VBS WH Analysis

Issues with the EWK W/Z samples

September 2nd, 2022

*P. Chang, L. Giannini, **J. Guiang**, F. Würthwein, Y. Xiang*

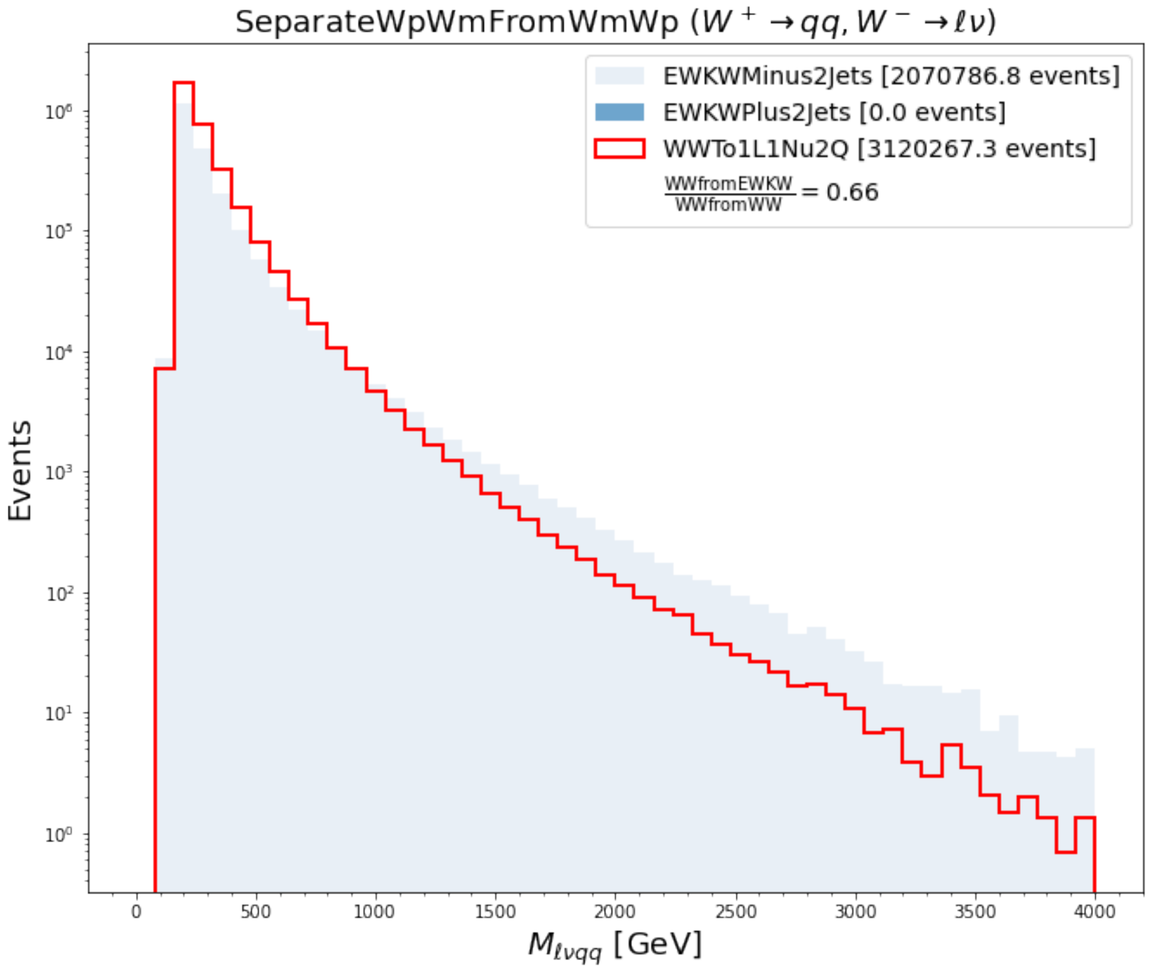
Overview

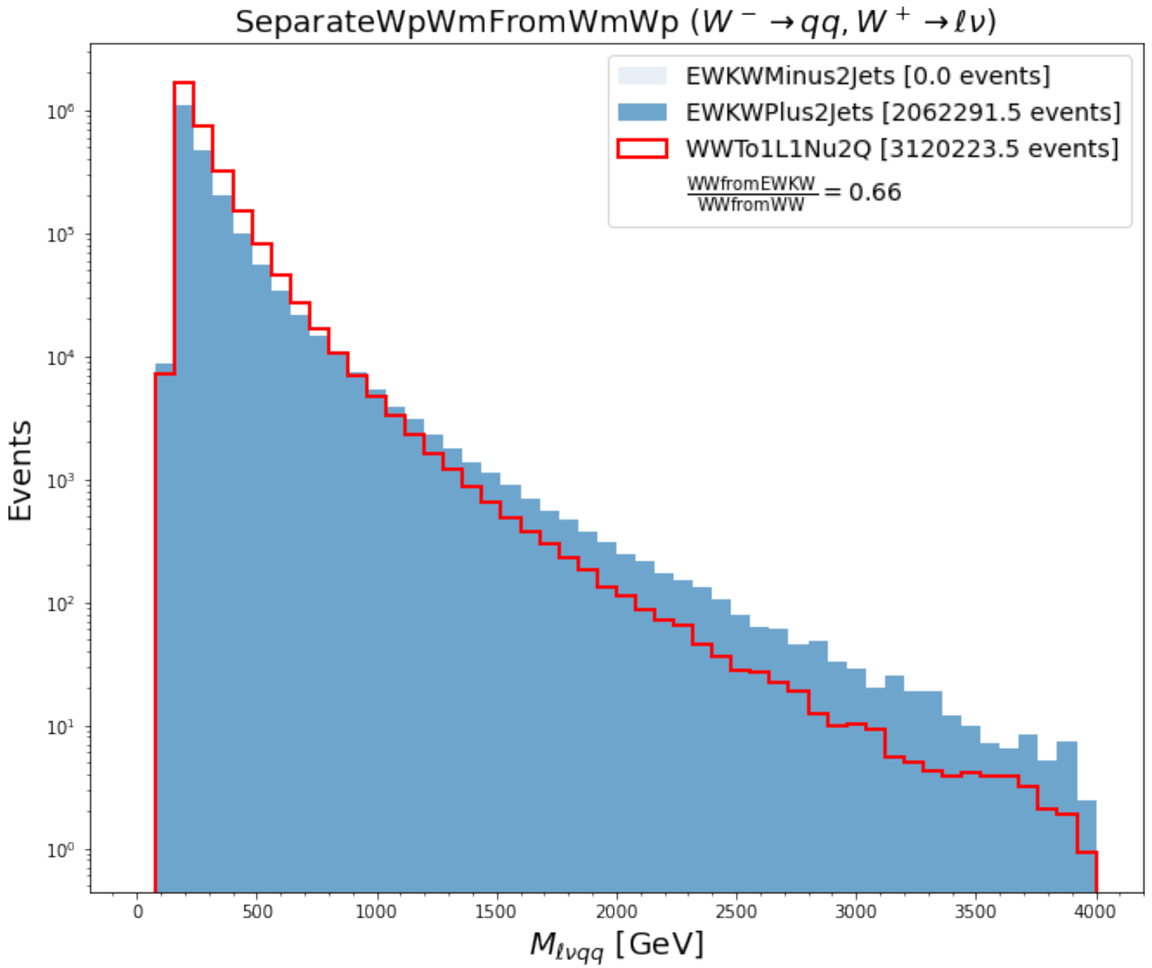
- New EWK samples add *a lot* of background
  - Entirely from EWK W± (W→ℓν)
  - Not just VBS W/Z: also have diboson events
- **Goal:** compare diboson contribution from EWK samples to dedicated diboson samples
  - For simplicity, compare only EWK W vs. WW sample
  - If good match: EWK samples may be correct
  - If bad match: EWK samples may be wrong

Sample	σ [pb]
EWK W^+, W→ℓν	39.33
EWK W^+, W→qq	10.67
EWK W^-, W→ℓν	32.26
EWK W^-, W→qq	10.67
EWK Z, Z→ℓℓ	6.22
EWK Z, Z→νν	10.72
EWK Z, Z→qq	10.67

Comparing Diboson Contributions

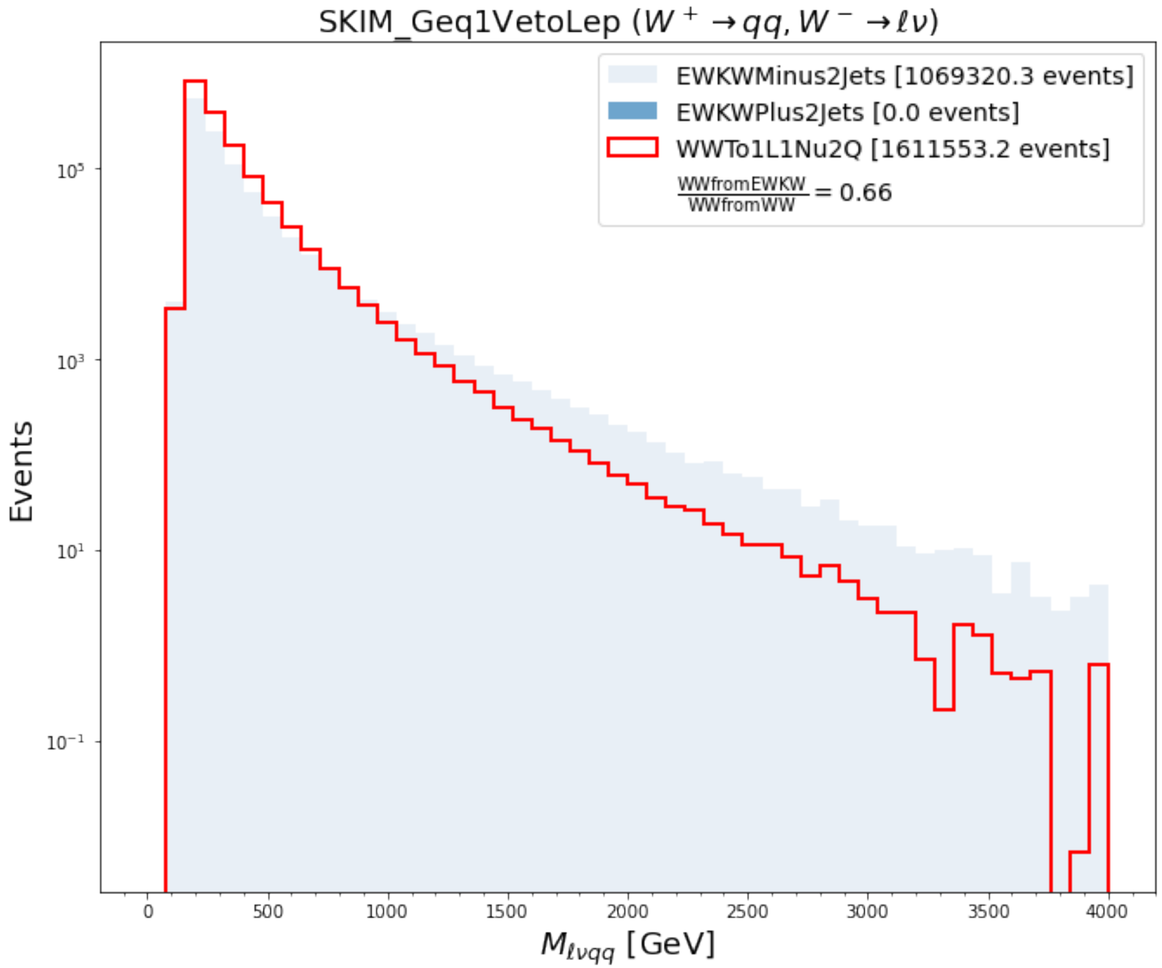
Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

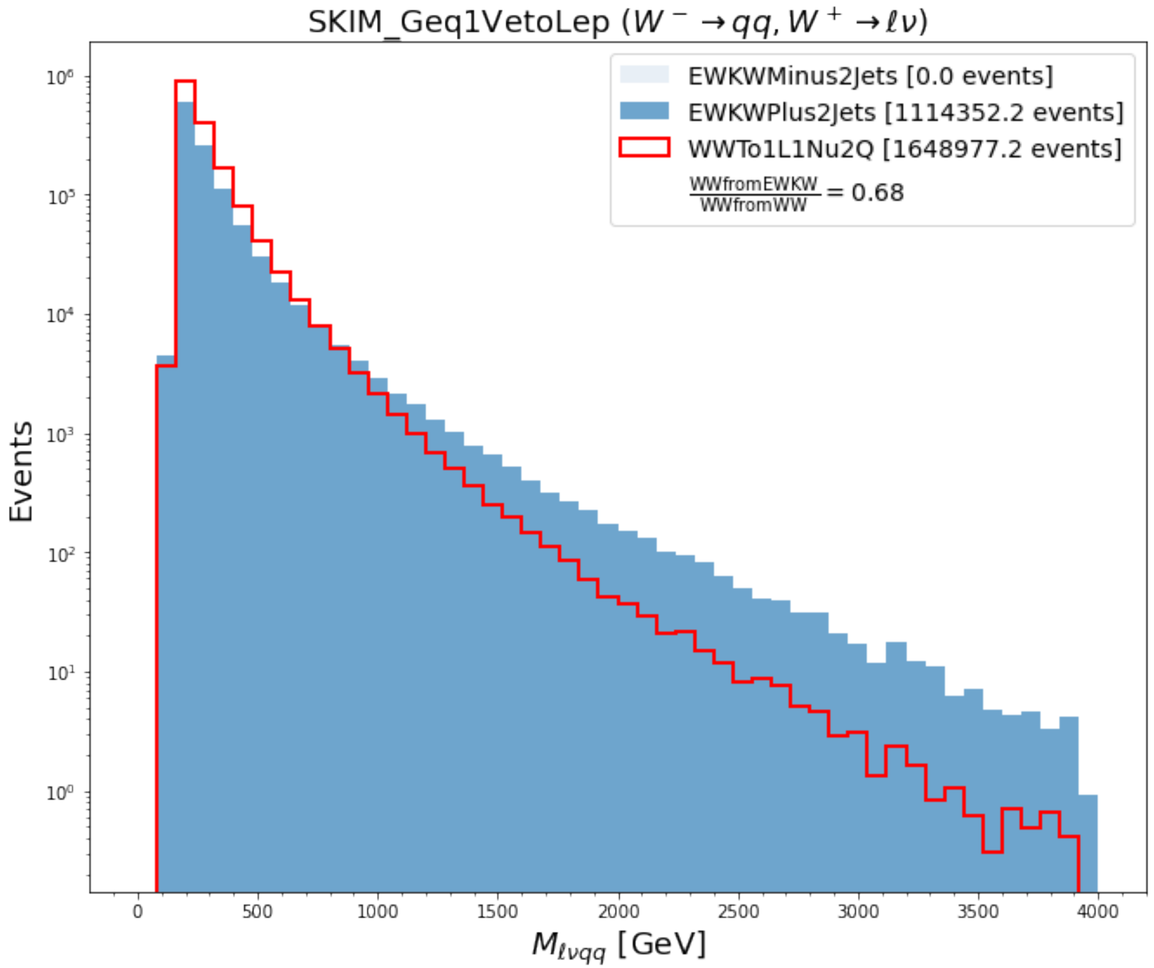




Comparing Diboson Contributions

Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

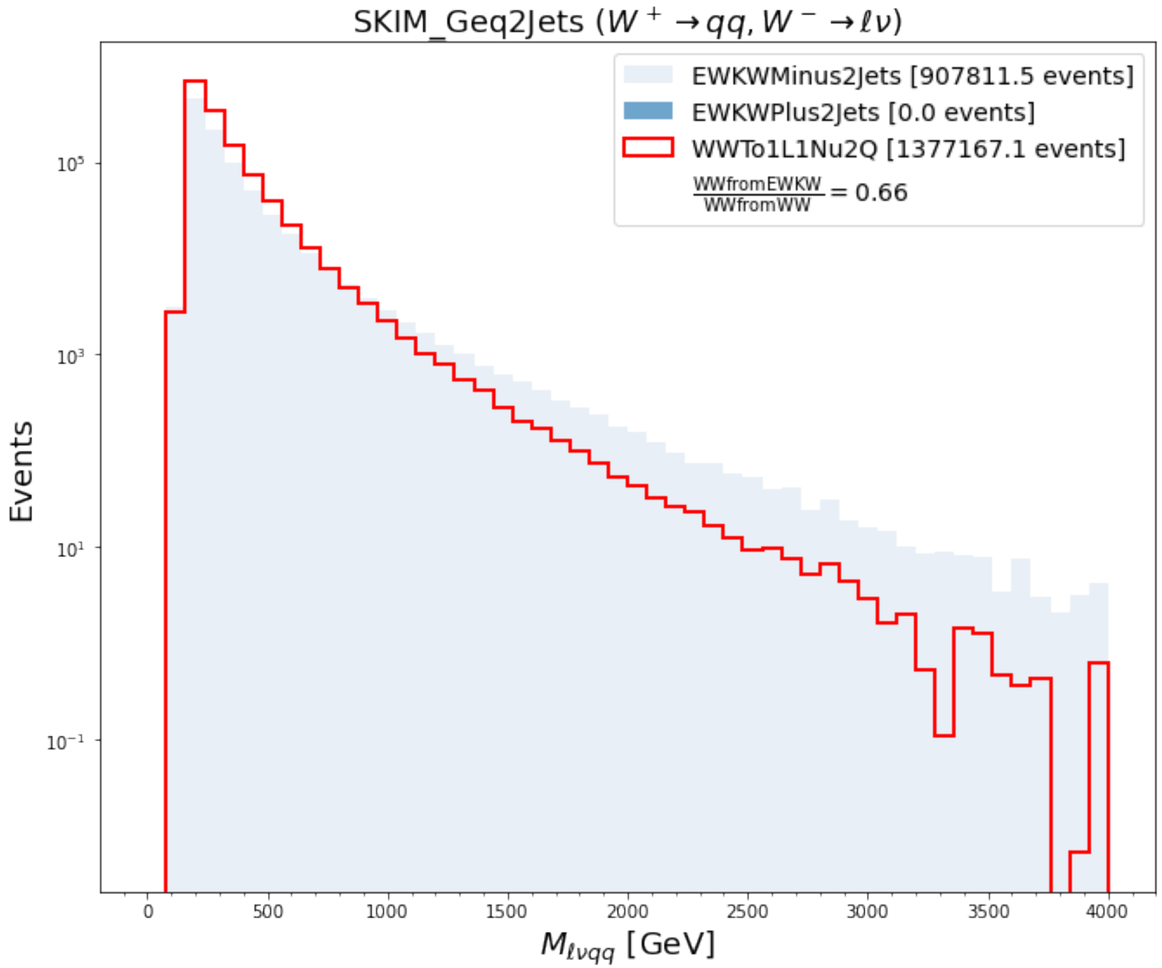


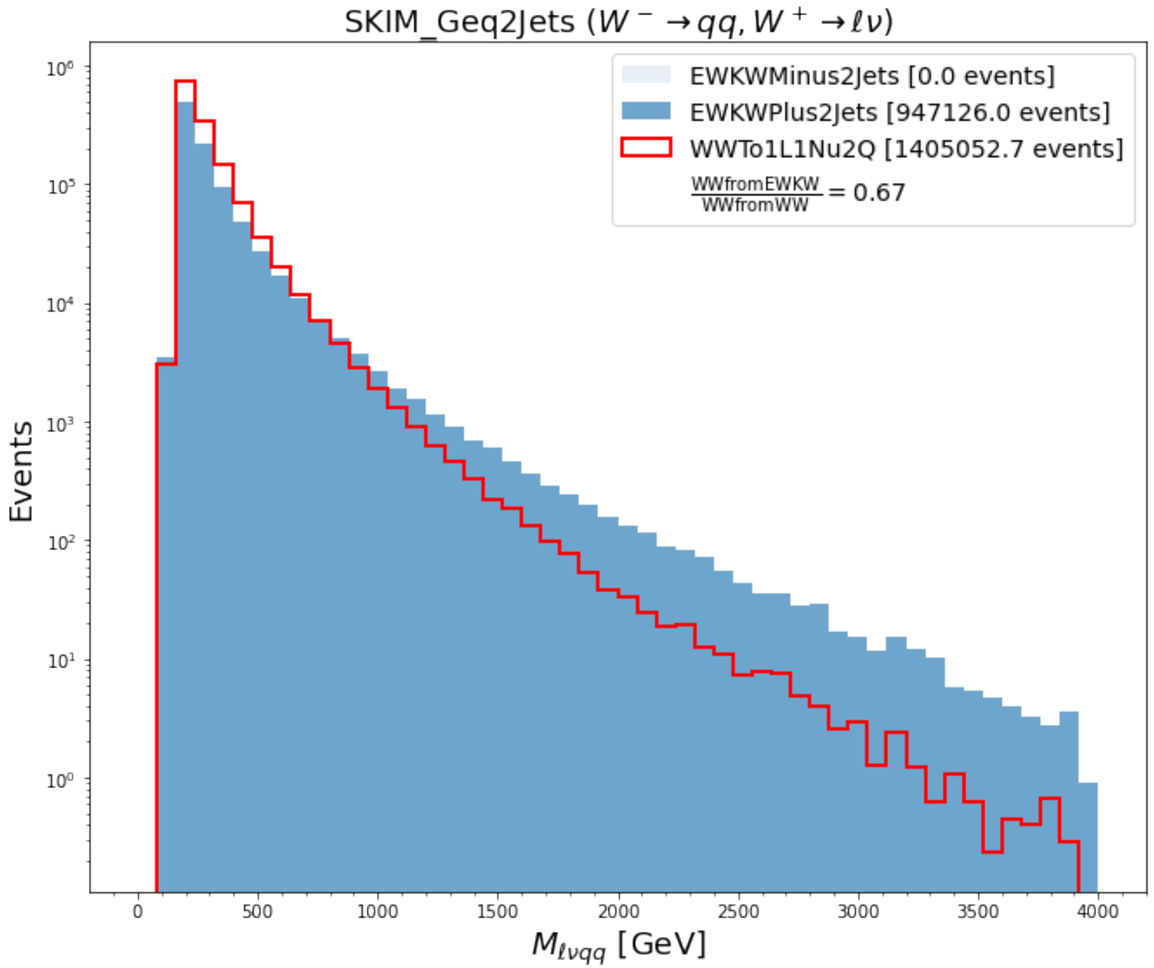


UC San Diego

Comparing Diboson Contributions

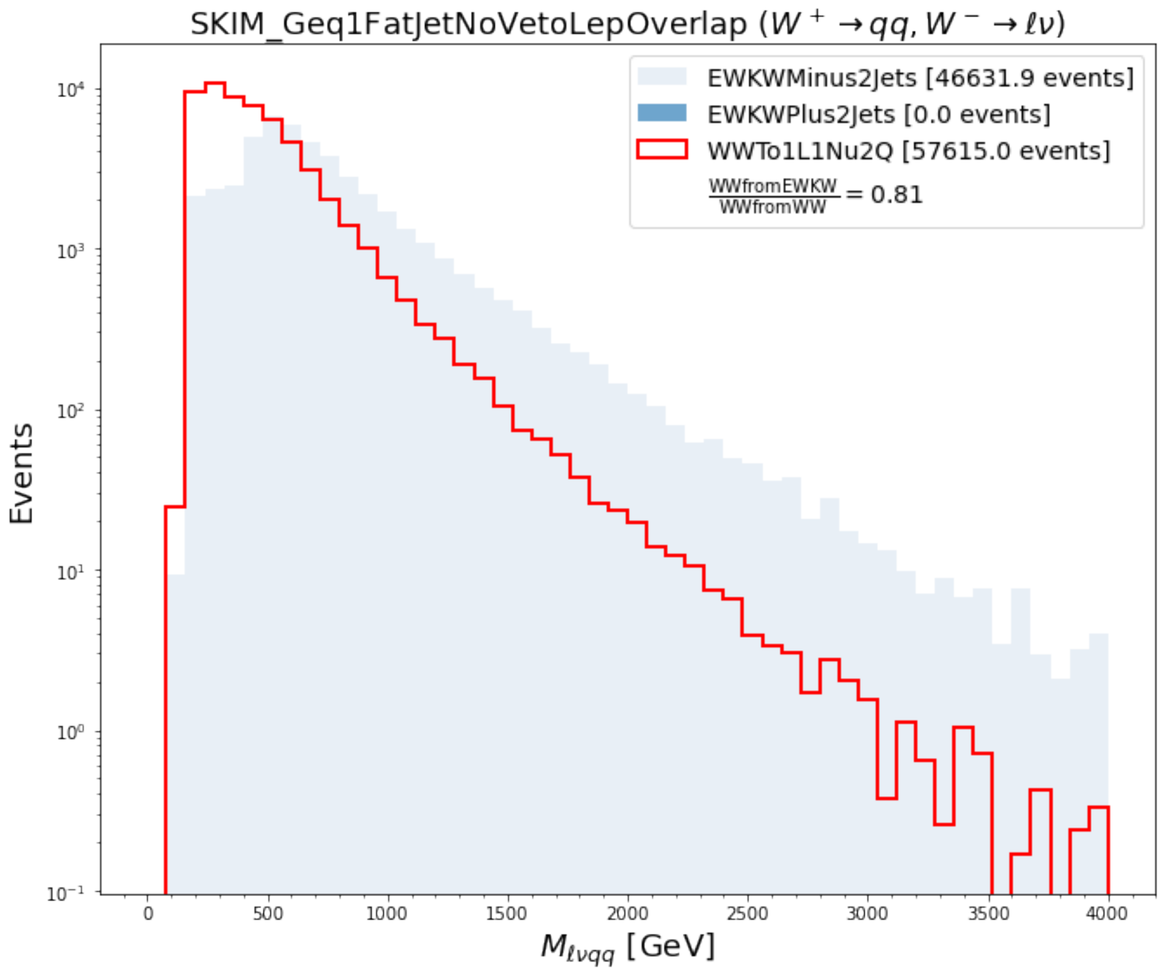
Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

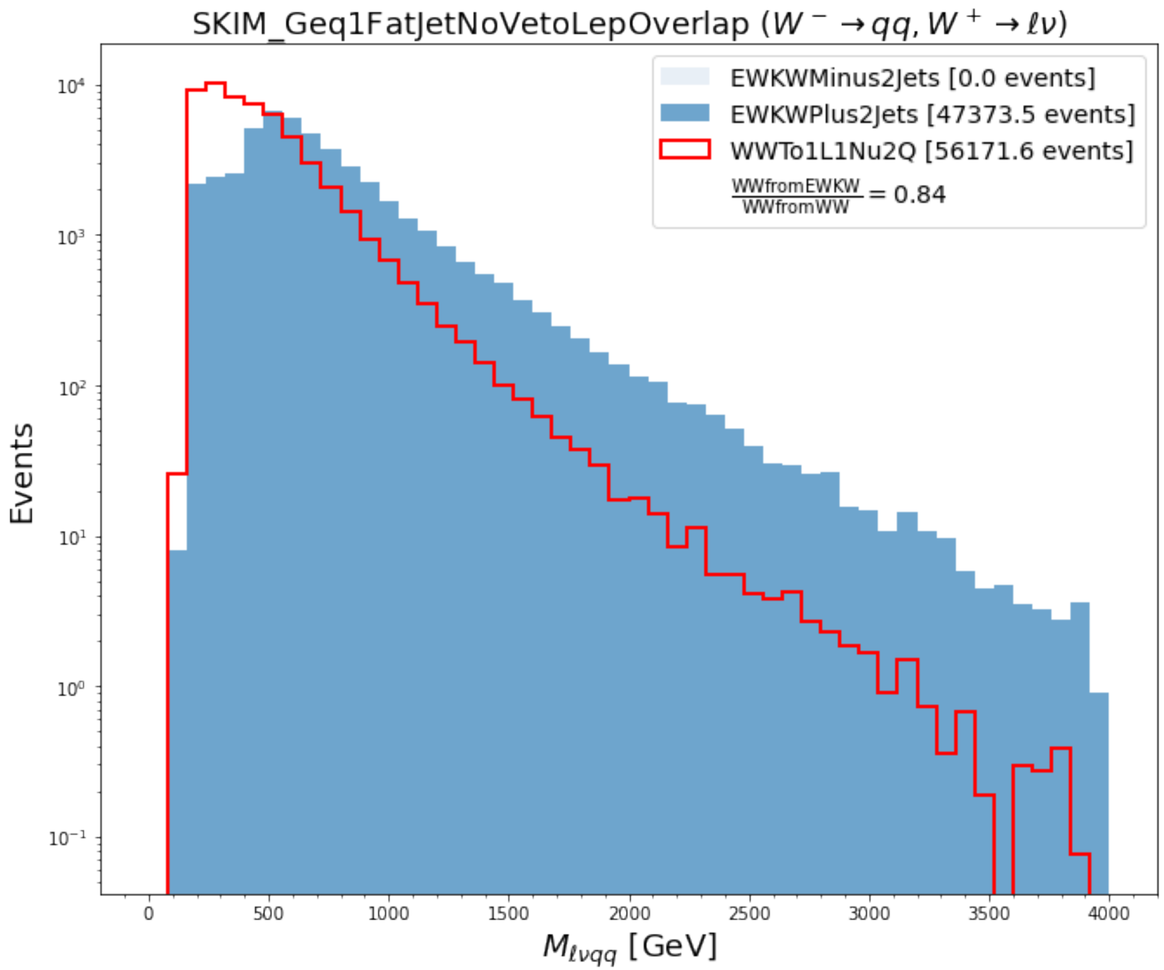




Comparing Diboson Contributions

Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

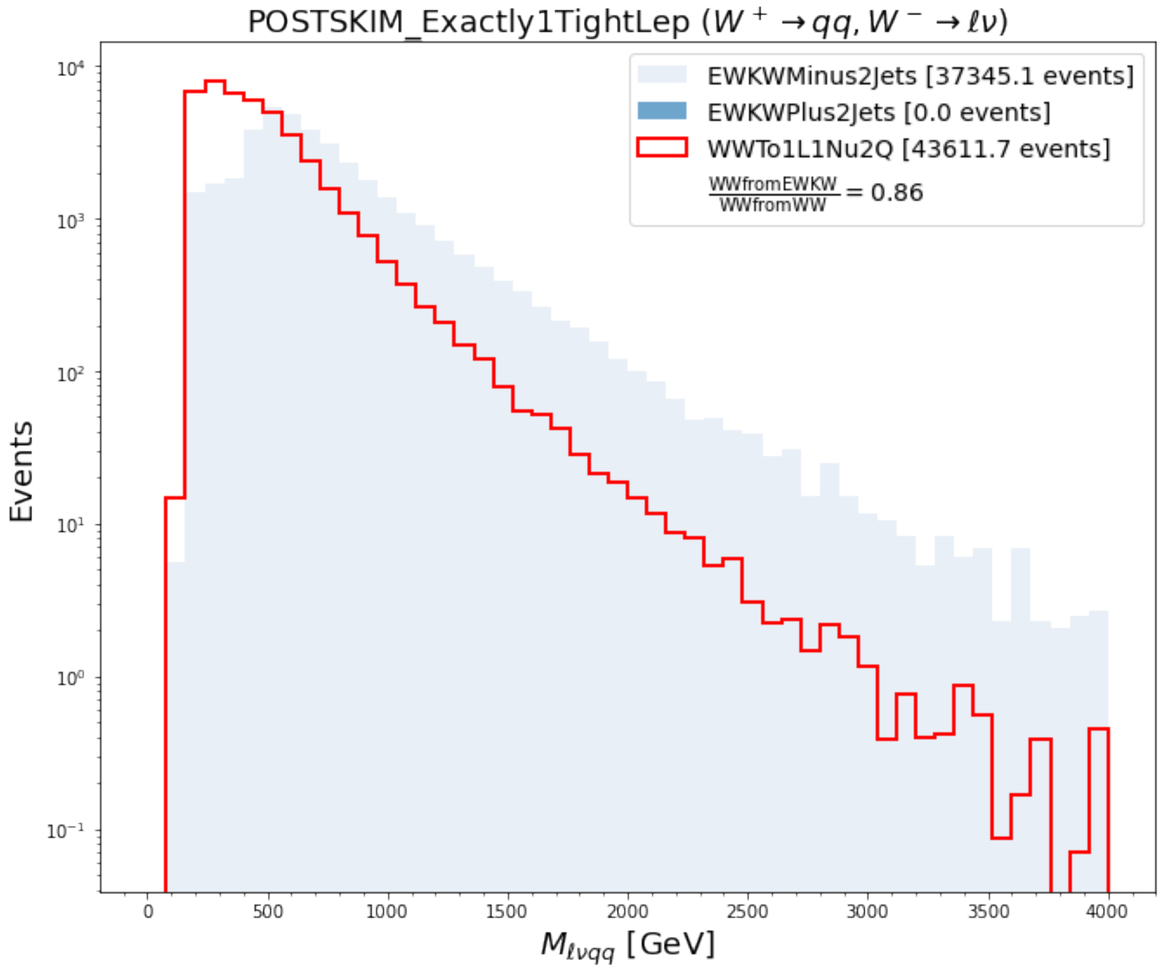


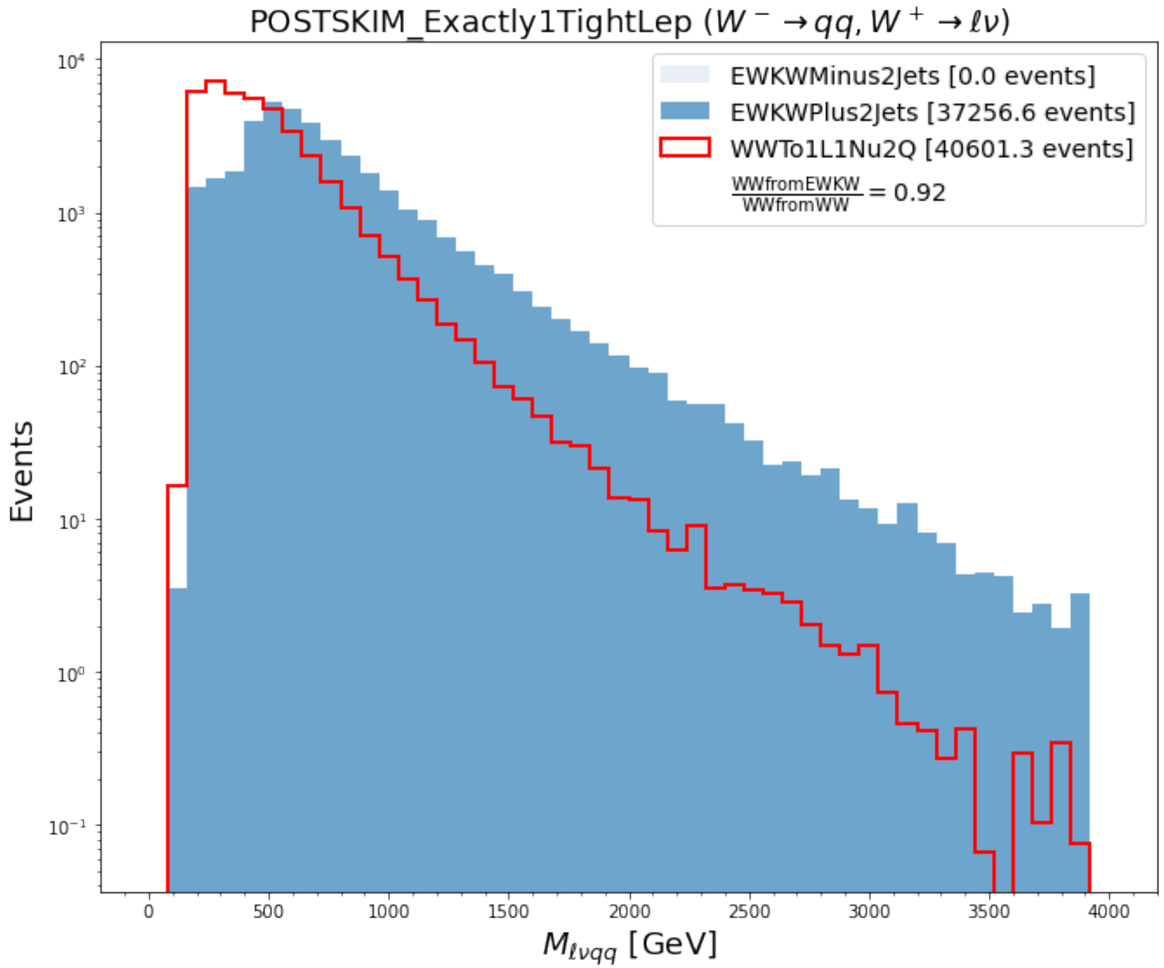


Cutoff starts to develop after fatjet selection

Comparing Diboson Contributions

Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

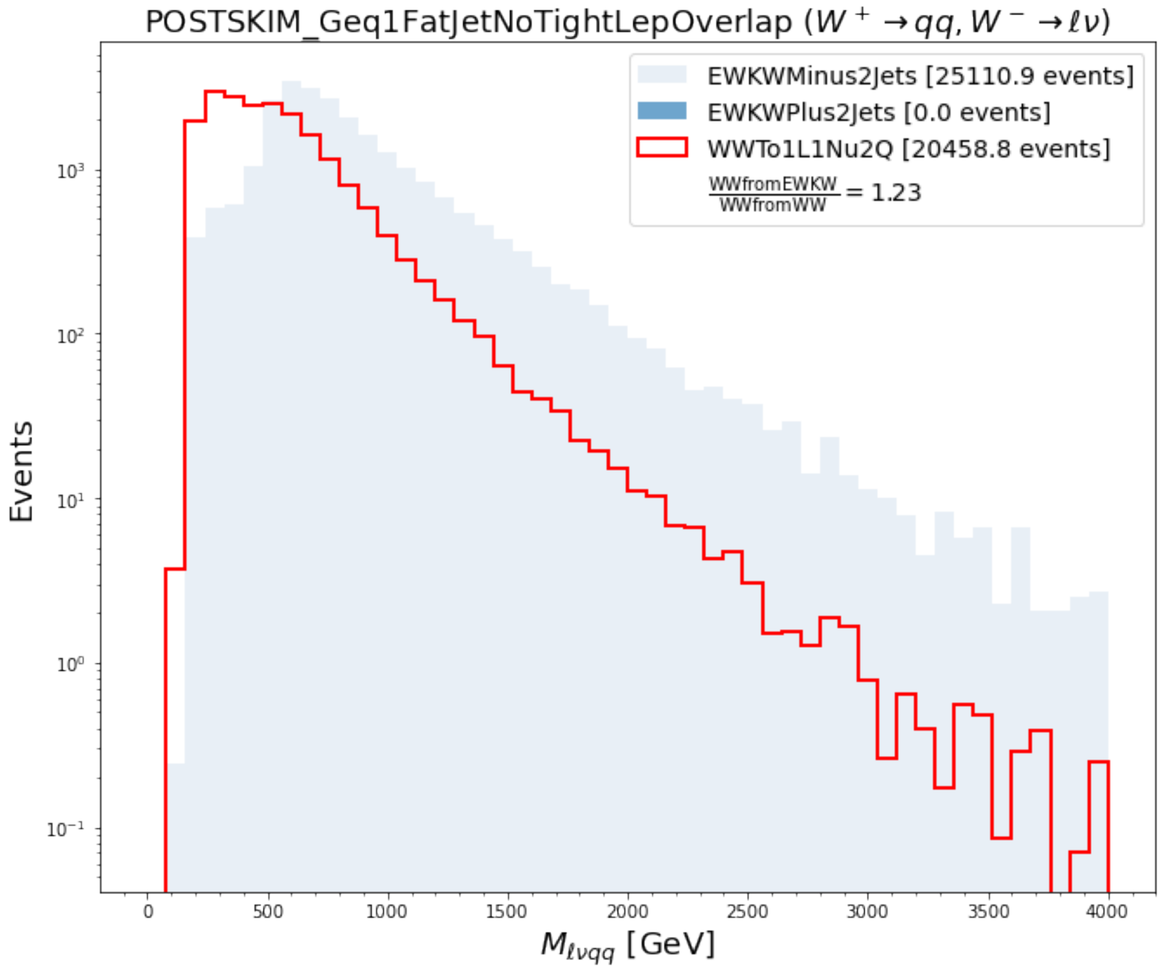


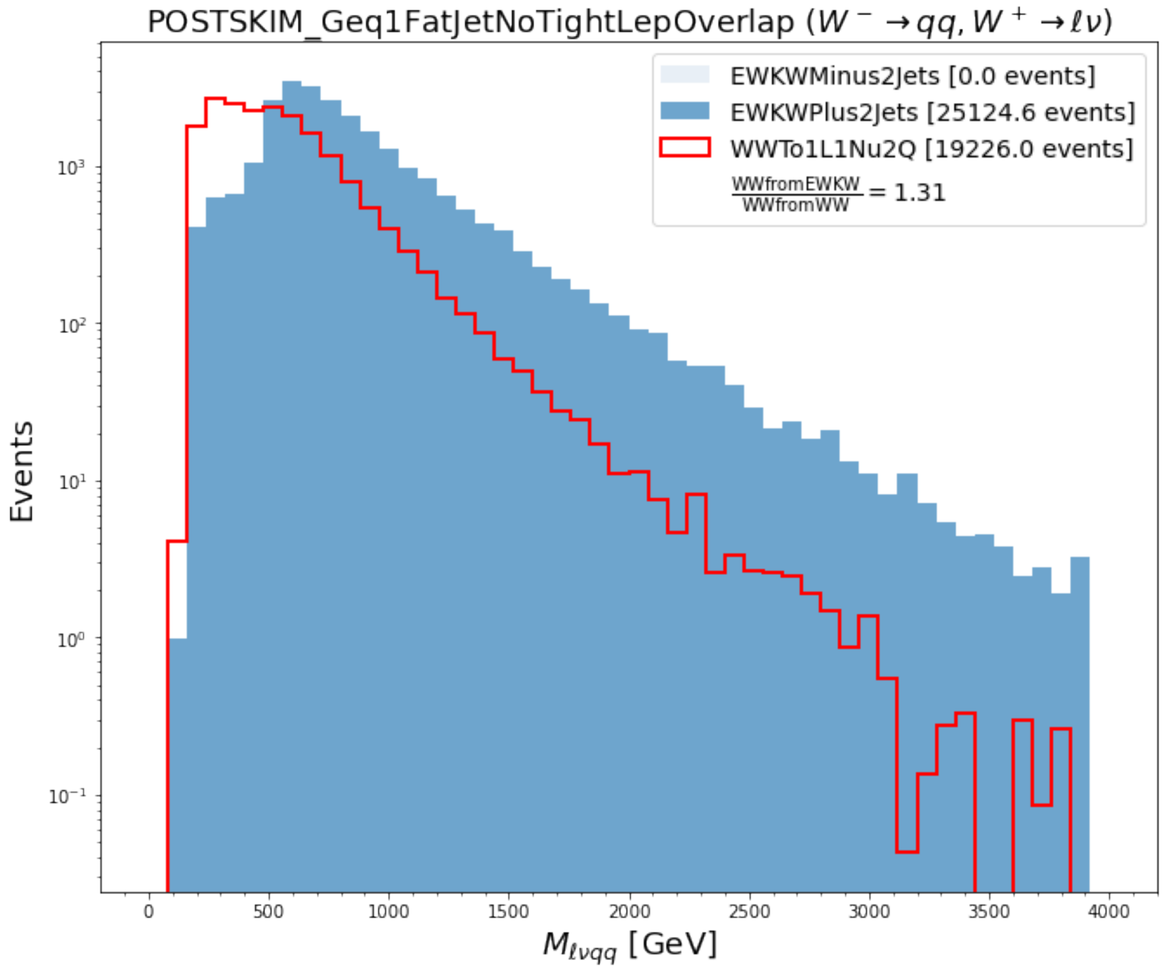


UC San Diego

Comparing Diboson Contributions

Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

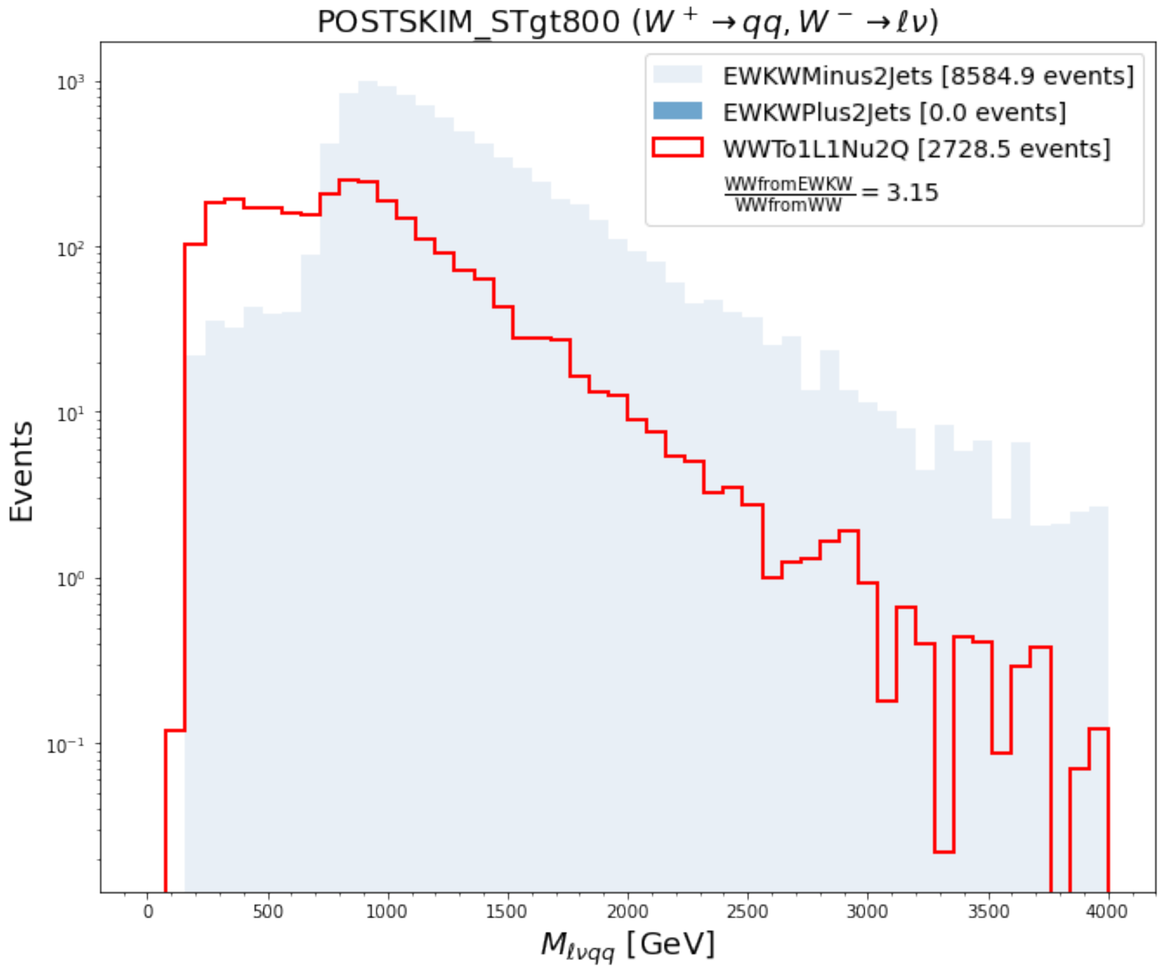


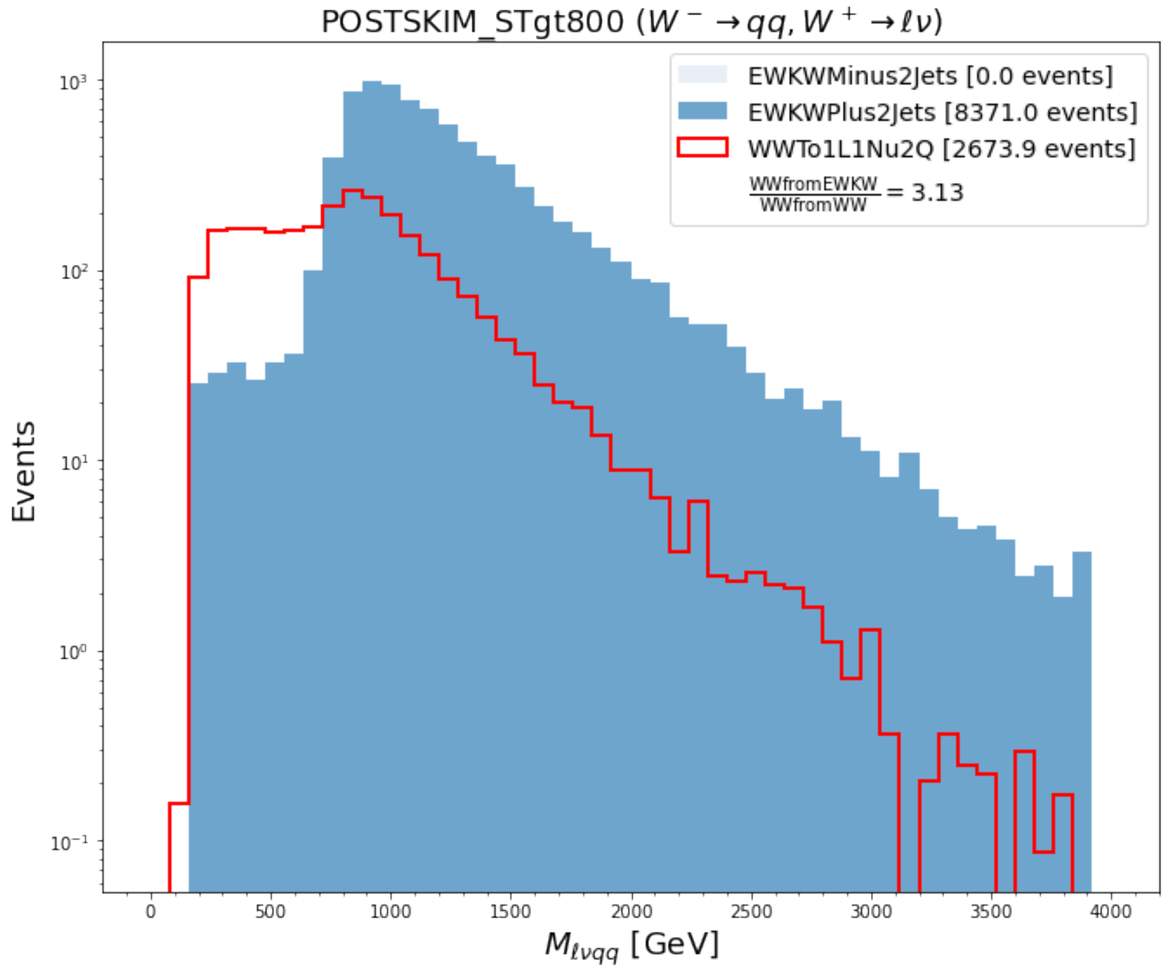


Cutoff gets worse after stricter fatjet selection

Comparing Diboson Contributions

Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$





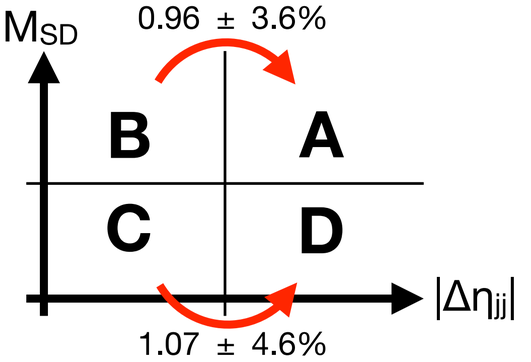
Cutoff at 700 GeV established after S_T selection

UC San Diego

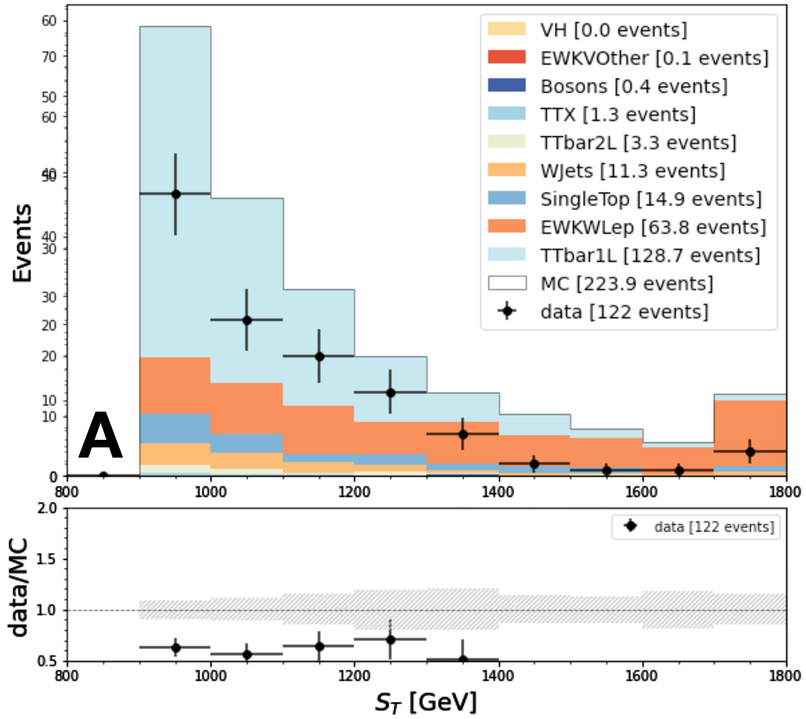
Returning to the Original Issue

Presel. (w/out $\Delta\eta_{jj}$ cut) AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9

Cut	Region	Bkg. (wgt)	Bkg. Err.*	Sig. (wgt)	Sig. Err.*	Data	Data Err.*
$\lvert\Delta\eta_{jj}\rvert$ > 4 AND M_{SD} ≥ 150 GeV	A	223.89	5.92	9.18	1.25	122	11.05
$\lvert\Delta\eta_{jj}\rvert$ ≤ 4 AND M_{SD} ≥ 150 GeV	B	232.05	5.59	1.00	0.45	179	13.38
$\lvert\Delta\eta_{jj}\rvert$ ≤ 4 AND M_{SD} < 150 GeV	C	174.37	4.85	14.07	1.55	142	11.92
$\lvert\Delta\eta_{jj}\rvert$ > 4 AND M_{SD} < 150 GeV (SR1)	D	186.76	6.90	371.20	8.16	—	—



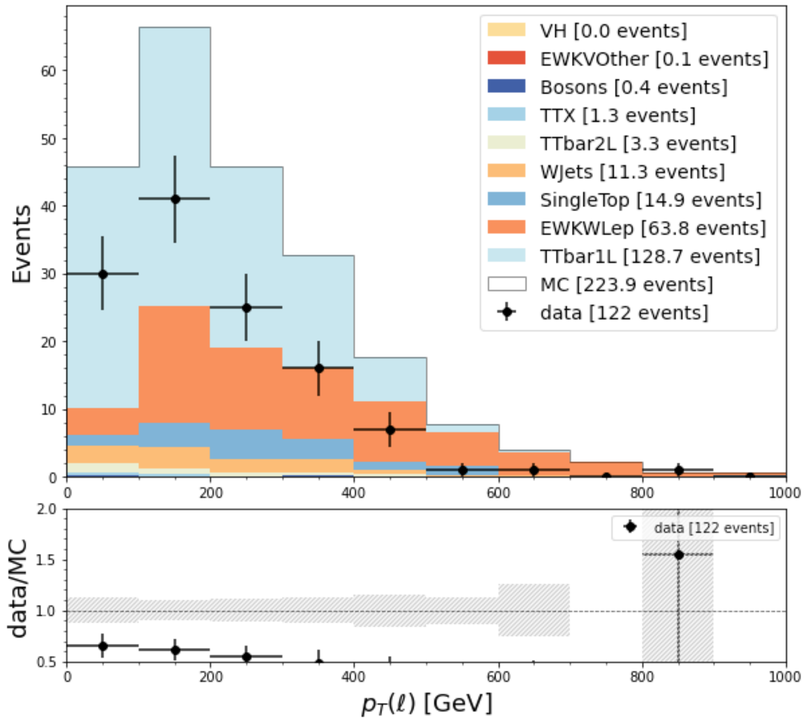
- Mostly problematic because it destroys data/MC agreement in Region A for SR1 ABCD
  - Preselection w/out $\Delta\eta_{jj}$ cut:
    M_{jj} > 500 GeV AND Hbb jet PNet Xbb score > 0.3
    AND AK4 b-jet veto
  - Especially bad at high S_T
- ST = $p_T(\ell)$ + MET + p_T(Hbb jet)
  - Is there a specific culprit?

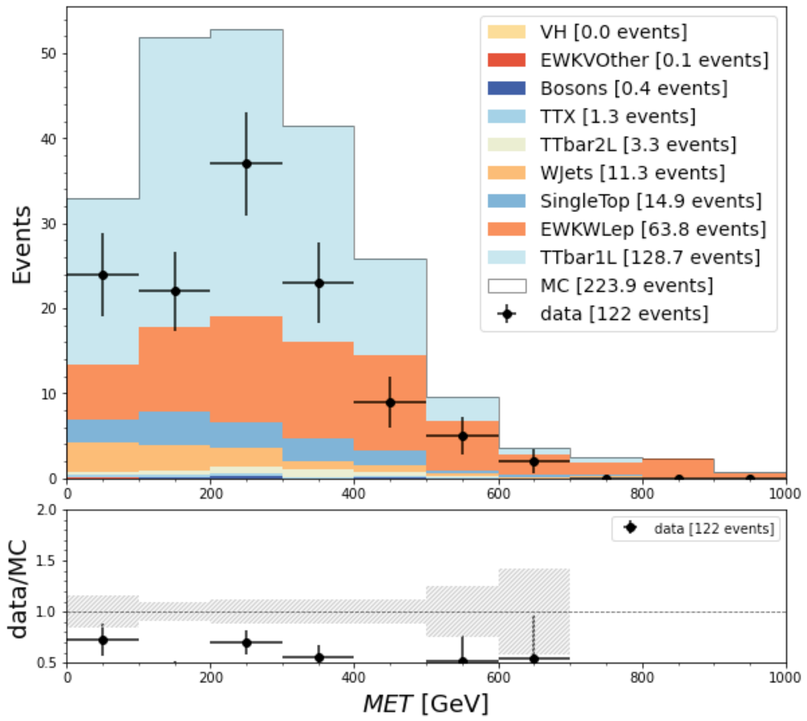


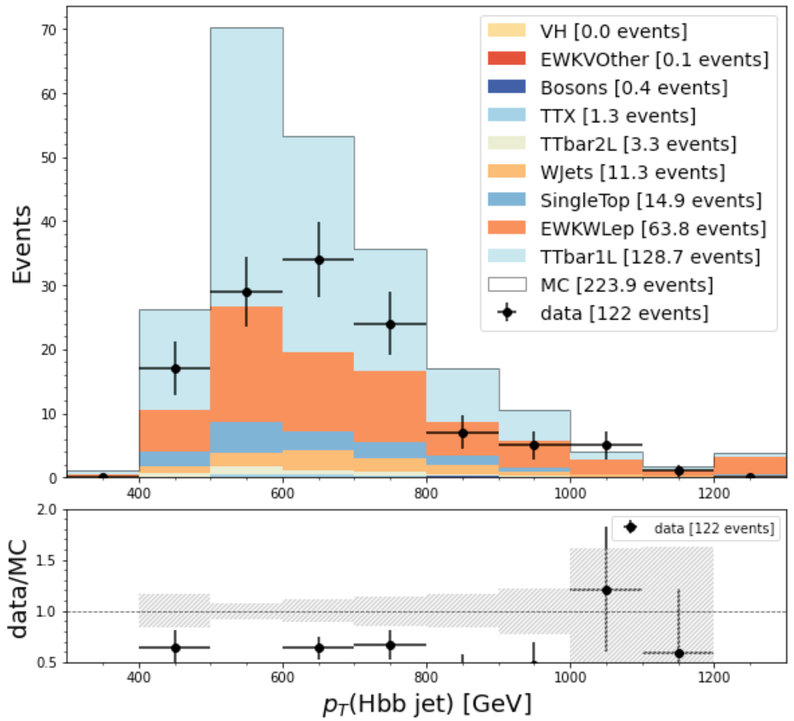
*err = $\sqrt{(\Sigma_i w_i^2)}$ for MC, $\sqrt{(\text{count})}$ for data

UC San Diego

Returning to the Original Issue

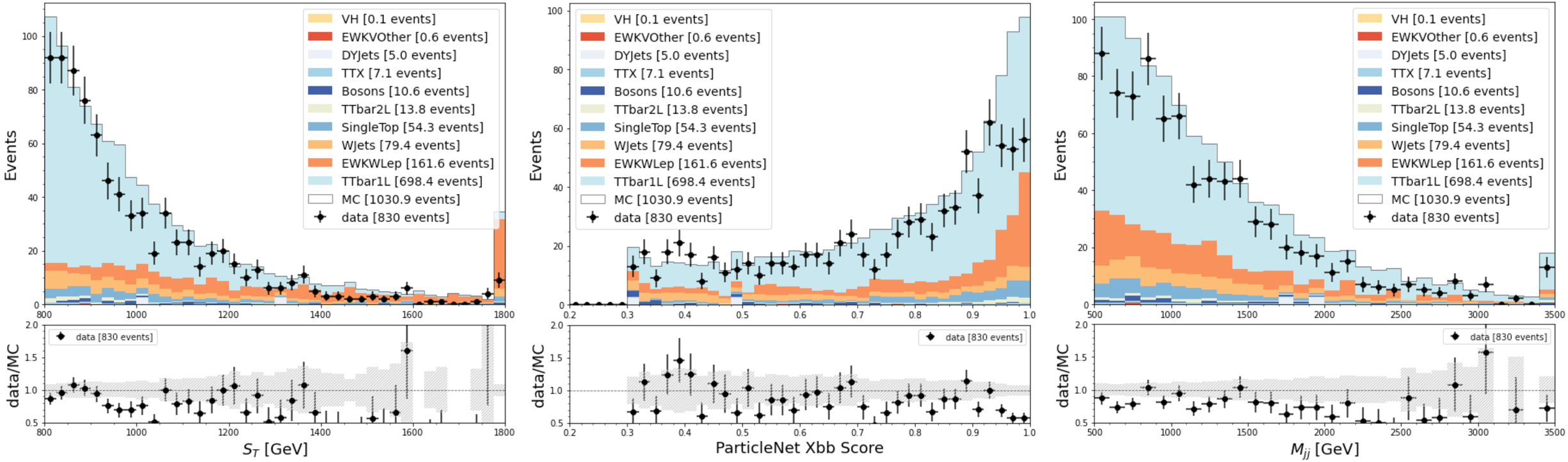






- Bad agreement across the board
- No smoking gun (i.e. good agreement outside of one specific region) here

Returning to the Original Issue

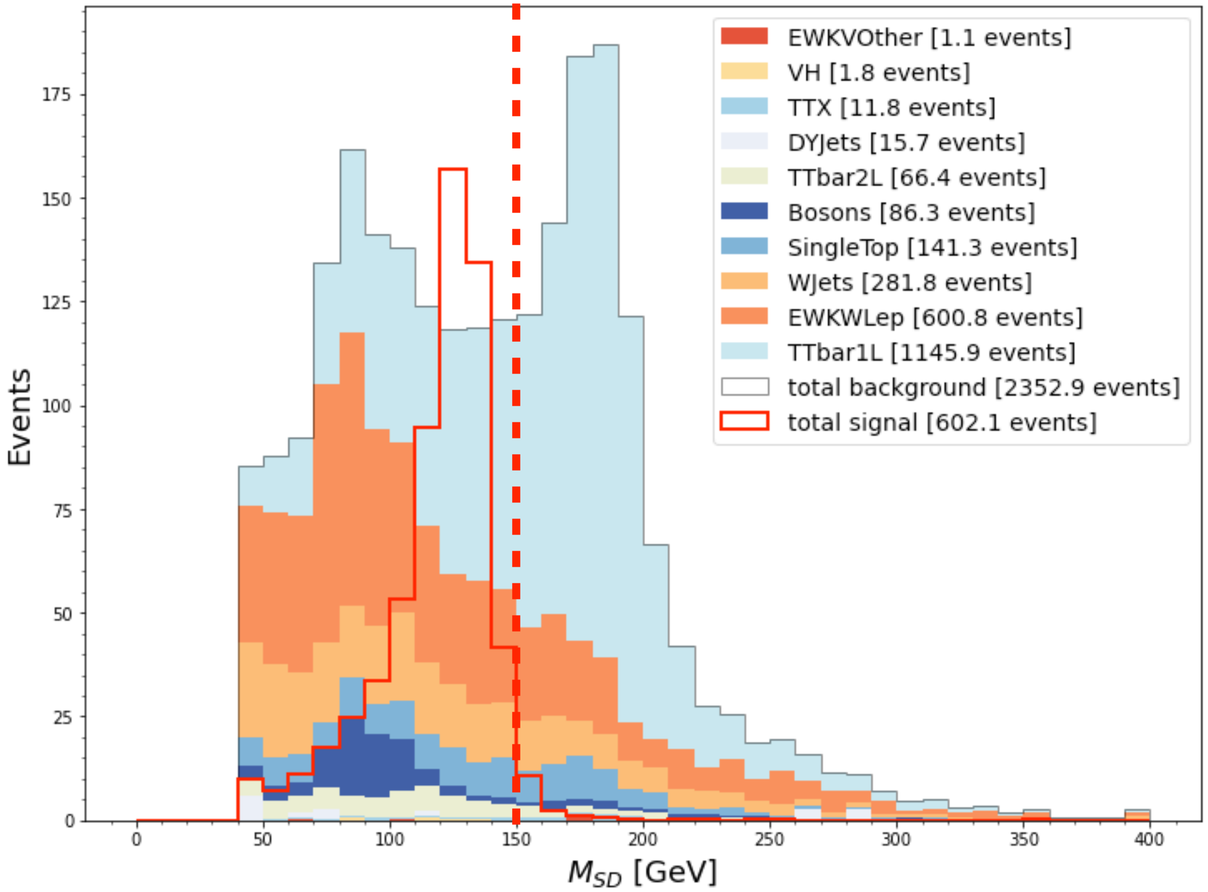


- Walk back S_T, PNet Xbb score, and M_{jj} cuts in Region A
- EWK W contribution leads to very bad data/MC agreement in high S_T, high Xbb score reg.

UC San Diego

Returning to the Original Issue

- Note: all of the data/MC plots we have looked at are in the M_{SD} sideband
- Much of the EWK W is concentrated in the M_{SD} SR
  - Can't check this region in data/MC without inverting $\Delta\eta_{jj}$ cut
  - Doesn't matter for ABCD, since Hbb-like regions (regions C and D) have this inverted $\Delta\eta_{jj}$ cut



Conclusions

- Weird feature at $M_{\ell\nu qq} \approx 700$ GeV in EWK W sample is develops after fatjet selection
- Must remember that the comparison to dedicated WW was to check whether EWK W sample made sense at all
  - Conclusion here is not clear to me
  - Otherwise, WW contamination in EWK W is…
    - …10-15% of EWK W causing SR1 ABCD issue
    - …50% of EWK W causing SR2 issue
    - …not enough to resolve our problem

Next Steps

- **Option 1:** double down on validating EWK W sample
  - Drop all reco cuts, compare data/MC cut-by-cut
  - Try to identify exactly where data/MC goes south and why
- **Option 2:** try to find phase space where EWK W is less problematic
  - Leonardo noticed EWK W VBS jet fakes Hbb selection
  - Could possibly refine Hbb/VBS selections to reduce EWK W contribution

Backup

*Professor
†Post-doc
‡Graduate Student
§Undergraduate Student

Who is involved?

- UC San Diego:
  A. Arora‡, P. Chang†, L. Giannini†, J. Guiang‡, F. Würthwein*, Y. Xiang‡, A. Yagil*

- U. Nebraska:
  F. Golf*

- Boston University:
  D. Spitzbart†, I. Suarez*

- UC Santa Barbara:
  C. Campagnari*

UC San Diego

University of Nebraska Lincoln

Boston University

UC Santa Barbara

CMS

UC San Diego

BSM Signal Models

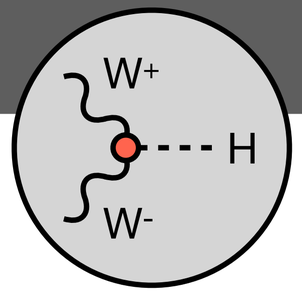
$\kappa_W = -1$

models/sm/couplings.py

```
GC_72 = Coupling(name = 'GC_72',
                 value = '(ee**2*complex(0,1)*vev)/(2.*sw**2)',
                 value = '-((ee**2*complex(0,1)*vev)/(2.*sw**2))',
                 order = {'QED':1})
```

models/sm/vertices.py

```
V_52 = Vertex(name = 'V_52',
              particles = [ P.W__minus__, P.W__plus__, P.H ],
              color = [ '1' ],
              lorentz = [ L.VVS1 ],
              couplings = {(0,0):C.GC_72})
```



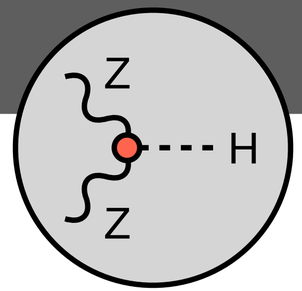
$\kappa_Z = -1$

models/sm/couplings.py

```
GC_81 = Coupling(name = 'GC_81',
                 value = 'ee**2*complex(0,1)*vev + … ',
                 value = '-(ee**2*complex(0,1)*vev + … )',
                 order = {'QED':1})
```

models/sm/vertices.py

```
V_69 = Vertex(name = 'V_69',
              particles = [ P.Z, P.Z, P.H ],
              color = [ '1' ],
              lorentz = [ L.VVS1 ],
              couplings = {(0,0):C.GC_81})
```

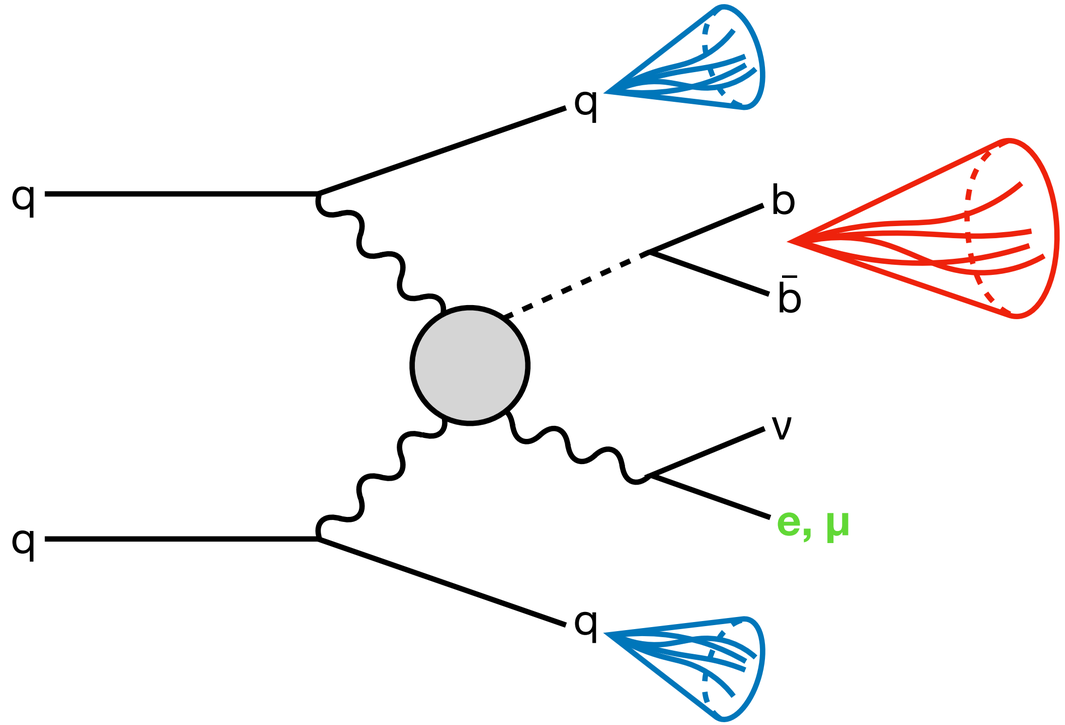


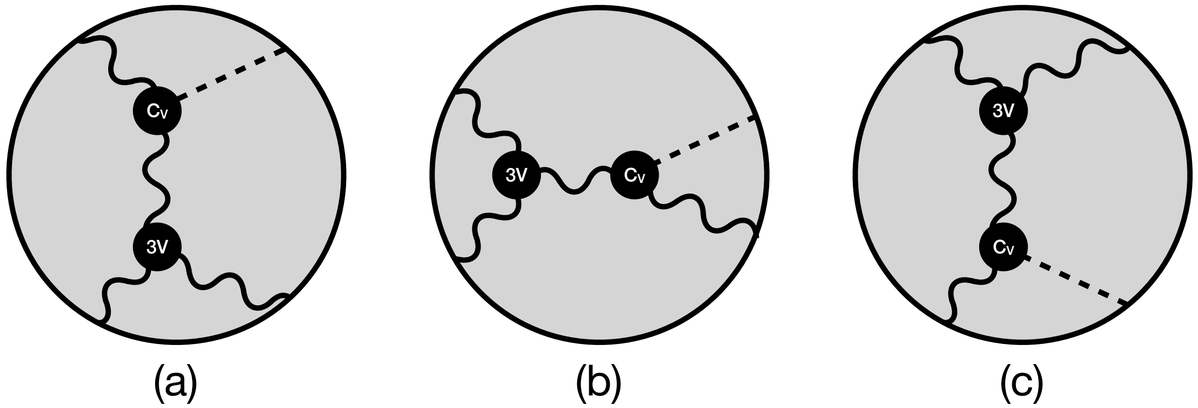
Only changed one line in SM Madgraph model!

UC San Diego

Target Final State

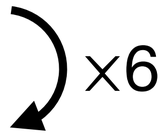
- Targeting **VBS** WH→**ℓνbb̄**
- Sensitive to $\kappa_V \rightarrow \lambda_{WZ} = \kappa_Z/\kappa_W$
  - Handle for ruling out $\lambda_{WZ} = -1$ (BSM)
- VBS WH BSM kinematics:
  - High-p_T H and W (high S_T)
  - VBS jets with large $\Delta\eta_{jj}$, M_{jj}





VBS WH Cross Sections

Model	σ [pb]
$\kappa_W = \kappa_Z = +1$ (SM)	0.075
$\kappa_W = -1$, $\kappa_Z = +1$	0.433
$\kappa_W = +1$, $\kappa_Z = -1$	0.433



- Setting $\kappa_W = -1$ or $\kappa_Z = -1$ equivalently enhances cross section by a factor of 6
- These numbers are taken from MadGraph*: `generate p p > w h j j QCD=0`
  - Includes gen-level filters (e.g. jet $p_T > 10$ GeV)
  - Generated 10,000 events for each to obtain xsec value
- **Optimizing for $\kappa_W = -1$** (kinematics are equivalent to $\kappa_Z = -1$)
  - Generated 100k UL NanoAOD events for 2016 pre-VFP, 2016 post-VFP, 2017, and 2018

*More details in backup

UC San Diego

Analysis Skim

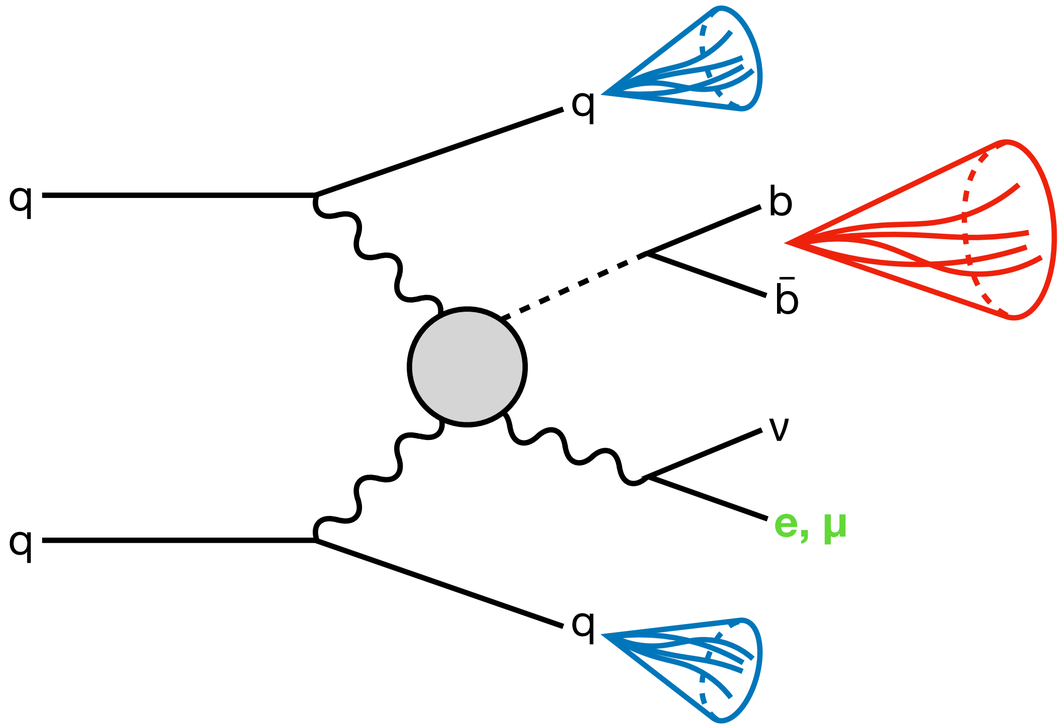
Yields scaled to lumi×σ, rounded for readability

Cut	VH	VV/VVV/VBSWZ	W+Jets	SingleTop	TTbar+X	TTbar1L	TTbar2L	TotalBkg	Eff.*	VBSWH (λ_{WZ} = -1)	Eff.*
Skim	232	6.5K	122K	7.6K	1K	86K	10K	**233K**	—	**2K**	—

Object	Selection
Leptons (μ, e)	== 1 loose AND == 1 tight*
Fat Jets	≥ 1 w/ p_T > 250 GeV AND mass > 50 GeV AND M_{SD} > 40 GeV AND ΔR(fat jet, tight lepton) > 0.8
Jets	≥ 1 w/ p_T > 20 GeV AND ΔR(jet, any veto lepton) > 0.4
Other	S_T > 800 GeV

*Using the ttH lepton ID

VBS WH Objects



Object	Selections
One **Lepton** (loose veto in postskim)	• Passes ttH tight ID • p_T > 40 GeV
One **Hbb** Ak8 Jet	• ΔR(Ak8 jet, any veto lep) > 0.8 • p_T > 250 GeV • mass > 50 GeV • M_{SD}* > 40 GeV • Has max(PNet Xbb vs. QCD score) • PNet Xbb vs. QCD score > 0.3

Two **VBS** Ak4 jets	• ΔR(Ak4 jet, any veto lep) > 0.4
	• ΔR(Ak4 jet, Hbb Ak8 jet) > 0.8
	• p_T > 30 GeV
	• Passes tight jet ID
	• For > 2 candidates**:
	  • All in same η-hemisphere: leading/trailing (in P) jets
	  • Else: leading (in P) jet from each η-hemisphere

*M_{SD} = soft drop mass
**Hereafter referred to as the “Max-E” selection

UC San Diego

Checking the Production Steps

- wmLHEGEN:
  https://cms-pdmv.cern.ch/mcm/public/restapi/requests/get_setup/HIG-RunIISummer20UL18wmLHEGEN-00832
- SIM:
  https://cms-pdmv.cern.ch/mcm/public/restapi/requests/get_setup/HIG-RunIISummer20UL18SIM-00481
- DIGIPREMIX:
  https://cms-pdmv.cern.ch/mcm/public/restapi/requests/get_setup/HIG-RunIISummer20UL18DIGIPremix-00481
- HLT:
  https://cms-pdmv.cern.ch/mcm/public/restapi/requests/get_setup/HIG-RunIISummer20UL18HLT-00481
- RECO:
  https://cms-pdmv.cern.ch/mcm/public/restapi/requests/get_setup/HIG-RunIISummer20UL18RECO-00481
- MiniAOD (v2):
  https://cms-pdmv.cern.ch/mcm/public/restapi/requests/get_setup/HIG-RunIISummer20UL18MiniAODv2-00571
- NanoAOD (v9):
  https://cms-pdmv.cern.ch/mcm/public/restapi/requests/get_setup/HIG-RunIISummer20UL18NanoAODv9-00525

Comparing Diboson Contributions

Proc Card	Notes

`import model sm-ckm_no_b_mass`
 	
`define l+ = e+ mu+ ta+`
`define vl = ve vm vt`
 	
`generate p p > l+ vl j j / t t~ h QCD=0`
 	
`output` **`ewkwplvjj_5f_LO`** `-nojpeg`
	• Used to generate EWKWPlus…WToLNu samples
	  • From here
	• **LO** sample
	• Both W's can be off-shell
	• 5-flavor scheme

`import model loop_sm`
 	
`define ell+ = e+ mu+ ta+`
`define ell- = e- mu- ta-`
 	
`generate p p > ell+ vl w- $$ t t~ H [QCD] @0`
`add process p p > ell+ vl w- j $$ t t~ H [QCD] @1`
`add process p p > ell- vl~ w+ $$ t t~ H [QCD] @2`
`add process p p > ell- vl~ w+ j $$ t t~ H [QCD] @3`
 	
`output` **`WWTo1L1Nu2Q01j_4f_NLO_FXFX`** `-nojpeg`
	• Used to generate WWTo1L1Nu2Q samples
	  • From here
	• **NLO** sample
	• One W can be off-shell, the **other** must be on-shell
	• 4-flavor scheme

Comparing Diboson Contributions

EWKWPlus Run Card (from **here**)

```
# Matching - Warning! ickkw > 1 is still beta
0 = ickkw ! 0 no matching, 1 MLM, 2 CKKW matching

# Minimum and maximum pt's
10 = ptj ! minimum pt for the jets
10 = pta ! minimum pt for the photons

# Maximum and minimum absolute rapidity
6.5 = etaj ! max rap for the jets
2.5 = etaa ! max rap for the photons

# Minimum and maximum DeltaR distance
0.1 = draj ! min distance between gamma and jet
0.1 = dral ! min distance between gamma and lepton

# Minimum and maximum invariant mass for pairs
10 = mmjj ! min invariant mass of a jet pair

# Minimum and maximum invariant mass for all letpons
50 = mmnl ! min invariant mass for all letpons (l+- and vl)

# Photon-isolation cuts, according to hep-ph/9801442
# When ptgmin=0, all the other parameters are ignored
0 = ptgmin ! Min photon transverse momentum

# maximal pdg code for quark to be considered as a light jet
5 = maxjetflavor ! Maximum jet pdg code

# Store info for systematics studies
T = use_syst ! Enable systematics studies
```

WWTo1L1Nu2Q Run Card (from **here**)

```
# Merging - WARNING! Applies merging only at the hard-event level.
3 = ickkw ! 0 no merging, 3 FxFx merging

# Cuts on the jets
1 = jetalgo     ! FastJet jet algorithm (1=kT, 0=C/A, -1=anti-kT)
1.0 = jetradius ! The radius parameter for the jet algorithm
15 = ptj        ! Min jet transverse momentum

# Photon-isolation cuts, according to hep-ph/9801442
# When ptgmin=0, all the other parameters are ignored
20 = ptgmin    ! Min photon transverse momentum
0.4 = R0gamma  ! Radius of isolation code
1.0 = xn       ! n parameter of eq.(3.4) in hep-ph/9801442
1.0 = epsgamma ! epsilon_gamma parameter of eq.(3.4) in hep-ph/9801442
.true. = isoEM ! isolate photons from EM energy (photons and leptons)

# Maximal PDG code for quark to be considered a jet when applying cuts
4 = maxjetflavor
```

All cuts that MadGraph ignores have been excluded here
unless one run card explicitly excludes something the other does not

Comparing Diboson Contributions

- NanoAOD stores LHE-level particles (right)
  - For both:
    - 0, 1 = incoming partons
    - 2, 3 = outgoing ℓ, v from W
  - For EWK W samples (LO), 6 particles:
    - 4, 5 = outgoing q, q from VBS or extra W/Z
  - For WW samples (NLO), ≥ 6 particles:
    - 4, 5 = outgoing q, q from W
    - 6, … = extra q's or q's from NLO

Index	Particle	Status
0	q	Incoming
1	q	Incoming
2	ℓ	Outgoing
3	v	Outgoing
4	q	Outgoing
5	q	Outgoing
6	q/g	Outgoing
…	q/g	Outgoing

LHE particle record

Comparing Diboson Contributions

- Strategy:
  1. Select ℓ, v (index 2, 3)
  2. Store $M_{\ell v}$, $charge_{\ell}$
  3. Select q, q (index 4, 5)
  4. Store M_{qq}, $charge_{q}$
  5. Require $M_{\ell v} \in [70, 90)$ GeV
  6. Require $M_{qq} \in [70, 90)$ GeV
  7. Require $|charge_{qq}| == 1$

  Selects WW events

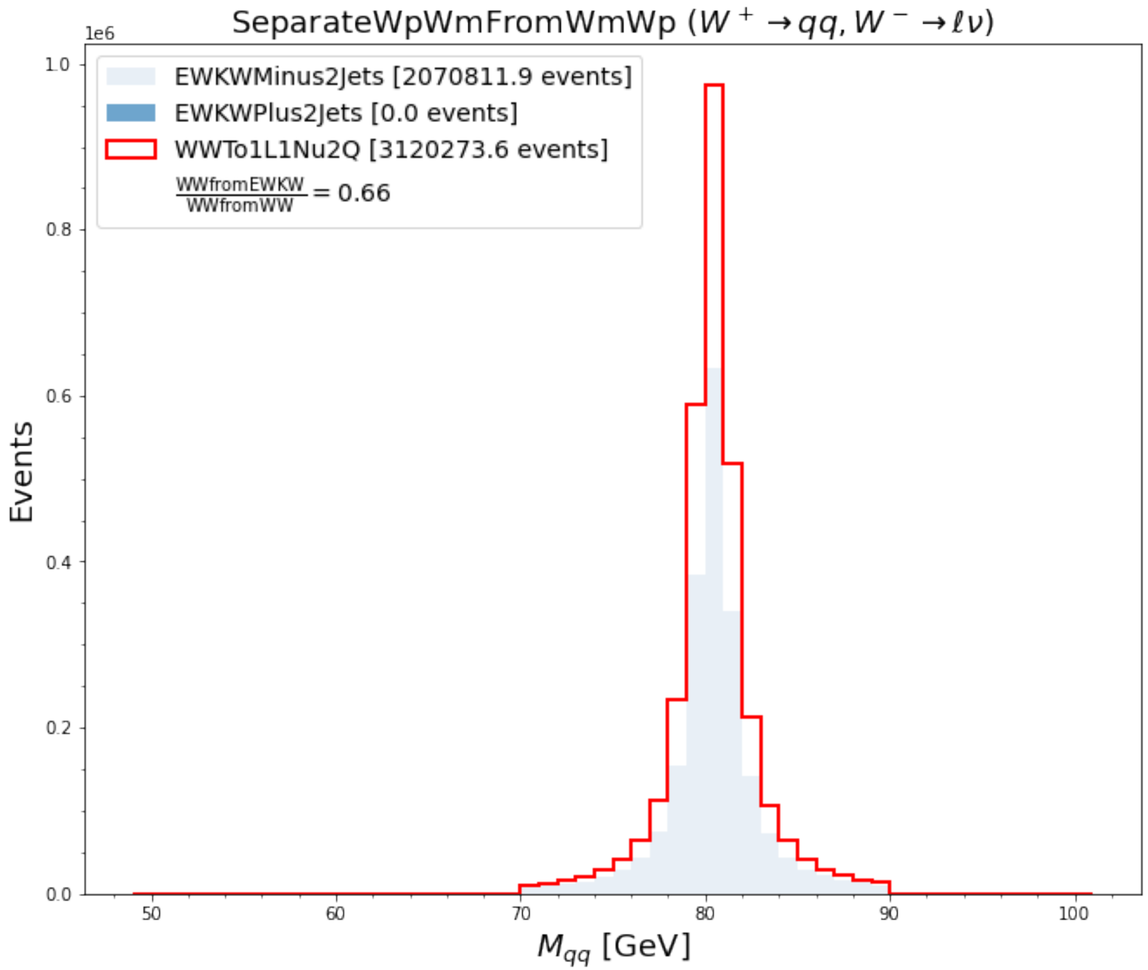
  8. Separate into W^-W^+, W^+W^- and compare

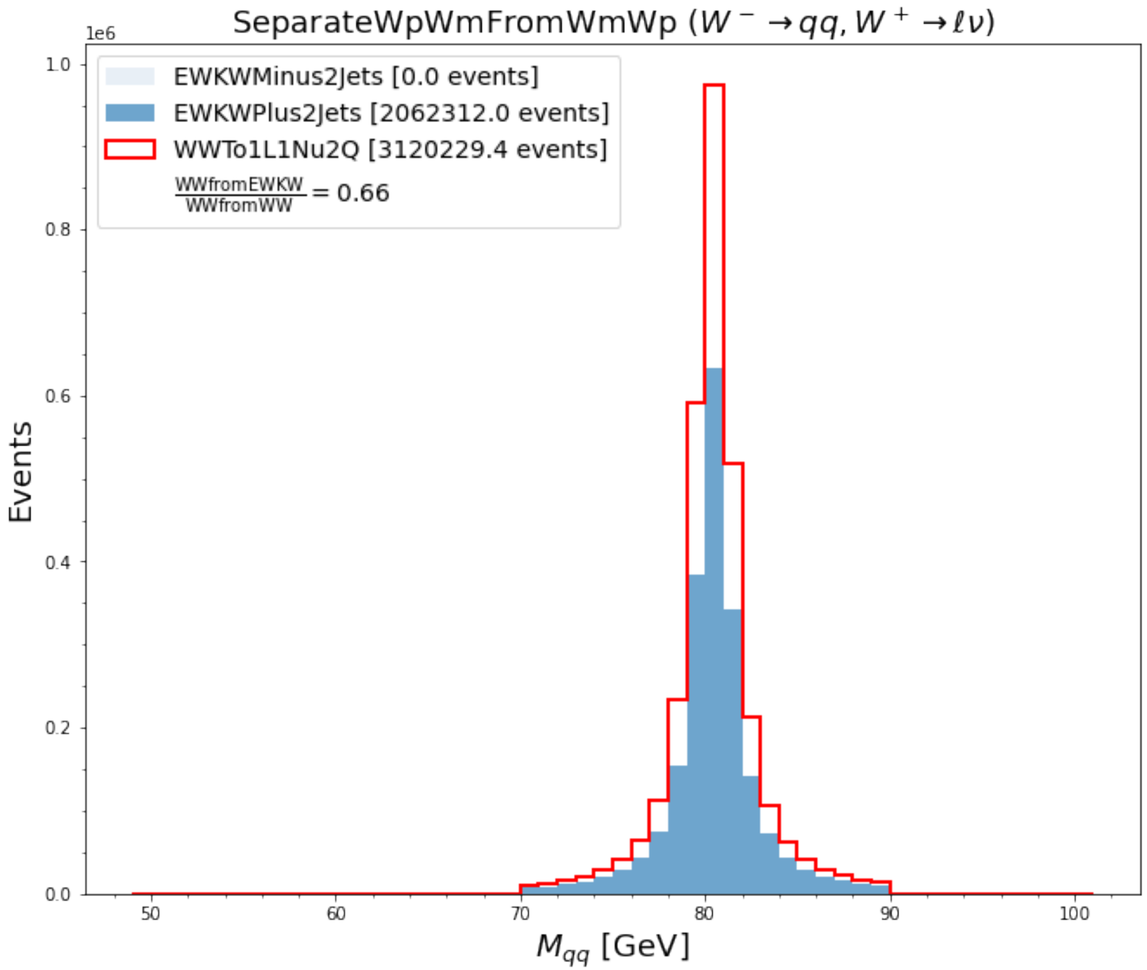
Index	Particle	Status
0	q	Incoming
1	q	Incoming
2	ℓ	Outgoing
3	v	Outgoing
4	q	Outgoing
5	q	Outgoing
6	q/g	Outgoing
…	q/g	Outgoing

LHE particle record

Comparing Diboson Contributions

Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

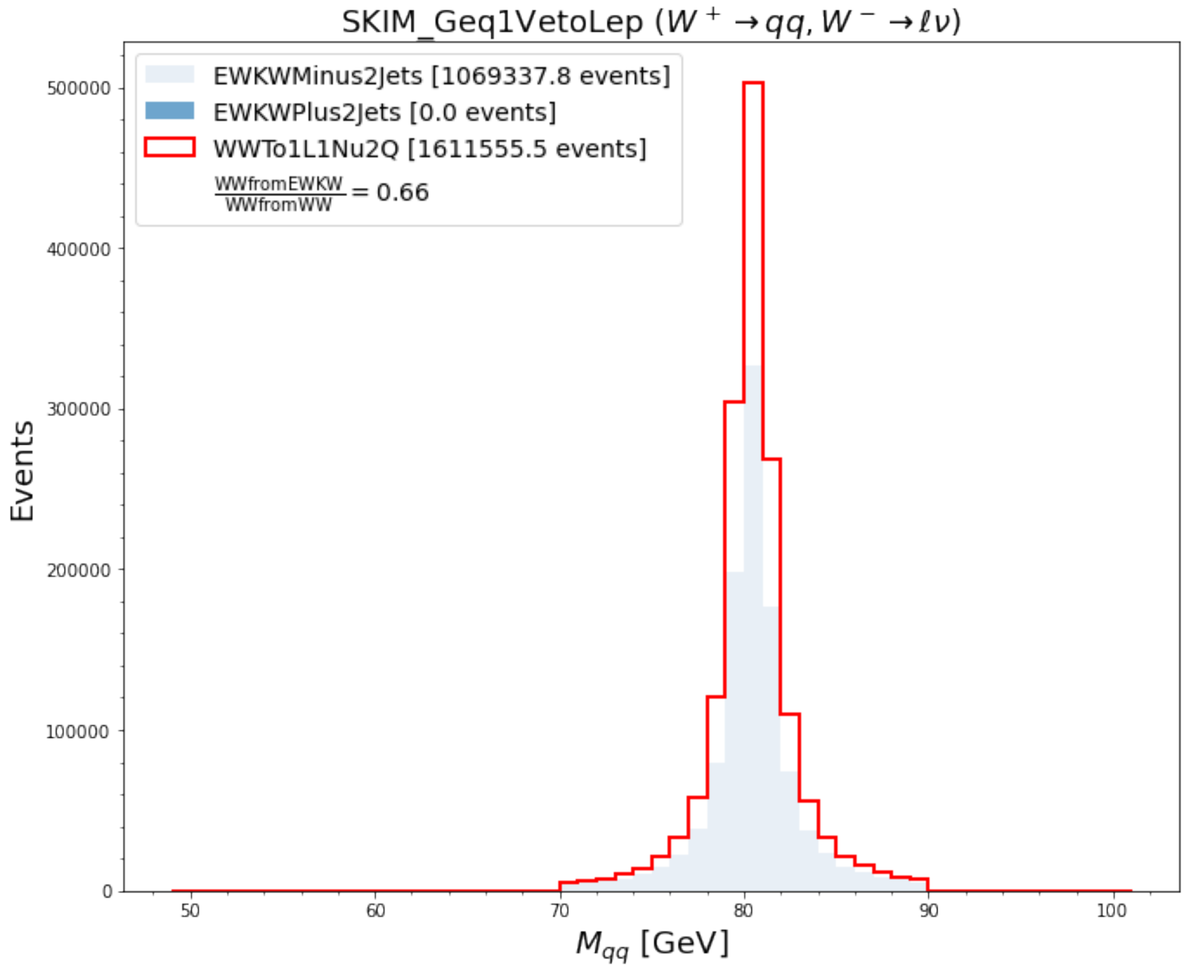


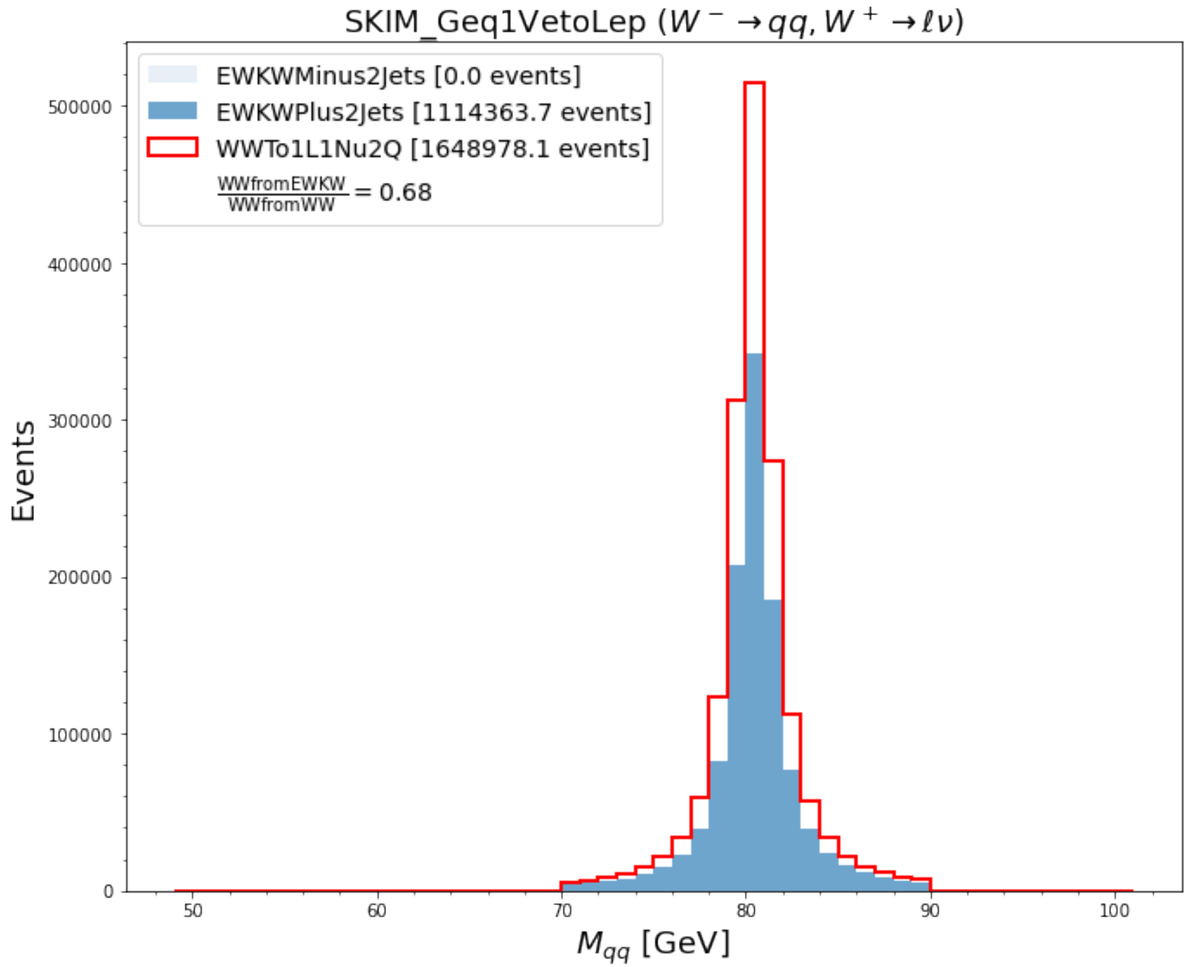


UC San Diego

Comparing Diboson Contributions

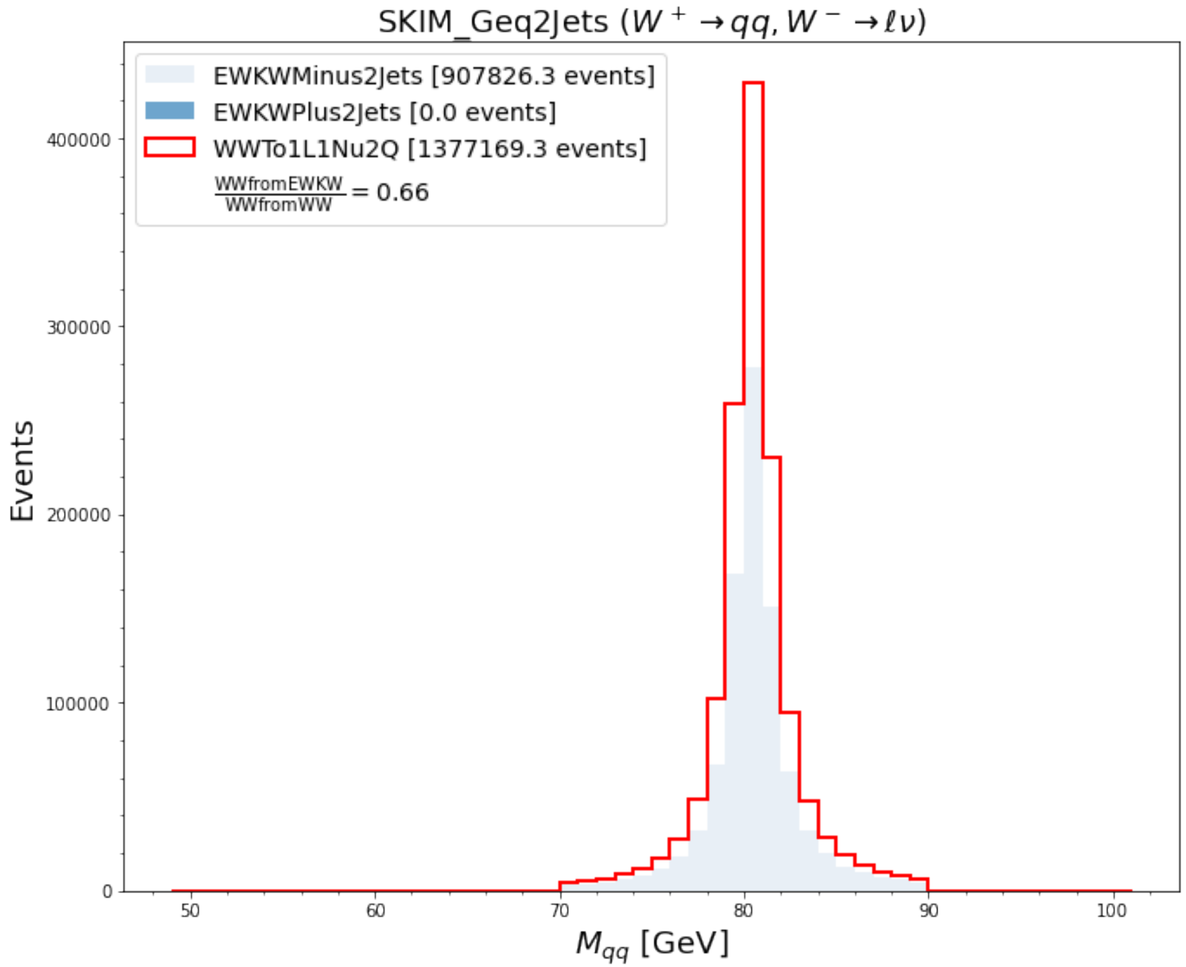
Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

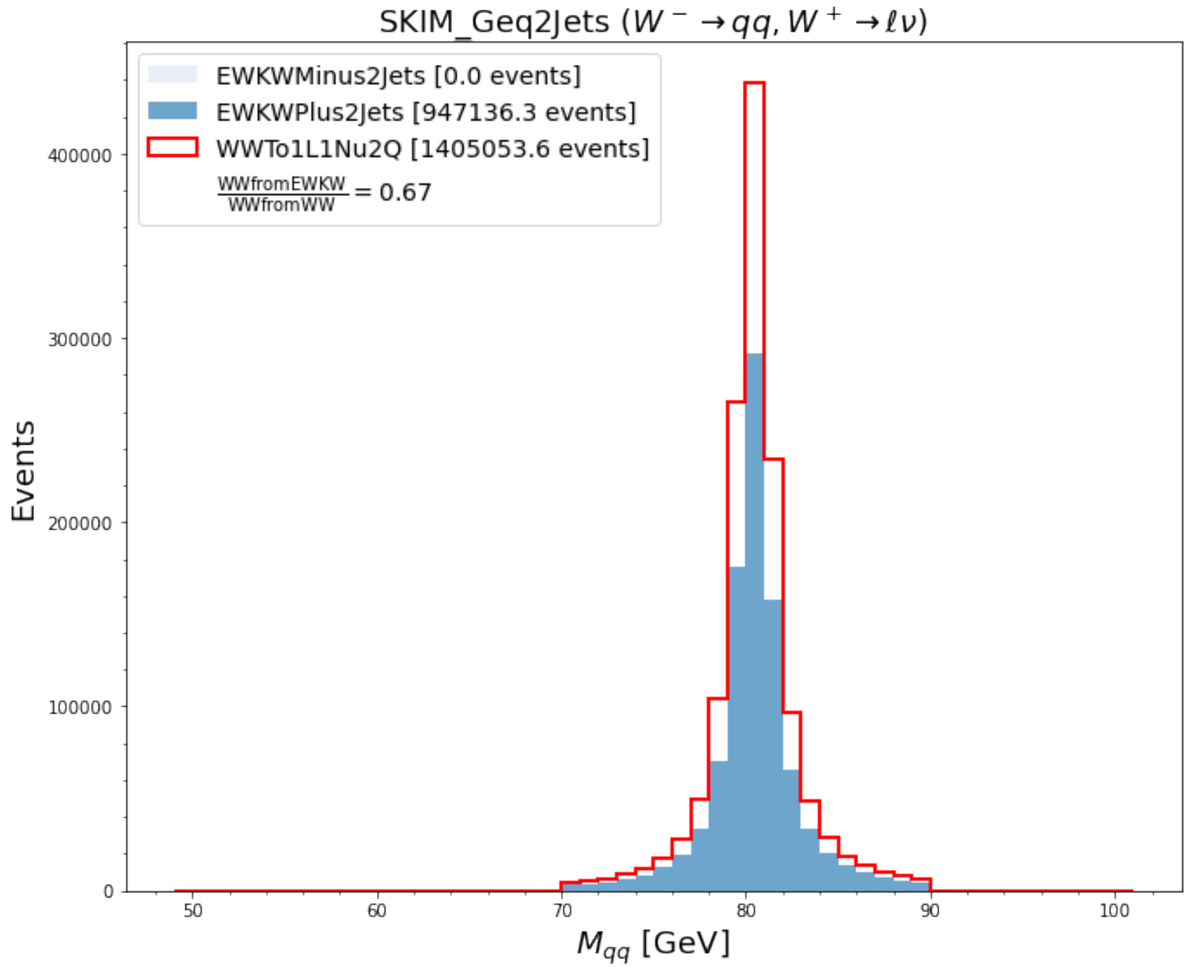




Comparing Diboson Contributions

Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

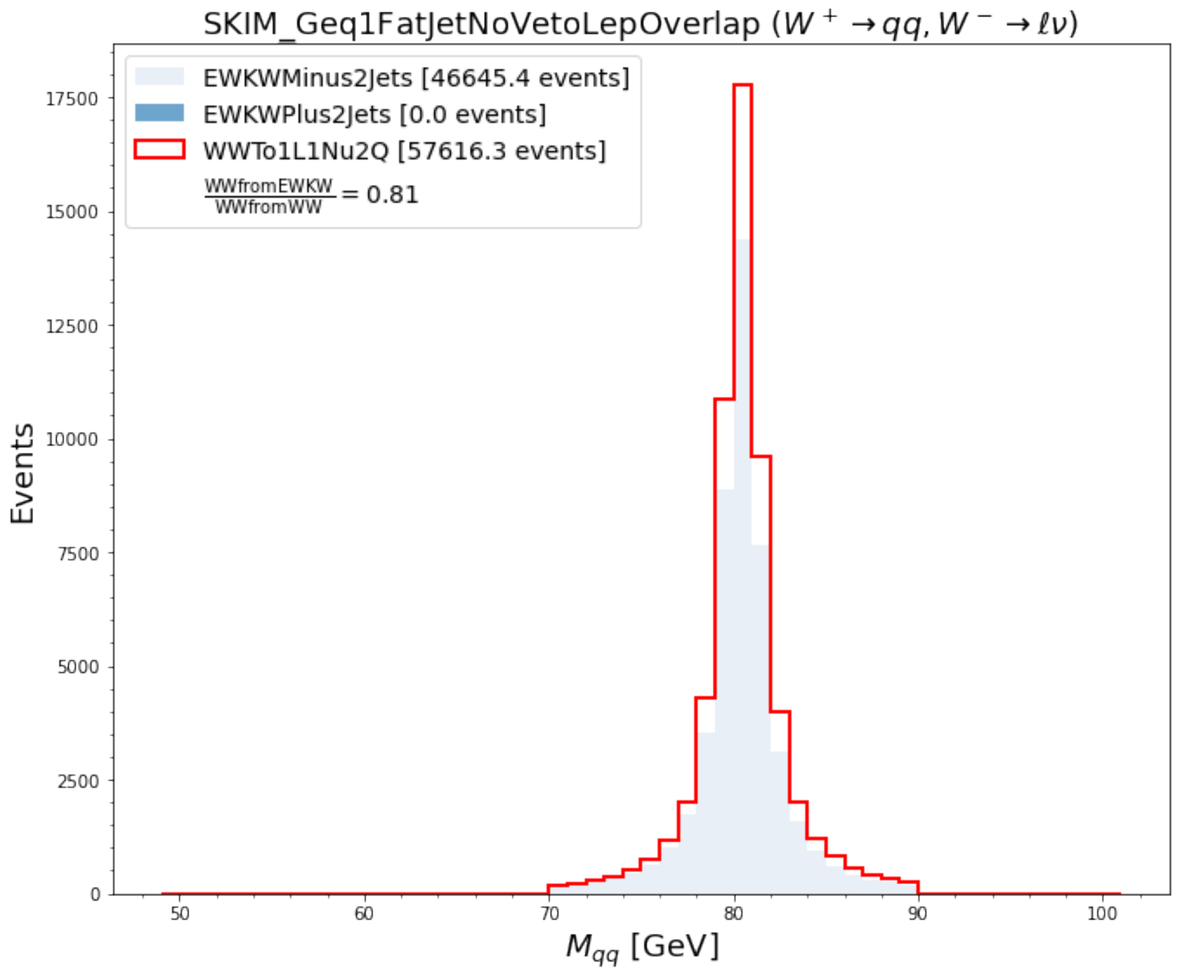


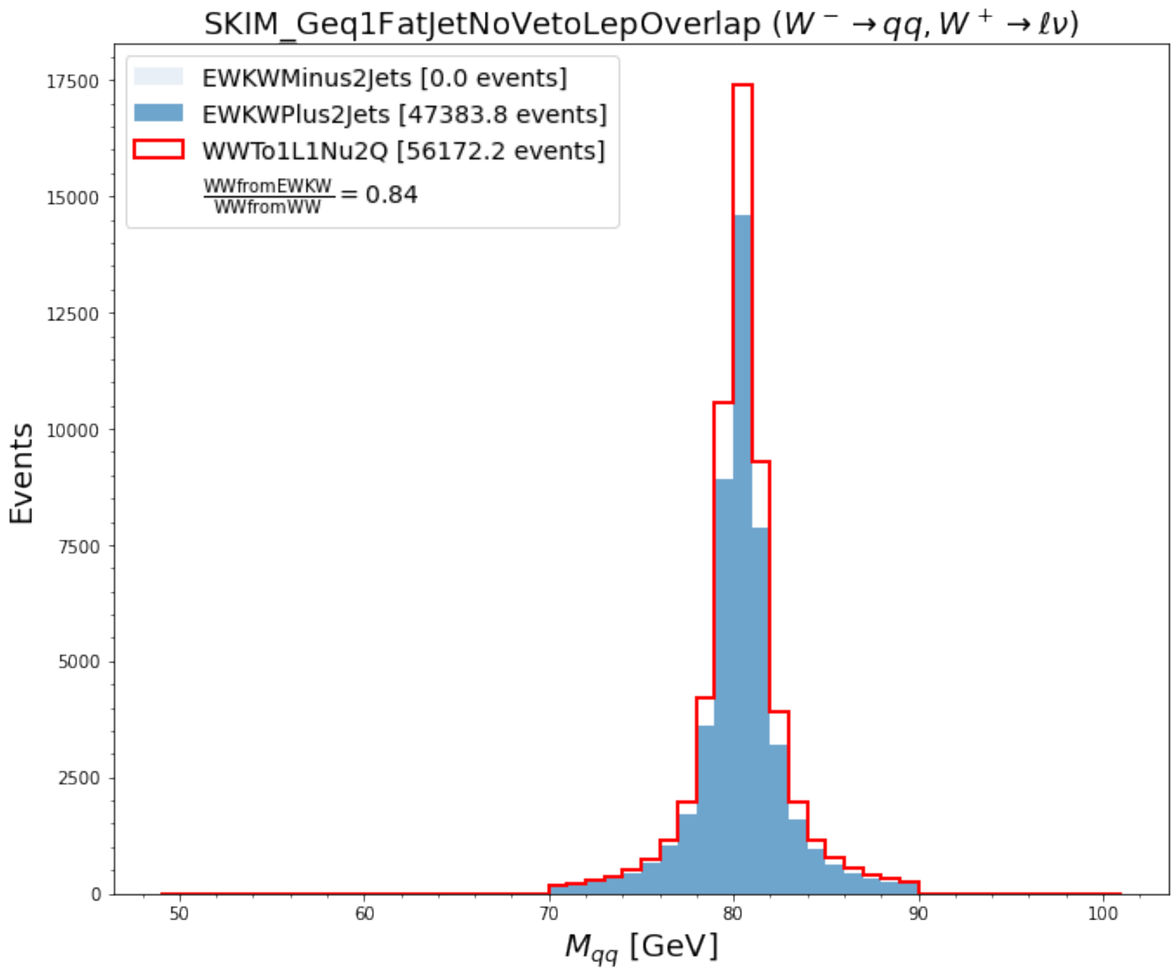


UC San Diego

Comparing Diboson Contributions

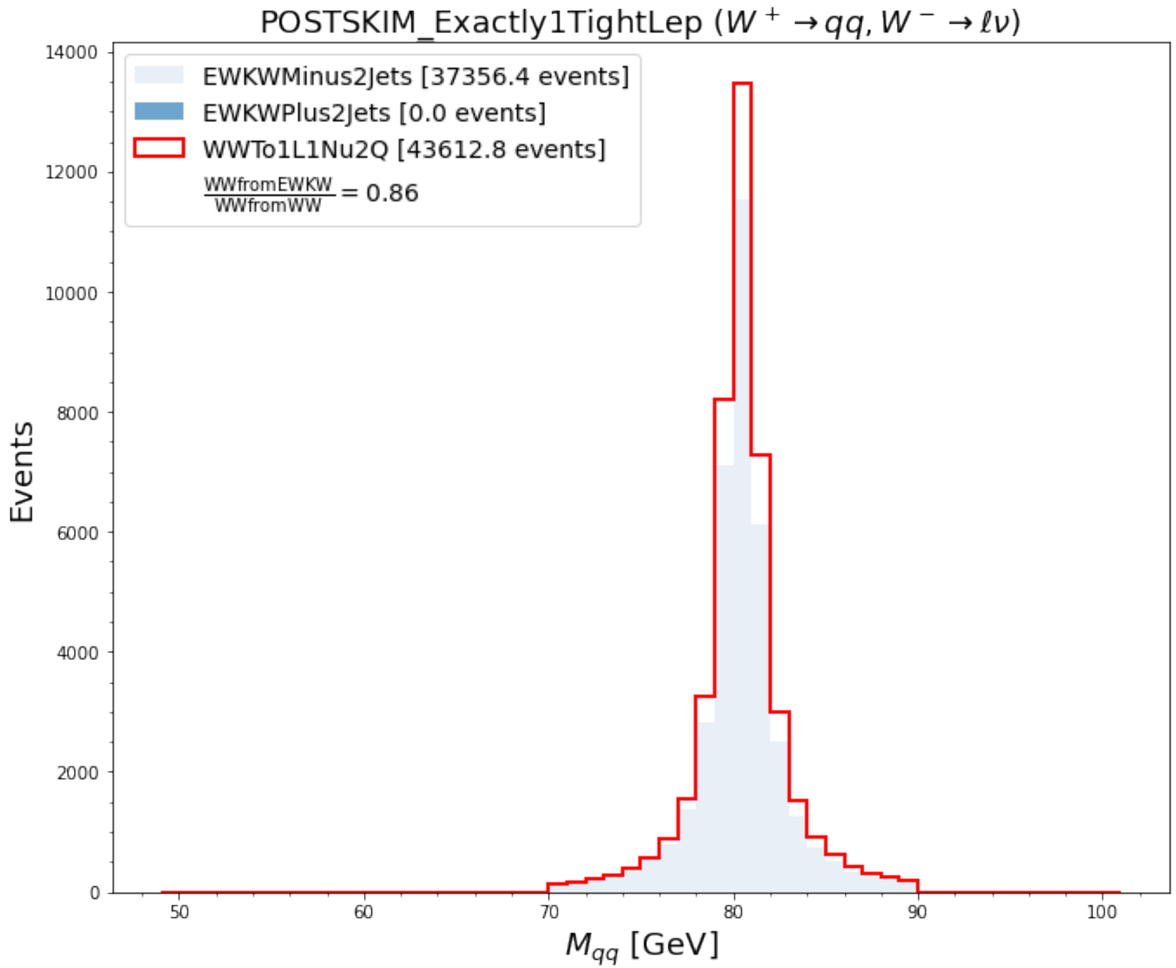
Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

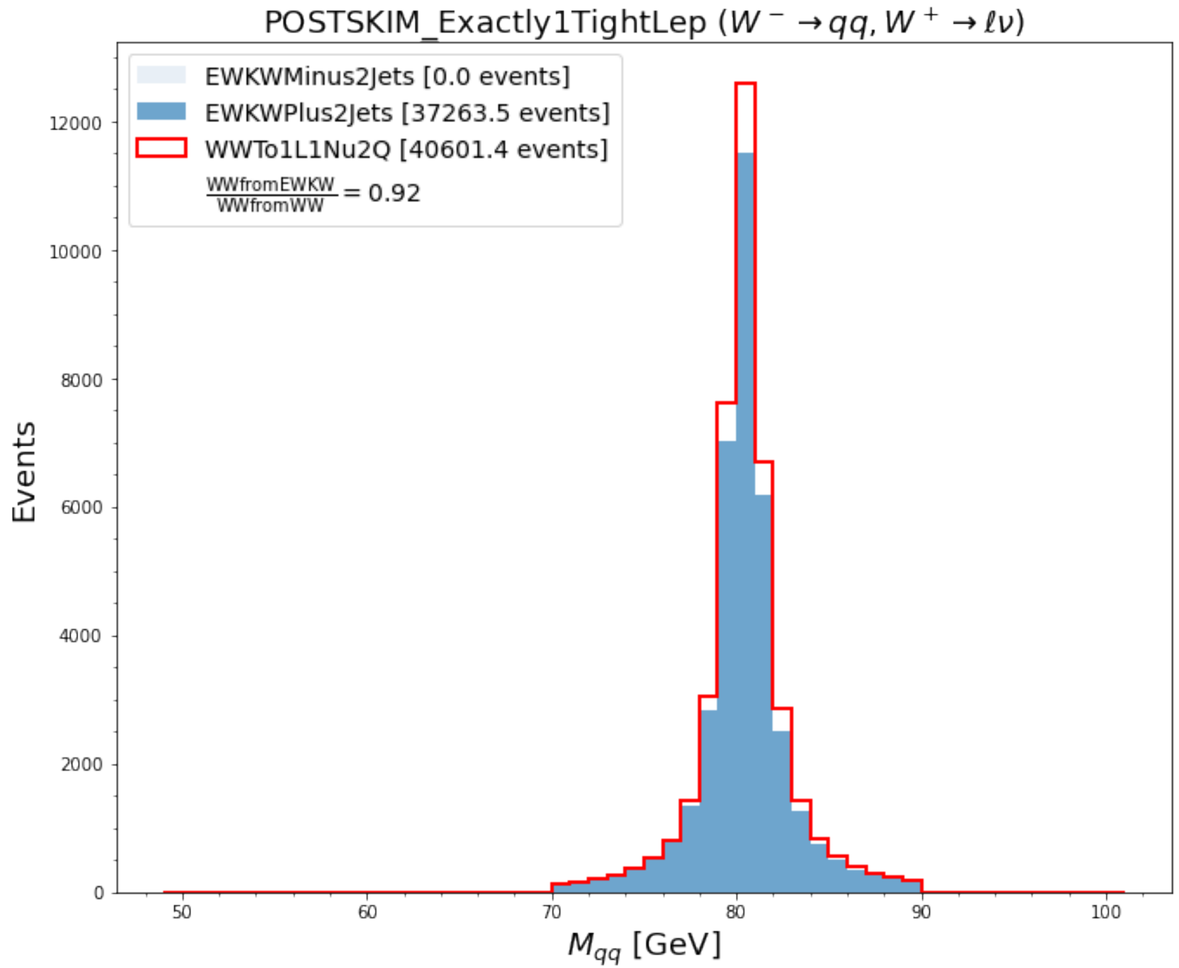




Comparing Diboson Contributions

Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

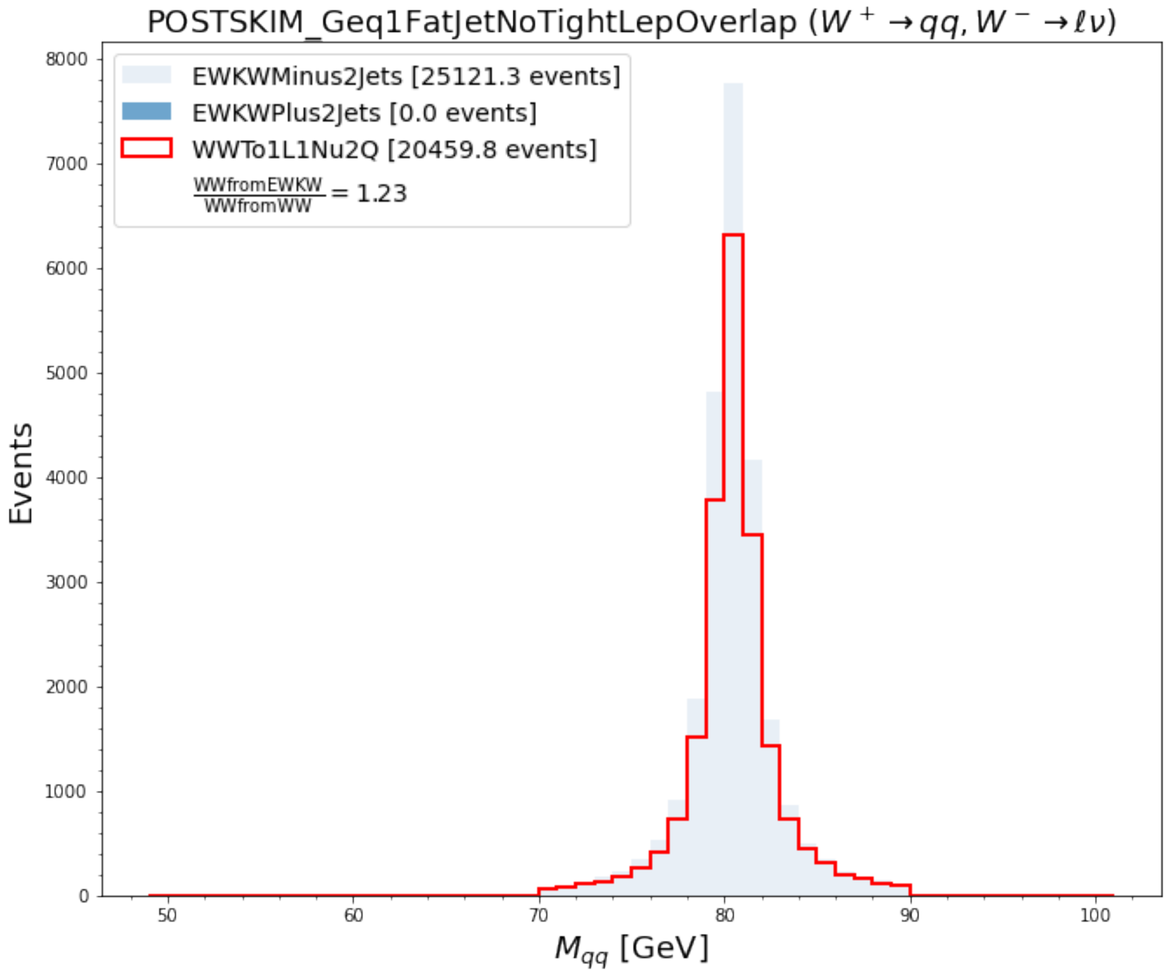


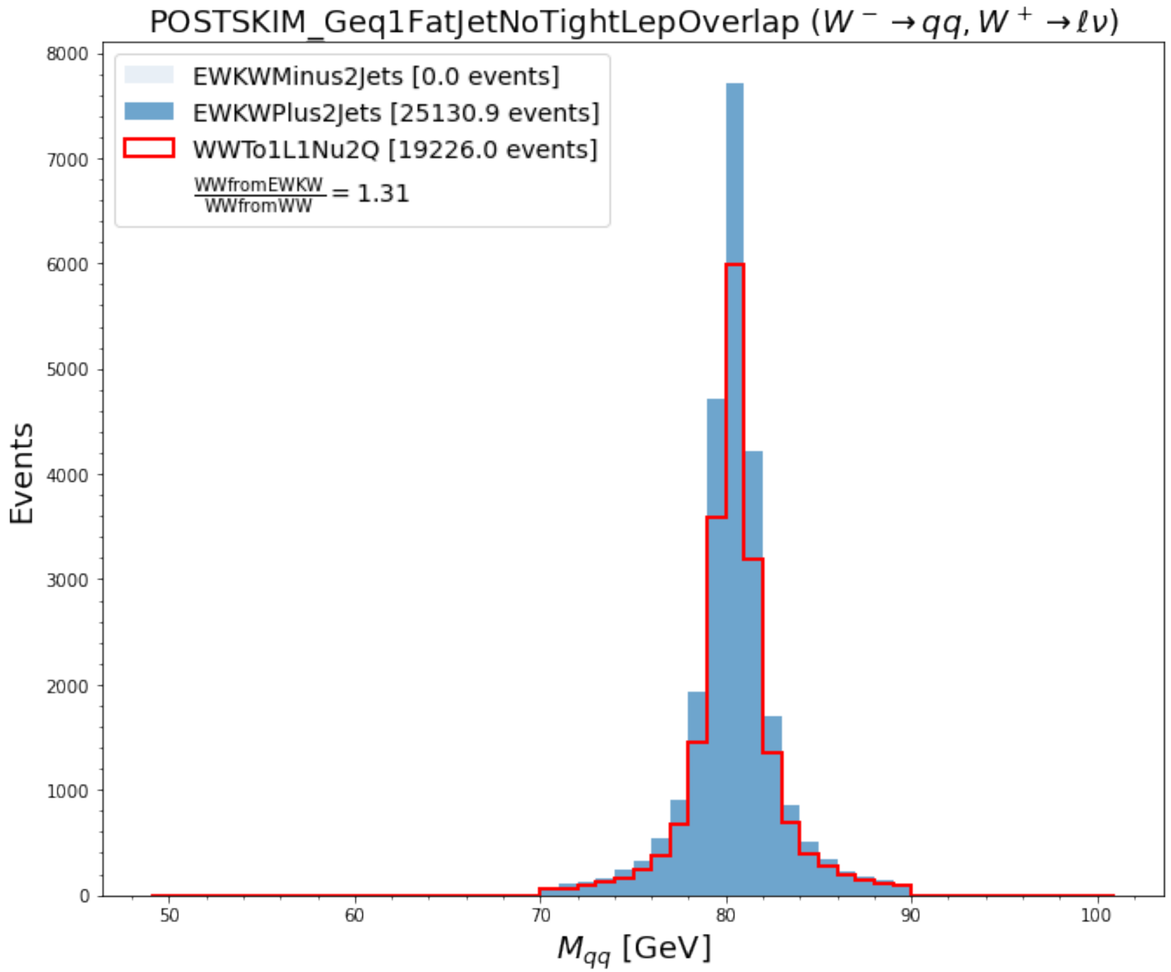


UC San Diego

Comparing Diboson Contributions

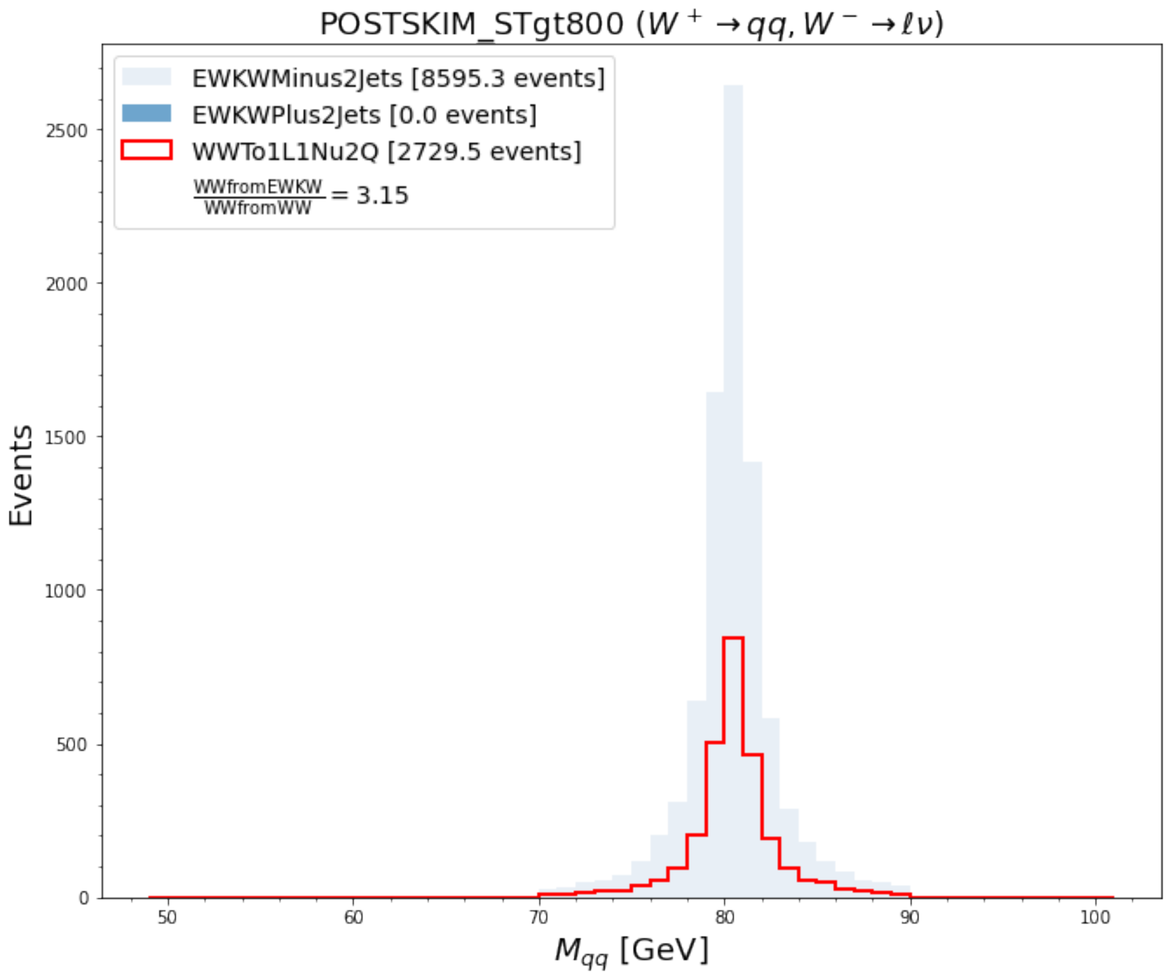
Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

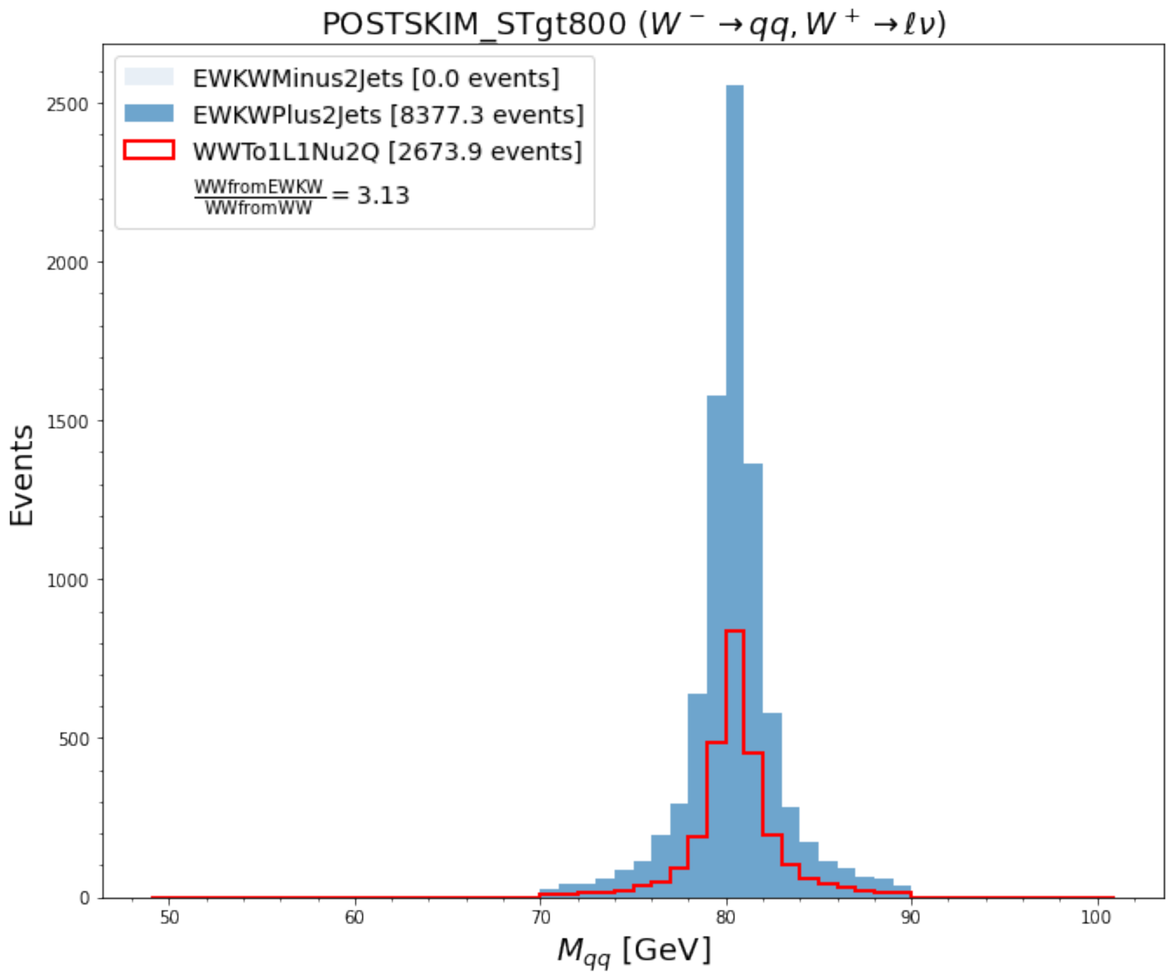




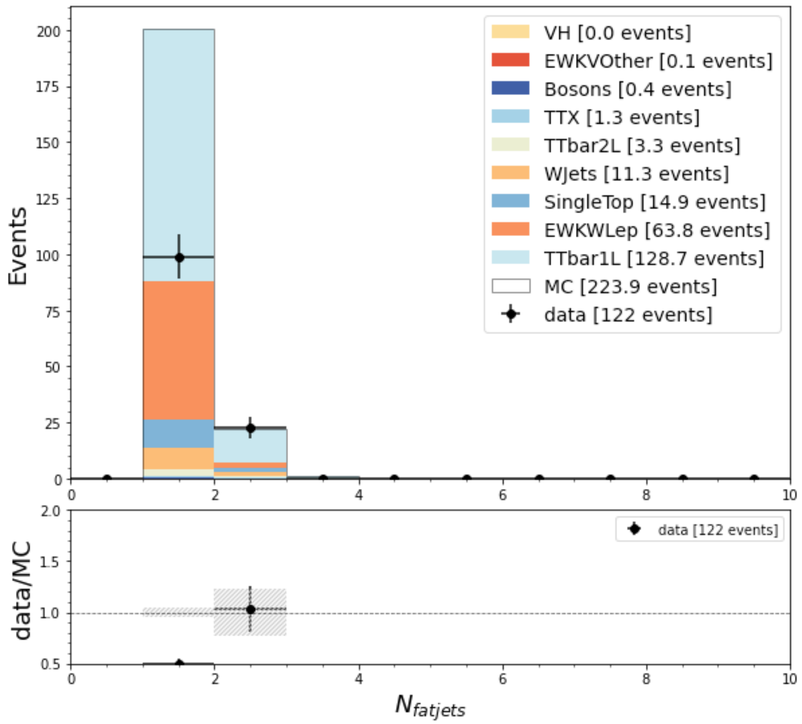
Comparing Diboson Contributions

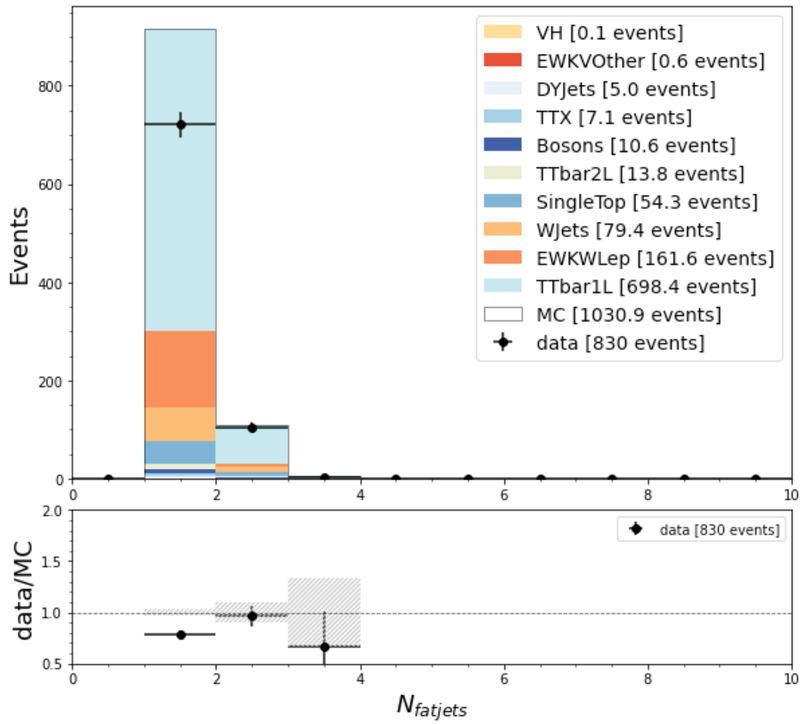
Overall Selection: $M_{\ell\nu} \in [70, 90)$ GeV AND $M_{qq} \in [70, 90)$ GeV AND $|\text{charge}_{qq}| == 1$

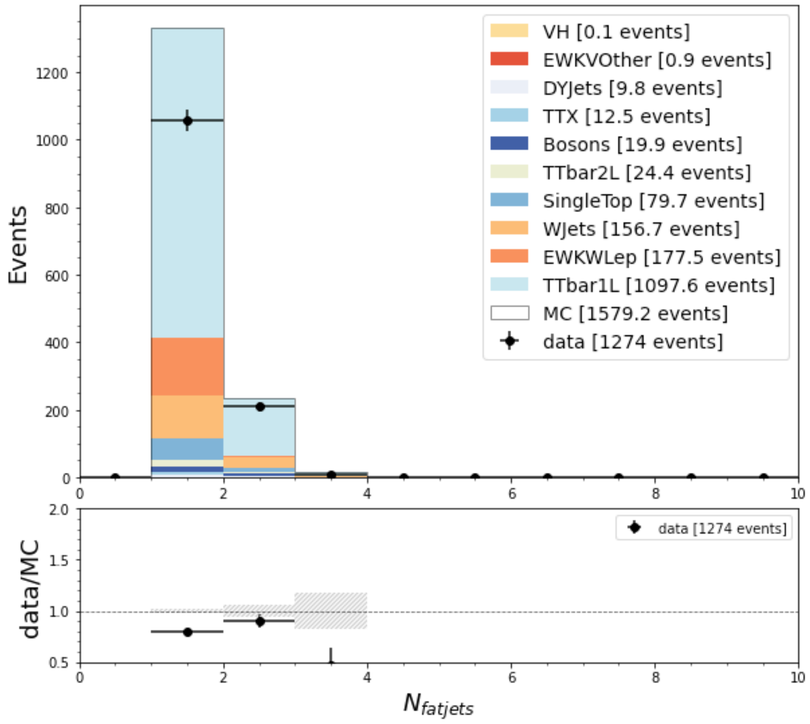




Returning to the Original Issue







- **Left:** Region A; **center:** Region A w/out S_T, Xbb, M_{jj} cuts; **right:** center w/out $\Delta\eta_{jj}$ cut
- Region A: Presel. w/out $\Delta\eta_{jj}$ cut, $S_T > 900$, Xbb > 0.9, $M_{jj} > 600$, and $|\Delta\eta_{jj}| > 4$
  - See slide 10

UC San Diego