CMS Internal
DL RunII_R_Zll_Klbdt_SM_3b_ll_bin_SB_Z
Sideband postfit
Events
p_T of Vector Boson (GeV)
Data/Bkgd.
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY
Data

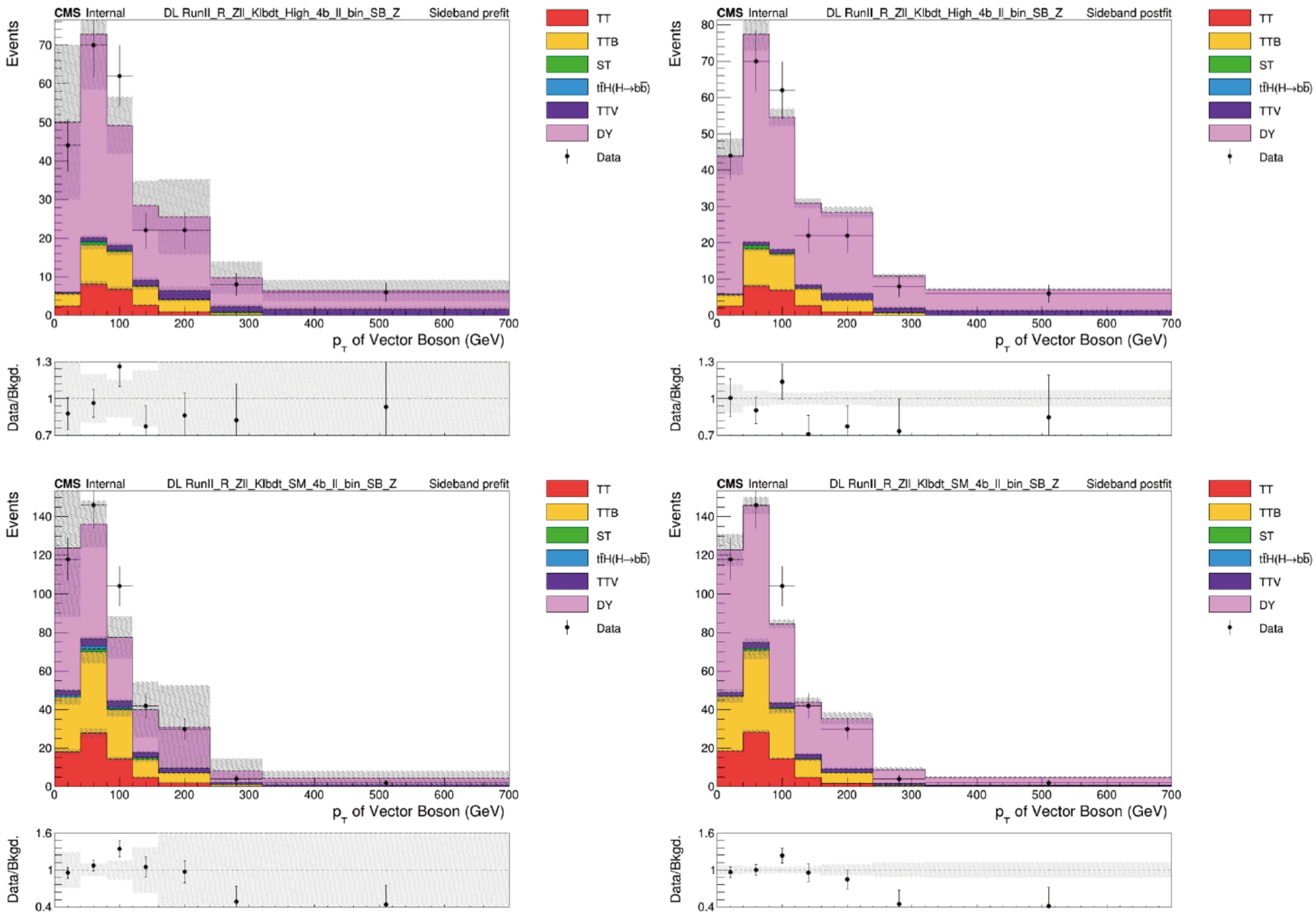
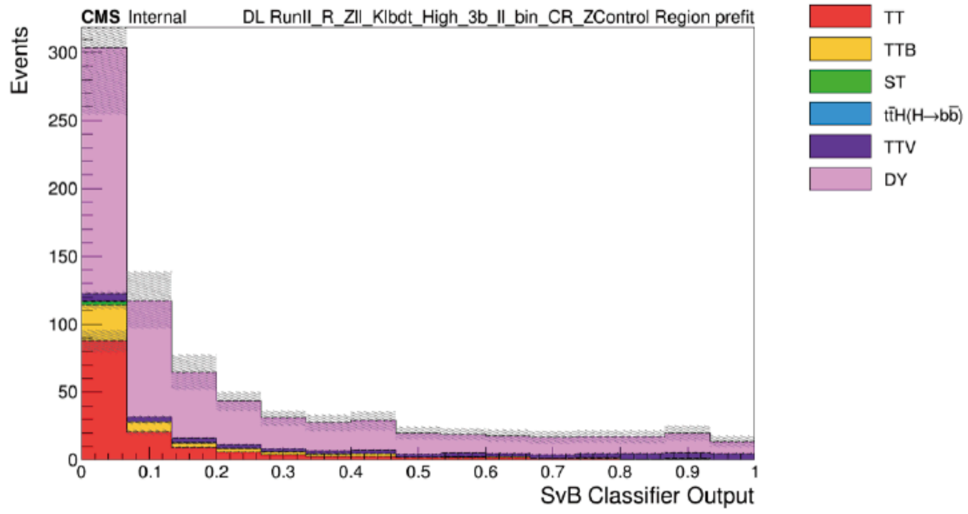
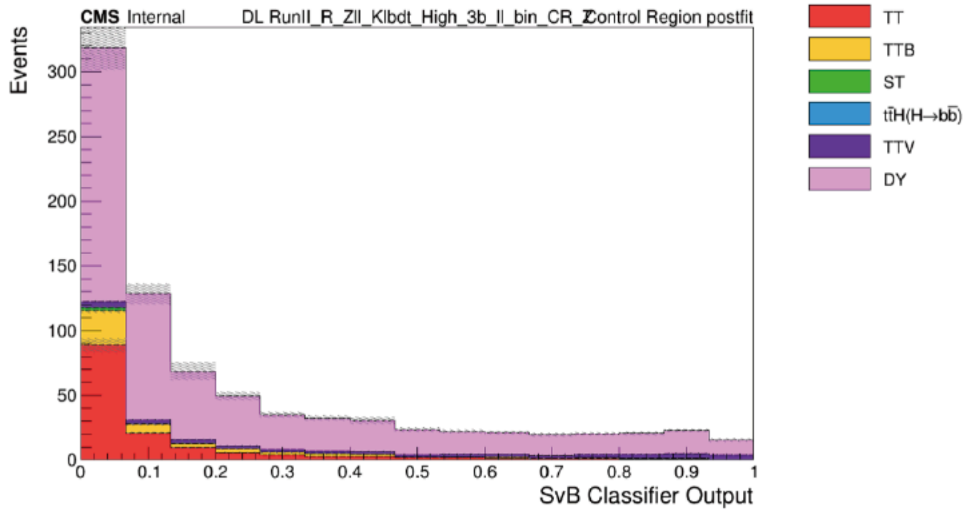
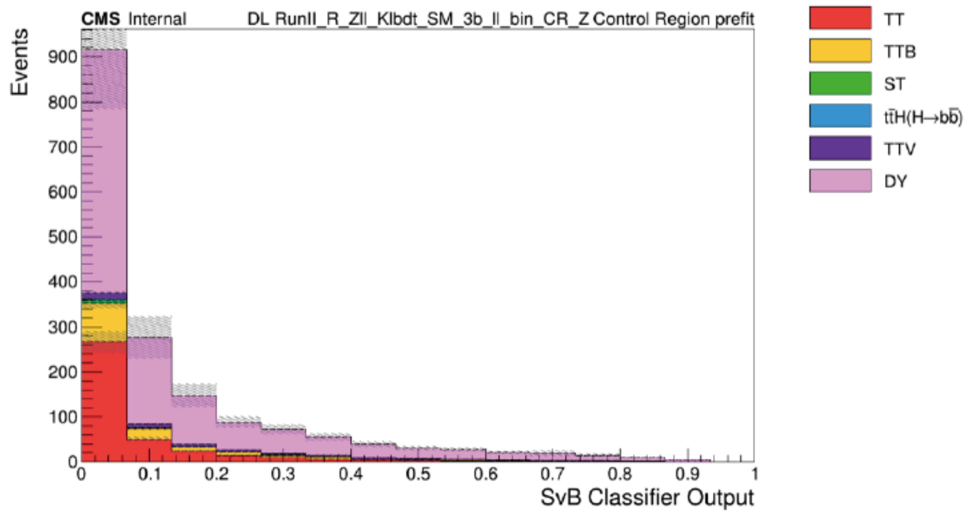
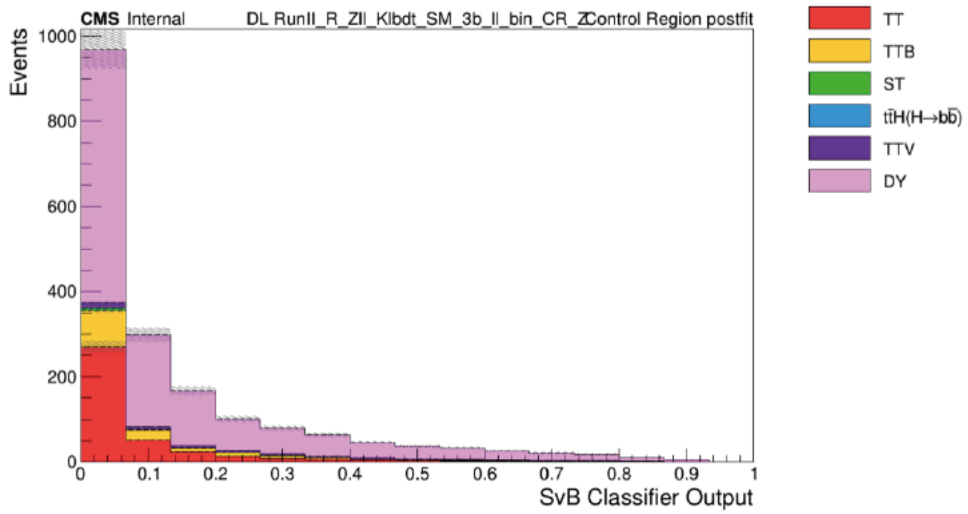


Figure 43: Pre-fit and Post-fit figures in rHH Side Band Region.


CMS Internal
DL RunII_R_Zll_Klbdt_High_3b_ll_bin_CR_ZControl Region prefit
Events
SvB Classifier Output
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY


CMS Internal
DL RunII_R_Zll_Klbdt_High_3b_ll_bin_CR_ZControl Region postfit
Events
SvB Classifier Output
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY


CMS Internal
DL RunII_R_Zll_Klbdt_SM_3b_ll_bin_CR_Z Control Region prefit
Events
SvB Classifier Output
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY


CMS Internal
DL RunII_R_Zll_Klbdt_SM_3b_ll_bin_CR_ZControl Region postfit
Events
SvB Classifier Output
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY

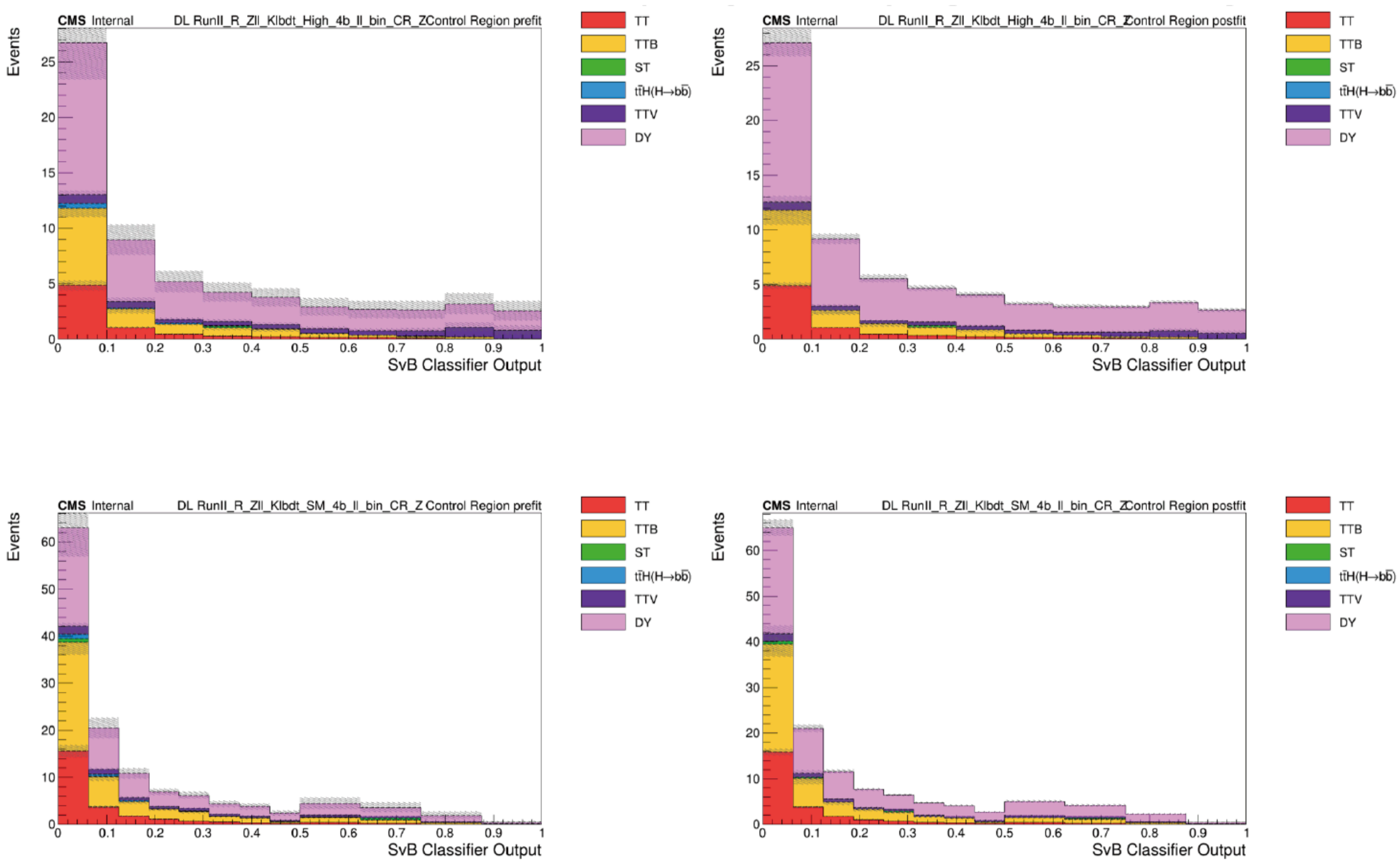
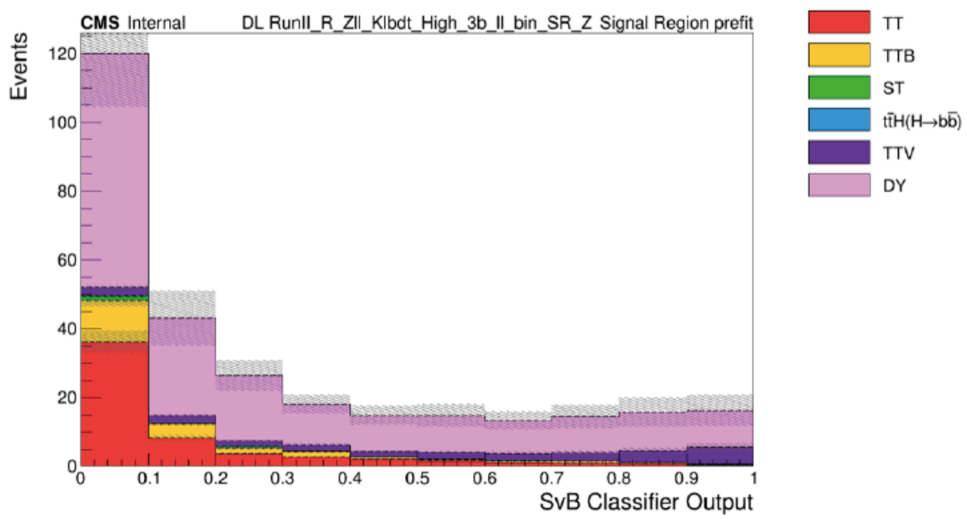
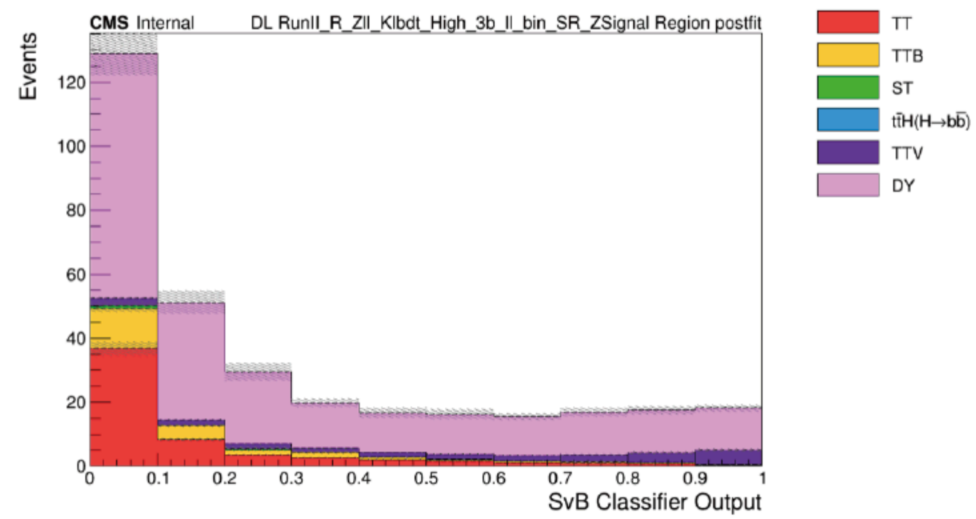
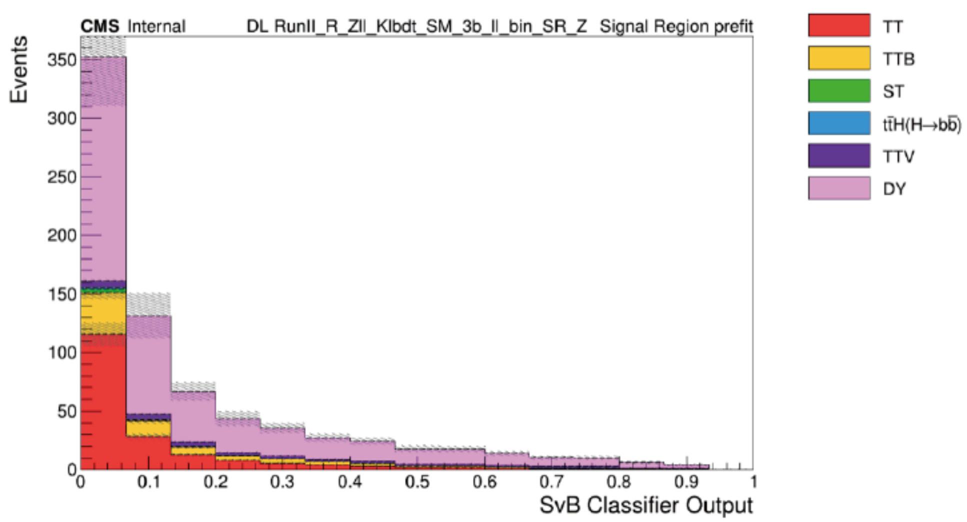
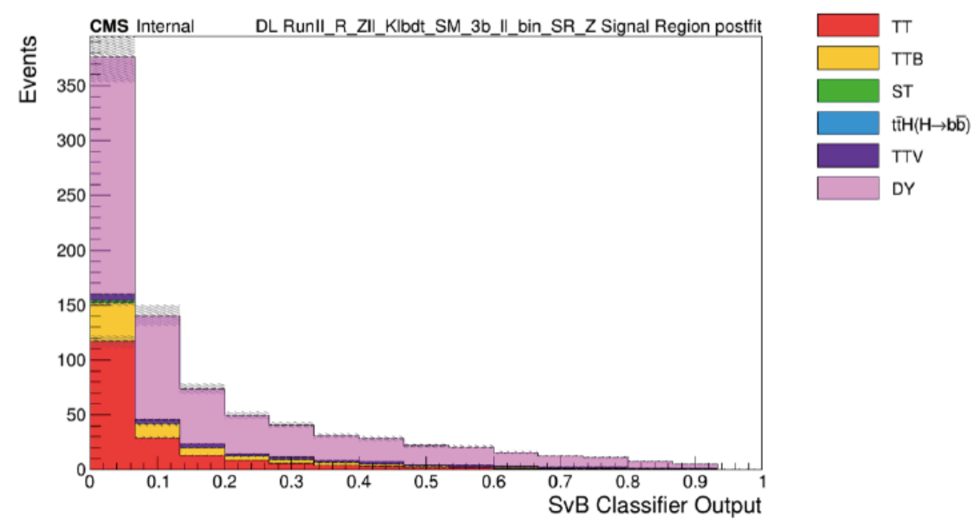


Figure 42: Pre-fit and Post-fit figures in rHH Control Region.


CMS Internal
DL RunII_R_Zll_Klbdt_High_3b_ll_bin_SR_Z Signal Region prefit
Events
SvB Classifier Output
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY


CMS Internal
DL RunII_R_Zll_Klbdt_High_3b_ll_bin_SR_ZSignal Region postfit
Events
SvB Classifier Output
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY


CMS Internal
DL RunII_R_Zll_Klbdt_SM_3b_ll_bin_SR_Z Signal Region prefit
Events
SvB Classifier Output
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY


CMS Internal
DL RunII_R_Zll_Klbdt_SM_3b_ll_bin_SR_Z Signal Region postfit
Events
SvB Classifier Output
TT
TTB
ST
tt̄H(H→bb̄)
TTV
DY

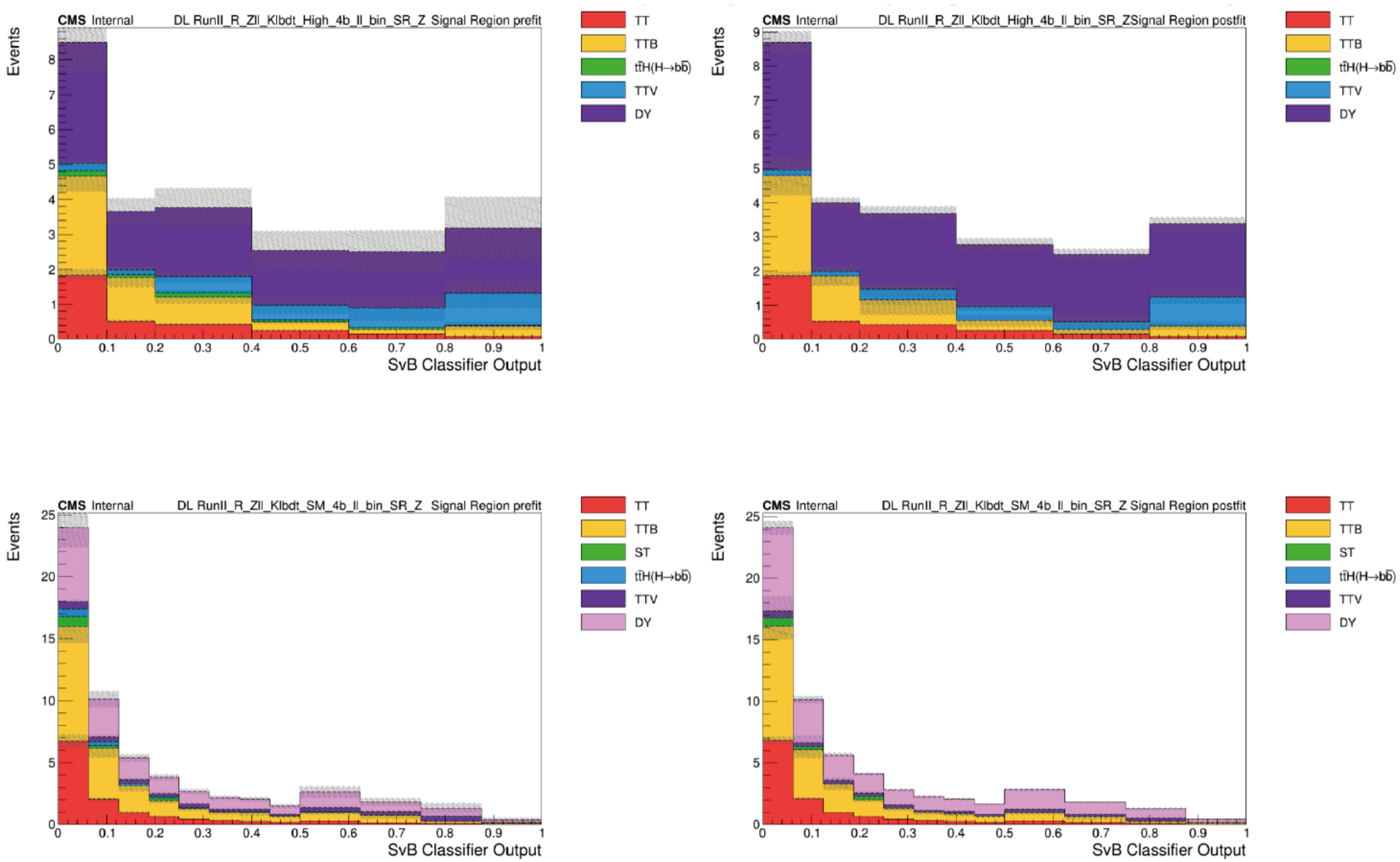
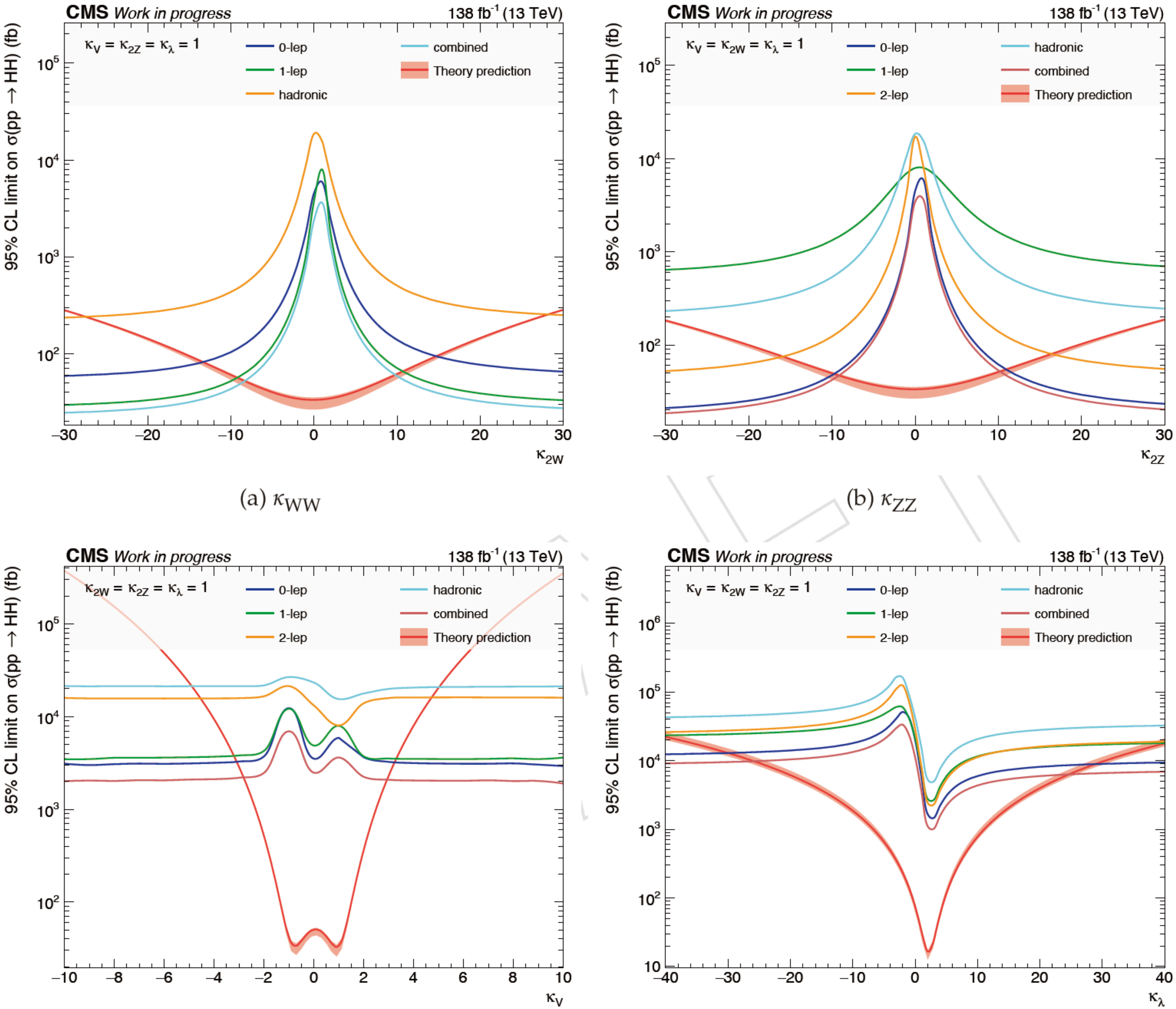


Figure 41: Pre-fit and Post-fit figures in rHH Signal Region.



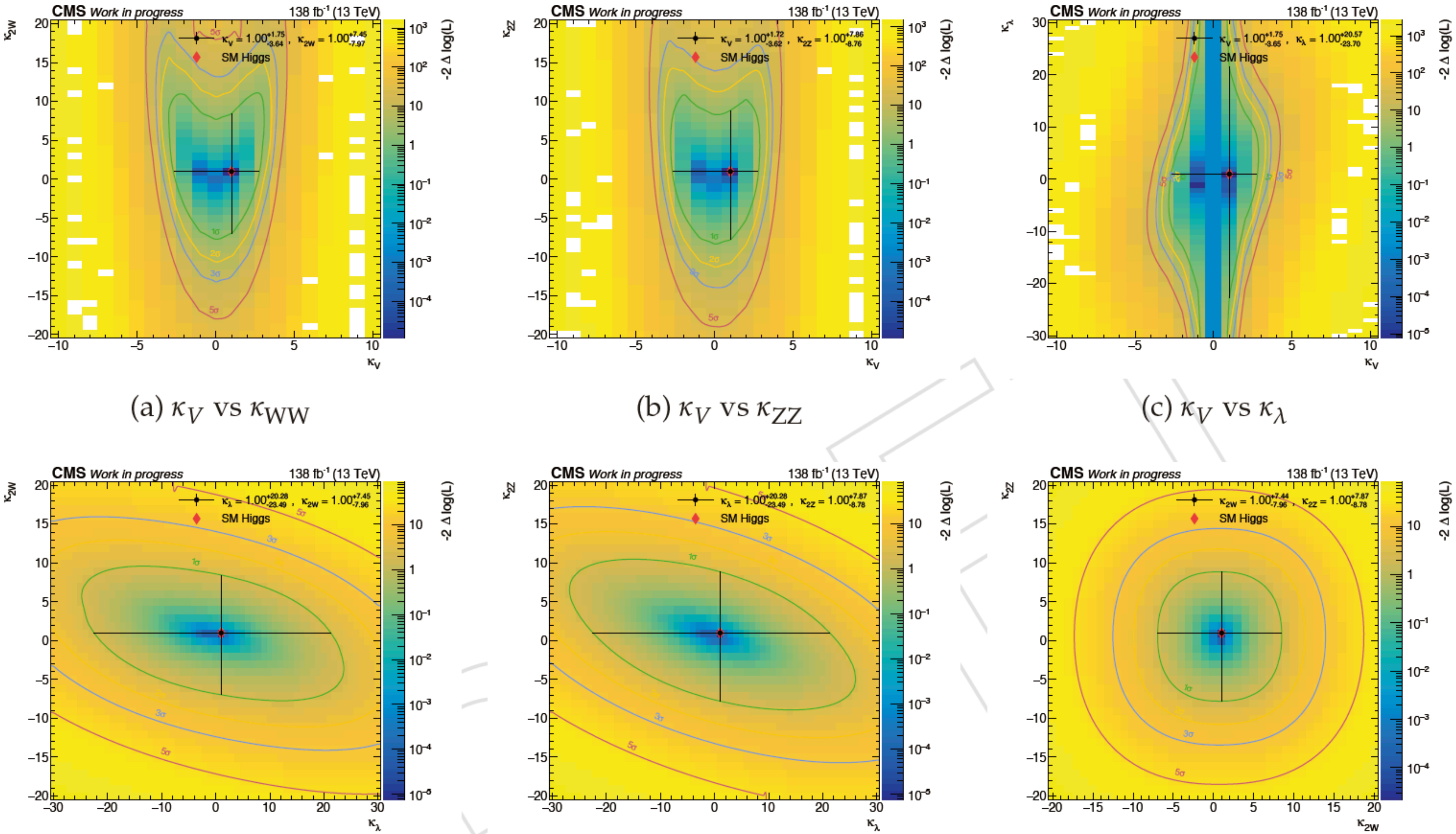
(a) κ_{WW} (b) κ_{ZZ}

(c) κ_V (d) κ_λ

Figure 130: Upper 95% CL limits on signal cross section scanned over the κ parameter of interest while fixing the other two to SM strength. x-axis is the scanned κ parameter and y-axis is the 95% CL upper limit on signal cross section. (a) Scans over κ_{WW}. (b) Scans over κ_{ZZ}. (c) Scans over κ_V. (d) Scans over κ_λ .

Table 39: 95% exclusion [lower, upper] limits on κ_V, κ_{WW}, κ_{ZZ} and κ_λ in each sub-channel and in the combined results. '-' sign means the fit didn't find an exclusion value on the POI within the scanned range shown in Figure 130.

Channel	κ_V	κ_{WW}	κ_{ZZ}	κ_λ
0-lep	[-3.09,3.23]	[-13.34,14.72]	[-9.76,10.96]	[-30.00,28.88]
1-lep	[-3.21,3.30]	[-9.24,10.71]	[-,-]	[-, -]
2-lep	[-4.54,4.72]	[-,-]	[-15.91,16.41]	[-,-]
Hadronic	[-4.89,5.02]	[-27.43,28.16]	[-,-]	[-, -]
Combination	[-2.82,2.93]	[-8.34,9.79]	[-9.16,10.23]	[-25.85,24.64]

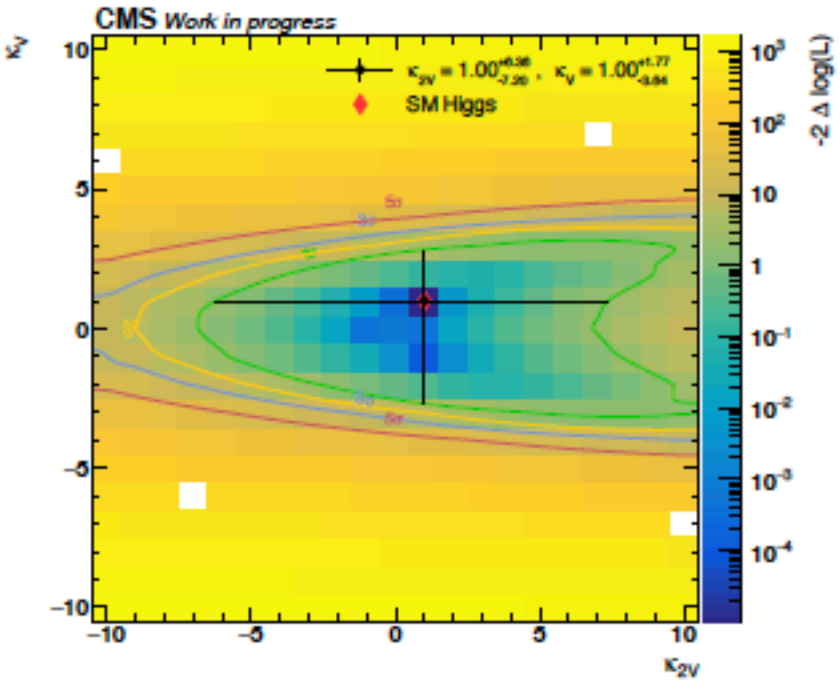


(a) κ_V vs κ_{WW} (b) κ_V vs κ_{ZZ} (c) κ_V vs κ_λ

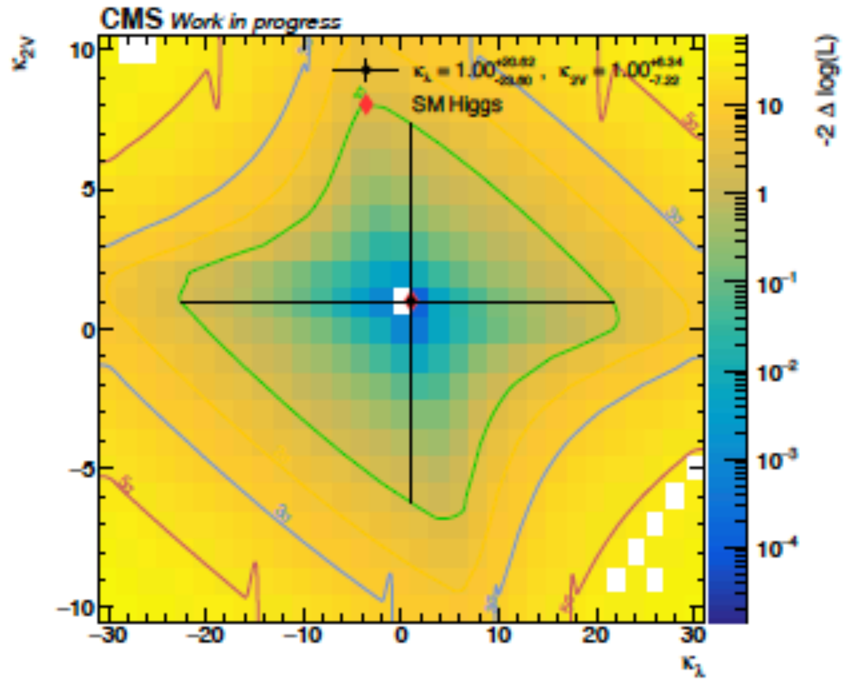
(d) κ_λ vs κ_{WW} (e) κ_λ vs κ_{ZZ} (f) κ_{WW} vs κ_{ZZ}

Figure 132: 2D likelihood profiles of two scan parameters as well as the position and errors of their best fit values while fixing the third parameter to its SM strength.

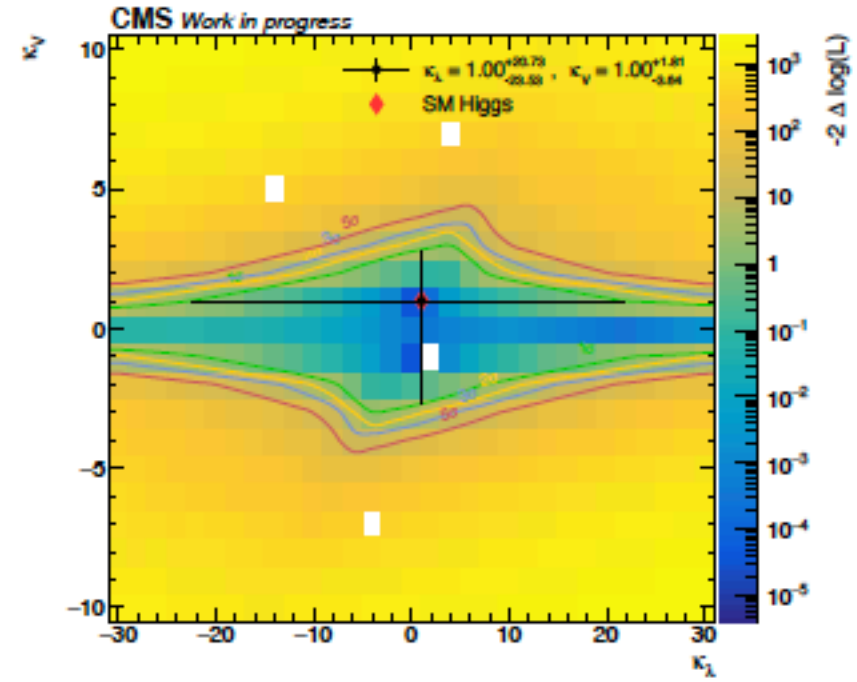
VHH Analysis backup



- K_v vs K_{VV}



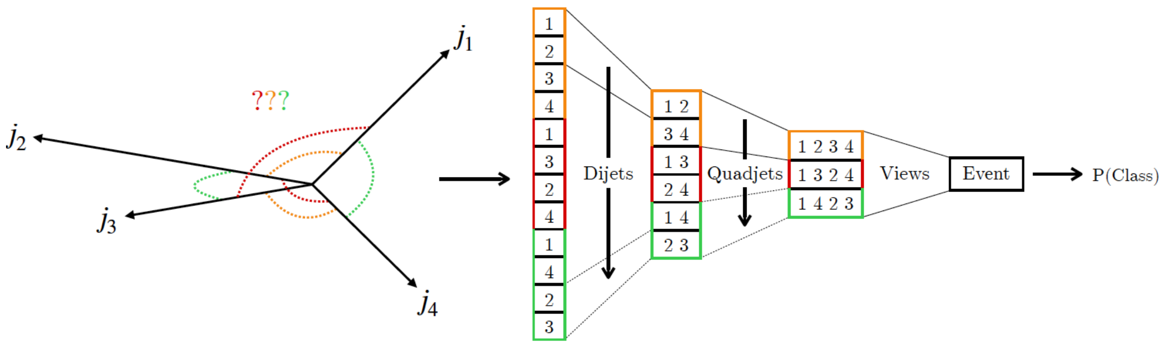
- K_{VV} vs K_λ



- K_v vs K_λ

Hadronic Channel

Machine Learning SvB Classifier



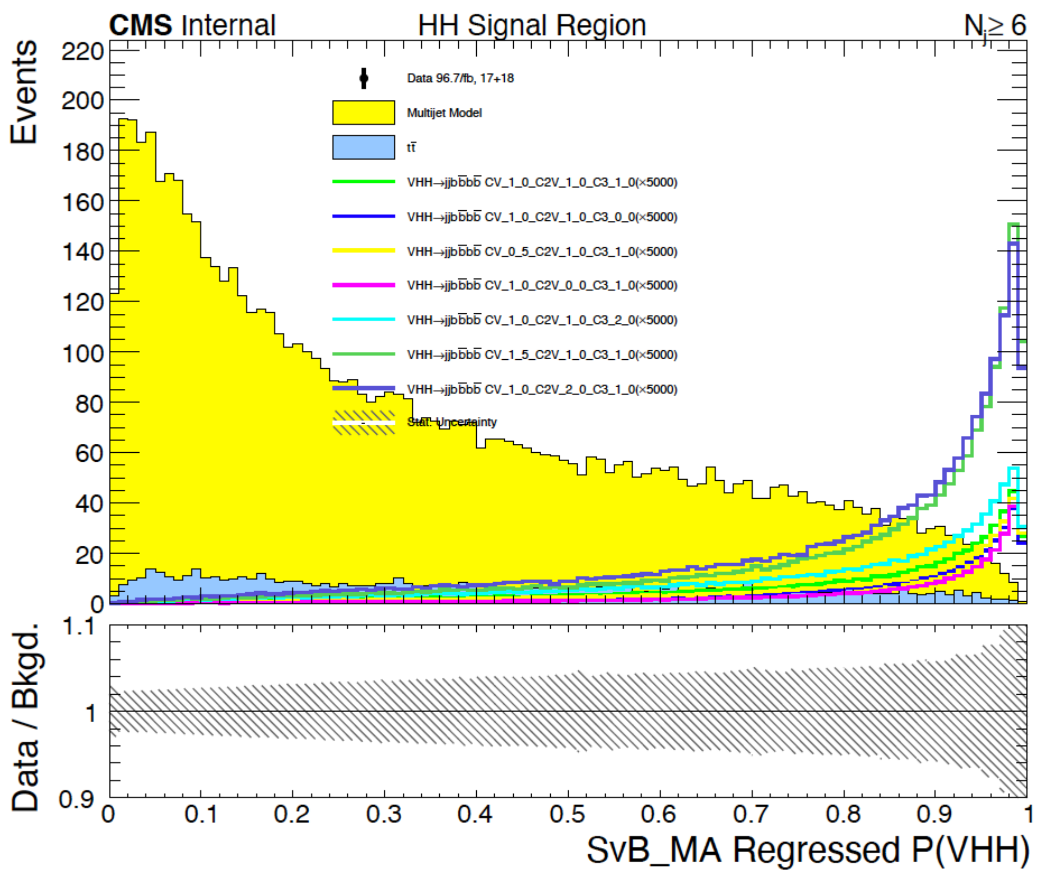


Figure: SvB Classifier Output

- Input: engineered jet features

$$pT, \eta, \phi, m_j, m_{jj}, m_{4j}, \Delta R_{jj}$$

- Architecture:
  - Hierarchical Combinatoric ResNet(Residual Neural Network):

Learn di-jets and qua-jets features and pick the correct combination to form 2 Higgs Bosons from 4 Higgs candidate jets.

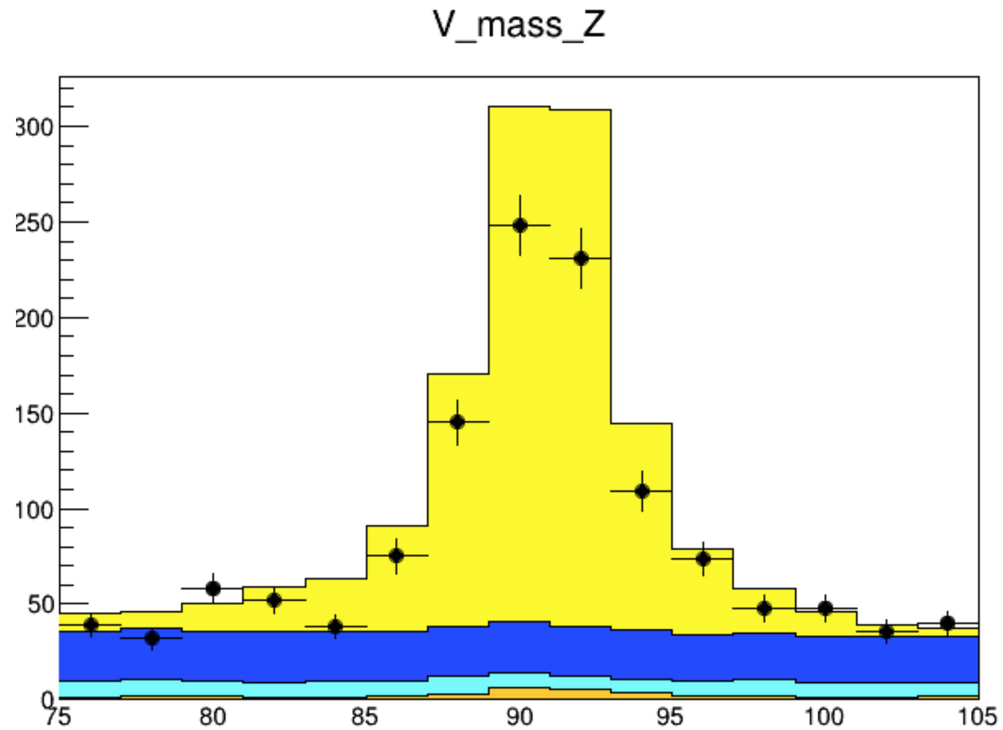
  - Multi-Head Attention Block:

Process other jets. Add di-jets features and expect a better performance.

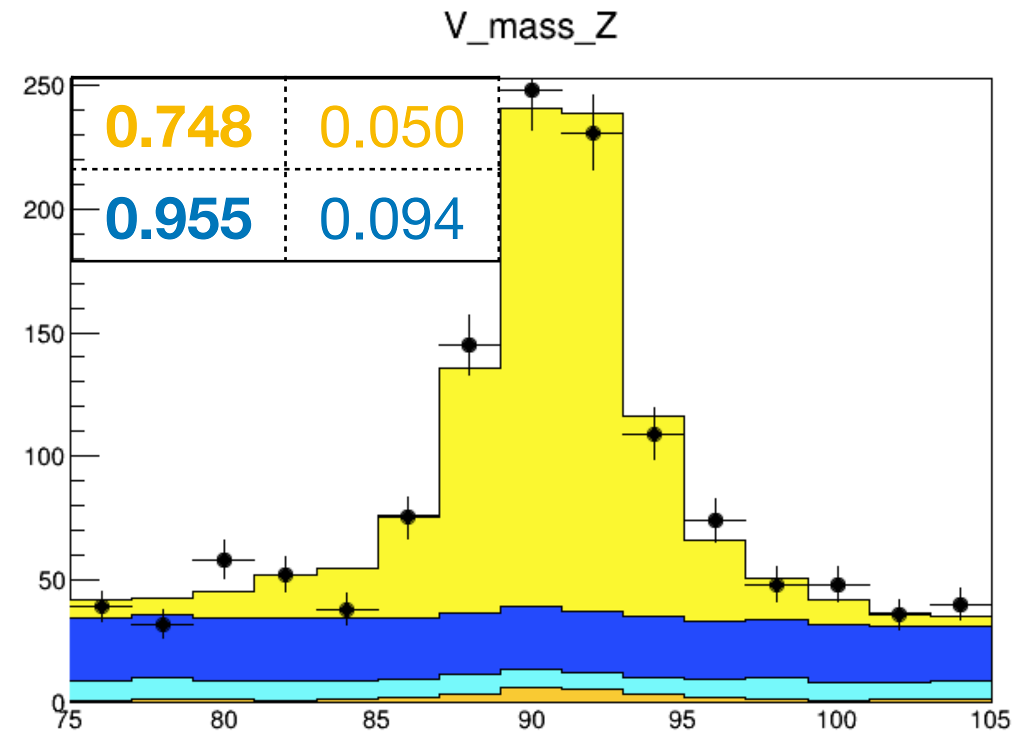
- Output: regressed probability that an event is VHH signal.

- Samples: The signal for training is both 2017 and 2018, ZHH snd WHH.

Data/MC from 3b or 4b In SB

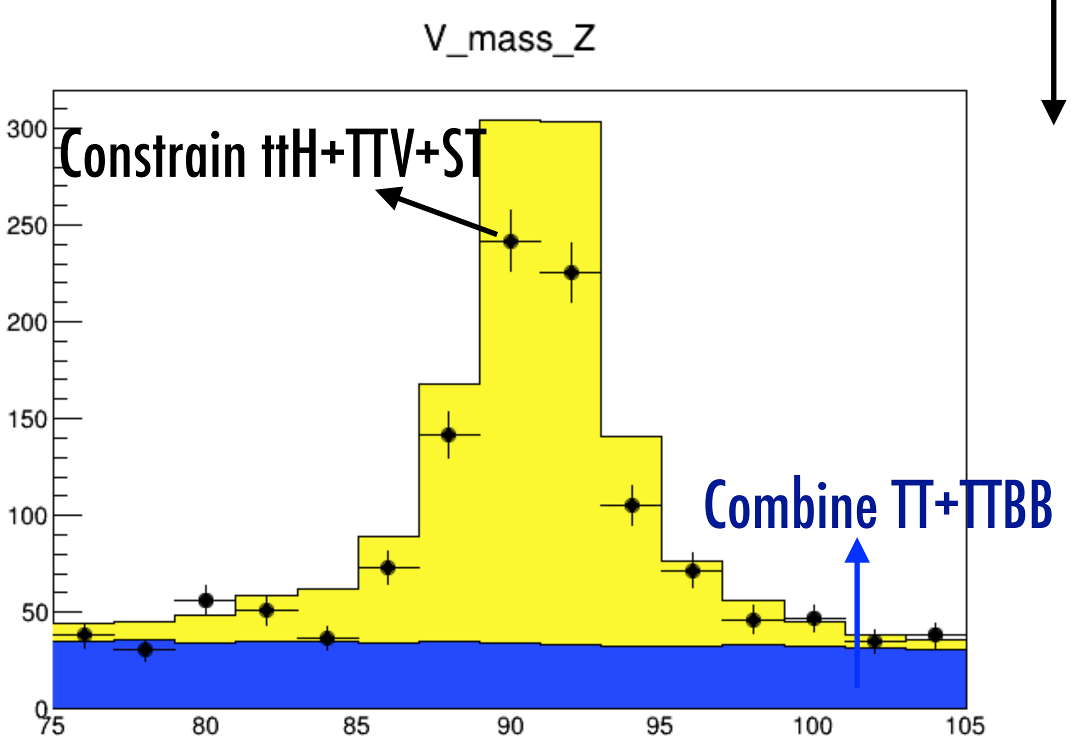


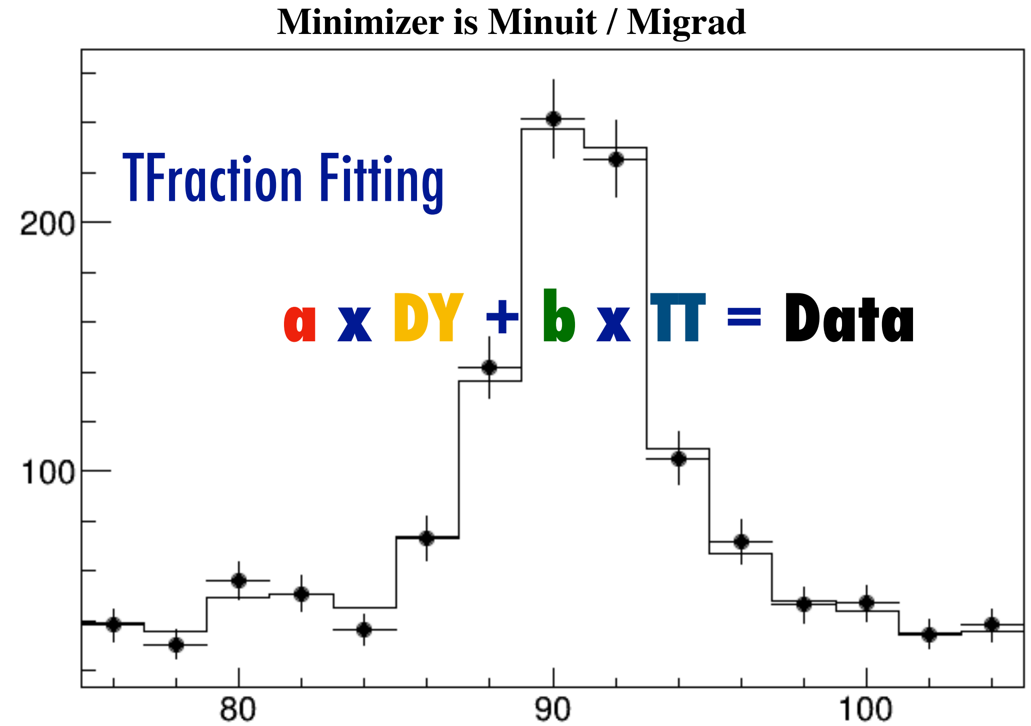
Original distribution



After SFs.

```
EXT PARAMETER                                 STEP         FIRST
NO.   NAME       VALUE          ERROR         SIZE         DERIVATIVE
 1   TT+TTBB     3.87960e-01   3.83189e-02   2.18548e-03  -1.24559e-03
 2   DY          6.12039e-01   4.14749e-02   1.96963e-03   3.22527e-04
                               ERR DEF= 0.5
```

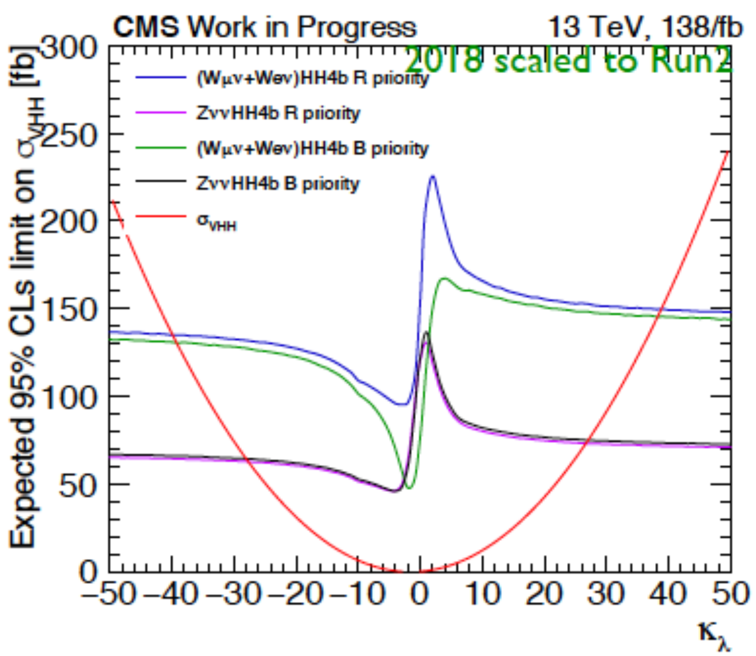


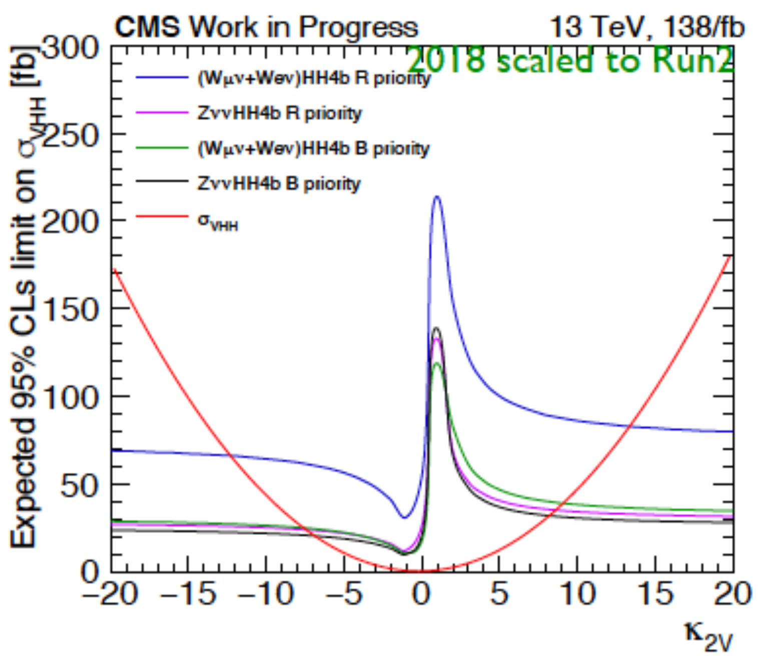


Re-weighted MC

Topology priority

- ★ Since we got 2 topologies, some events can be classified as both. An inevitable question is: How to decide which event should be included as which topology?
- ★ 2 choices are tested, put all the common events into the one topology at the same time. In the plots below, R priority means put them into Resolved, B priority means Boosted topology
- ★ Apparently we should prioritize the Boosted topology in both channels
  - Mainly because the number of the pure Boosted events are too small

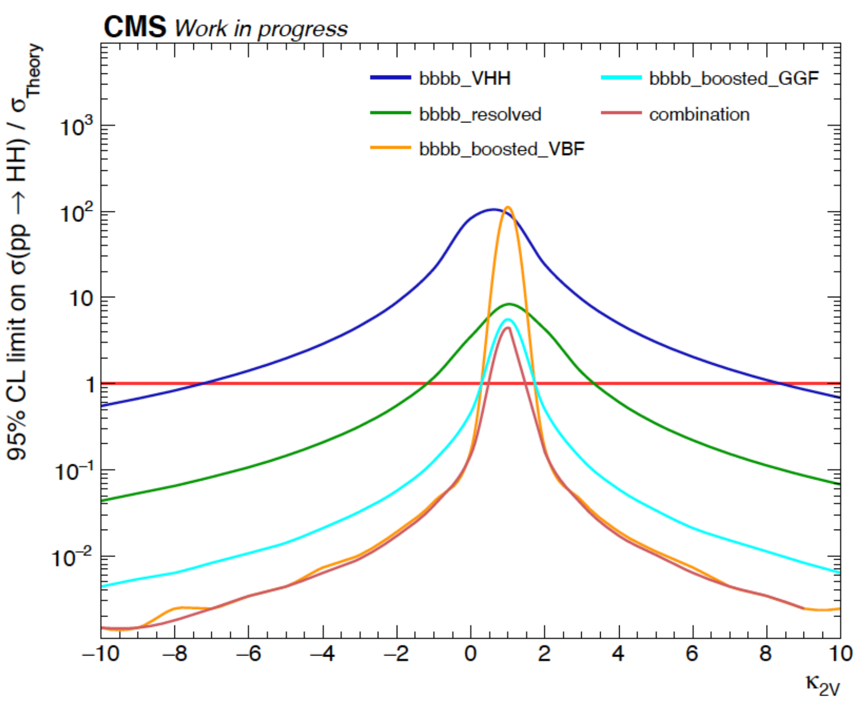


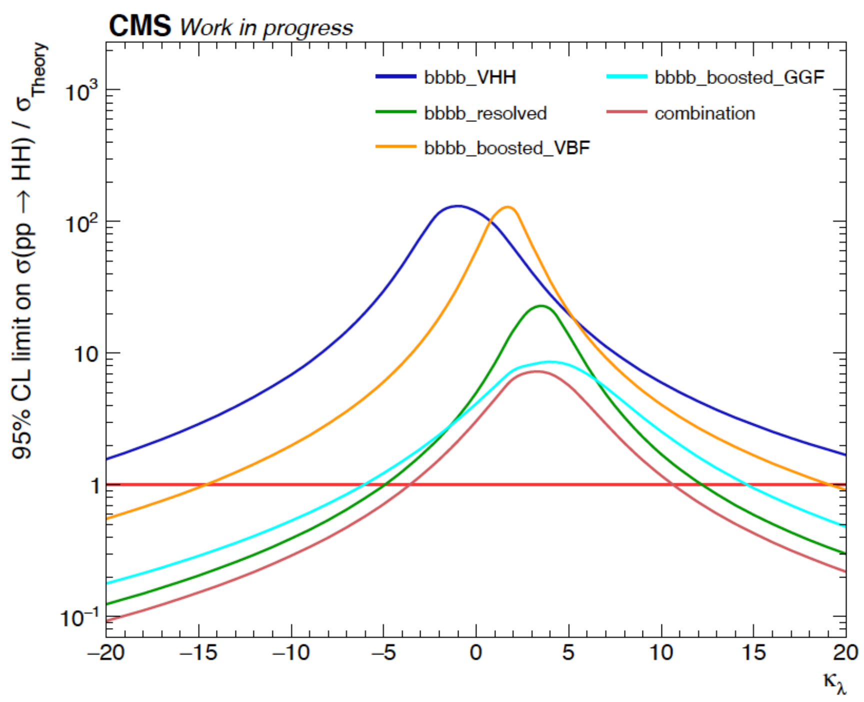


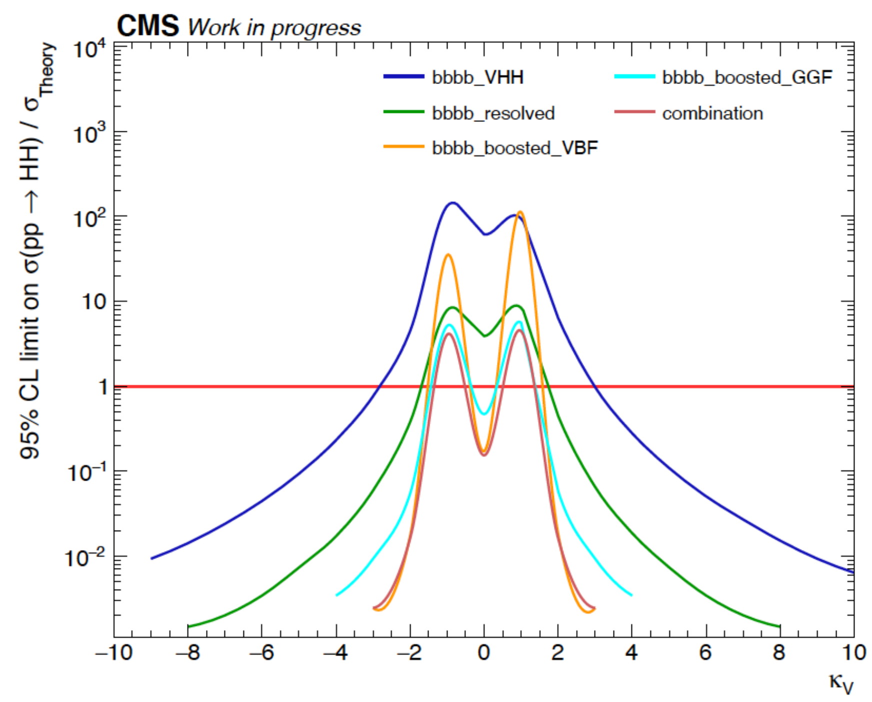
Prioritize the Boosted topology

28

New Result - Comparison to Other 4b Analysis







- Upper limit scan over VHH of κ_{VV} parameter in different 4b channels
- Upper limit scan over VHH of κ_{λ} parameter in different 4b channels
- Upper limit scan over VHH of κ_{V} parameter in different 4b channels

- Expected upper limits comparison between VHH and ggF/VBF channels
- VHH SM cross section is of orders smaller then resolved HH, about same order to boosted VBF, reasonable results shown in plots
- VHH tends to be more boosted at SM point, comparable result with boosted VBF channel

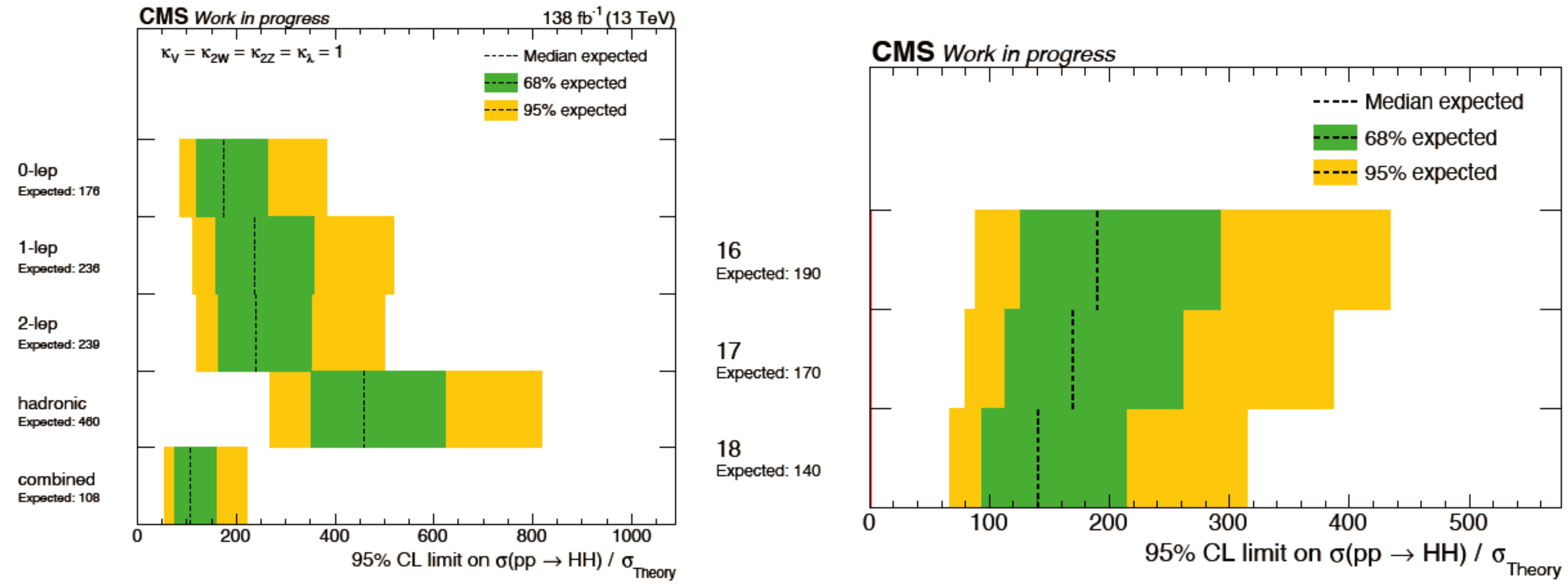


Figure 133: Upper limits on VHH production cross section at SM coupling point divided by theoretical prediction separated by channels (left) and by data years (right).

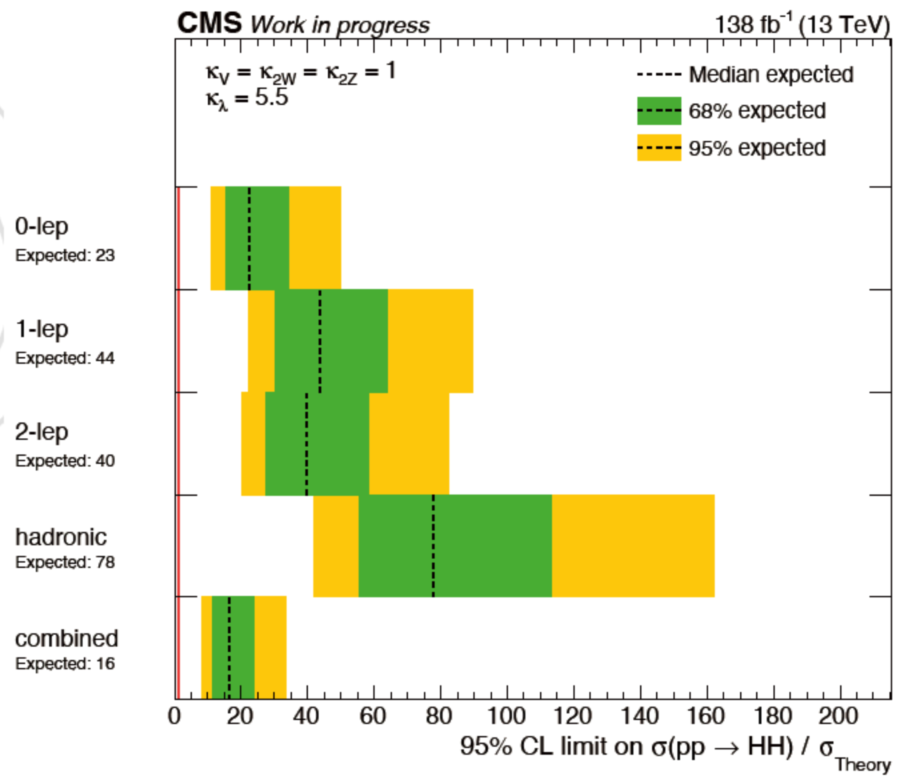


Figure 134: Upper limits on VHH production cross section at $\kappa_\lambda = 5.5$ coupling point divided by theoretical prediction separated by channels and combination.