Status update on $A \to Zh \to \ell\ell b\bar{b}$ analysis.

Paolo Checchia, Tommaso Dorigo,
Stefano Lacaprara, Mia Tosi, Alberto Zucchetta

INFN Padova

Hbb meeting,
CERN, 23 August 2013

Introduction

Introduction

- Interesting channel in 2HDM
- Including, but not limited to, SUSY, $m_h + m_Z \lesssim m_A \lesssim 2m_{top}$, decay $A \to Zh_{125} \to \ell\ell b\bar{b}$
  - also in MSSM at low $\tan\beta$, not allowed in m_h^{max} benchmark scenario given $m_h = 125\ GeV$
  - possible in other scenarios if $M_{SUSY} \gg 1\ TeV$
- Production via gluon fusion process.
  - Also $b\bar{b}$ associated production possible (not yet considered)

signature two resonant ℓ; $\ell = e, \mu$;
- two resonant $b - tag$ jets;
- reconstruct $\ell\ell b\bar{b}$ invariant mass;

background mostly Z+bb, $t\bar{t}$

What's new

- Moved to CMSSW_5_3_11
  - GlobalTags: Data `FT_53_V21_AN5` MC`START53_V7G`
  - Jet energy scale fixed was a bug in cfg for PU treatment
  - MET type1 corrected for MC
  - new b-tagger: CSV retrained and the "supercombined" NOT used, yet
- Moved to Data ReReco22Jan13
- Some new MC samples,
  - including $A \to Zh \to \ell\ell bb$ signals from full simulation
  - more statistics for $Z +$ jets background
- Consolidate control regions
  - Simultaneous scale factor fit from 3/4 CRs

2HDM Scan

- Rui Santos kindly provided the 2HDM (type-I [left] and type-II [center]) scans for $\sigma \times \mathcal{B}$ for $A \to Zh$ modes.
- https://twiki.cern.ch/twiki/bin/view/CMS/Higgs/HiggsExotics2HDM
  - $\sigma \times \mathcal{B}(A \to Zh) \times \mathcal{B}(h \to bb) \sim \mathcal{O}(1\ pb)$
  - $\times\mathcal{B}(Z \to \ell\ell) \approx 0.07$ not included in 2DHM plots!
  - $\sigma \times \mathcal{B}(A \to Zh \to \ell\ell bb) \sim \mathcal{O}(100\ fb)$

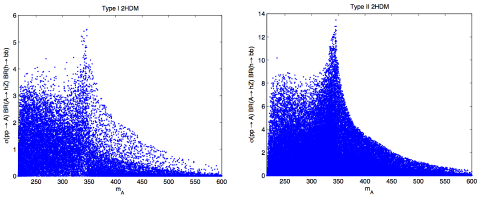


FIG. 7: The pseudoscalar scalar mass *versus* $\sigma(pp \to A)BR(A \to hZ)BR(h \to bb)$. Left - type I and right - type II. Cross sections in pbarn and masses in GeV. The peak at $\sim$ 375 GeV, corresponding to the "opening" of the $t\bar{t}$ channel in the gluon-gluon production of A

2HDM Scan

- Rui Santos kindly provided the 2HDM (type-I [left] and type-II [center]) scans for $\sigma \times \mathcal{B}$ for $A \to Zh$ modes.
- https://twiki.cern.ch/twiki/bin/view/CMS/Higgs/HiggsExotics2HDM
  - $\sigma \times \mathcal{B}(A \to Zh) \times \mathcal{B}(h \to bb) \sim \mathcal{O}(1\ pb)$
  - $\times \mathcal{B}(Z \to \ell\ell) \approx 0.07$ not included in 2DHM plots!
  - $\sigma \times \mathcal{B}(A \to Zh \to \ell\ell bb) \sim \mathcal{O}(100\ fb)$
- [right] expected sensitivity (from feasiblity studies) for this analysis
- We are in the correct ballpark (limits overlayed by hand!)
- It could be worth to look at wider m_A range: $m_A = 200 \div 500$

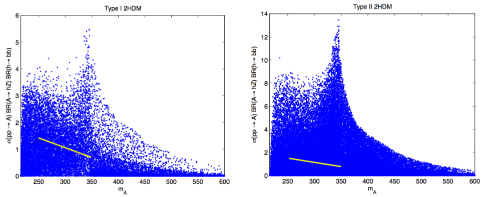
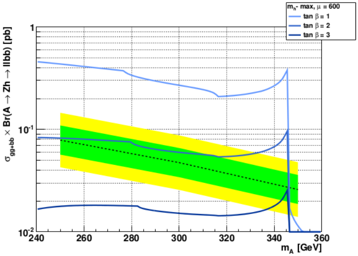


FIG. 7: The pseudoscalar scalar mass *versus* $\sigma(pp \to A)BR(A \to hZ)BR(h \to bb)$. Left - type I and right - type II. Cross sections in pbarn and masses in GeV. The peak at $\sim$ 375 GeV, corresponding to the "opening" of the $t\bar{t}$ channel in the gluon-gluon production of A



Data ReReco 22Jan

CMSSW_5_3_11

HLT paths:

- HLT_Mu17_Mu8 OR
- HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIso

Dataset	Triggers	$L[fb^{-1}]$
/DoubleMuParked/Run2012A-22Jan2013-v1	567 697	0.912
/DoubleMuParked/Run2012B-22Jan2013-v1	12 313 533	4.508
/DoubleMuParked/Run2012C-22Jan2013-v1	13 922 018	7.228
/DoubleMuParked/Run2012D-22Jan2013-v1	12 636 904	7.446
Total DoubleMuParked	39 440 152	20.094
/DoubleElectron/Run2012A-22Jan2013-v1	1 167 639	0.912
/DoubleElectron/Run2012B-22Jan2013-v1	5 905 466	4.511
/DoubleElectron/Run2012C-22Jan2013-v1	9 357 957	7.267
/DoubleElectron/Run2012D-22Jan2013-v1	6 226 511	7.446
DoubleElectron Total	22 657 573	20.136

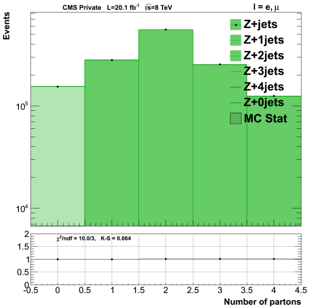
MonteCarlo samples

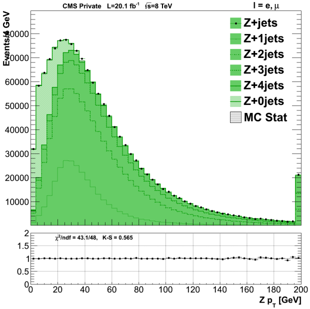
Dataset	Events	Triggers	Trigger ϵ	σ
QCD_Pt_20_MuEnrichedPt_15_TuneZ2star_8TeV_pythia6	21 484 602	338 281	1.59%	$364 \cdot 10^6$
DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball	30 459 503	8 880 856	29.16%	3,503.71
DYJetsToLL_M-10To50_TuneZ2Star_8TeV-madgraph	37 835 275	331 795	0.88%	11,050.00
DY1JetsToLL_M-50_TuneZ2Star_8TeV-madgraph	24 045 248	7 692 924	31.99%	666.30
DY2JetsToLL_M-50_TuneZ2Star_8TeV-madgraph	21 852 156	7 404 195	33.88%	214.97
DY3JetsToLL_M-50_TuneZ2Star_8TeV-madgraph	11 015 445	3 872 159	35.15%	60.69
DY4JetsToLL_M-50_TuneZ2Star_8TeV-madgraph	6 402 827	2 320 782	36.25%	27.36
DYJetsToLL_PtZ-50To70_M-50_TuneZ2Star_8TeV-madgraph	21 507 862	7 670 504	35.66%	105.70
DYJetsToLL_PtZ-70To100_M-50_TuneZ2Star_8TeV-madgraph	13 177 933	5 148 083	39.07%	62.95
DYJetsToLL_PtZ-100_M-50_TuneZ2Star_8TeV-madgraph	15 173 463	6 724 067	42.87%	39.08
DYJetsToLL_PtZ-180_M-50_TuneZ2Star_8TeV-madgraph	1 555 476	807 523	51.91%	5.42
ZbbToLL_massive_M-50_TuneZ2star_8TeV-madgraph-pythia6_tauola	14 129 304	5 916 544	41.87%	76.75
WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball	57 709 905	85 365	0.15%	37,509.00
T_s-channel_TuneZ2star_8TeV-powheg-tauola	259 961	6 221	2.39%	3.79
T_t-channel_TuneZ2star_8TeV-powheg-tauola	99 876	1 631	1.63%	56.40
T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	497 658	34 417	6.92%	11.10
Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola	139 974	3 265	2.33%	1.76
Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola	1 935 072	33 047	1.71%	30.70
Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	493 460	34 328	6.96%	11.10
TTJets_FullLeptMGDecays_8TeV-madgraph-tauola	12 011 428	4 886 110	40.68%	23.64
TTWJets_8TeV-madgraph	196 046	32 975	16.82%	0.23
TTZJets_8TeV-madgraph_v2	210 160	35 763	17.02%	0.21
WW_TuneZ2star_8TeV_pythia6_tauola	10 000 431	330 481	3.30%	33.61
WZ_TuneZ2star_8TeV_pythia6_tauola	10 000 283	475 852	4.76%	12.63
ZZ_TuneZ2star_8TeV_pythia6_tauola	9 799 908	830 053	8.47%	5.20
ZH_ZToLL_HToBB_M-125_8TeV-powheg-herwigpp	999 462	455 572	45.58%	0.02
GluGluToAToZhToLLBB_mA-250_mh-125_8TeV-pythia6-tauola	300 000	147 653	49.22%	-
GluGluToAToZhToLLBB_mA-300_mh-125_8TeV-pythia6-tauola	300 000	154 646	51.55%	-
GluGluToAToZhToLLBB_mA-350_mh-125_8TeV-pythia6-tauola	299 272	159 499	53.30%	-
Total	272 697 805			

Blue=new dataset all datasets `Summer12_DR53X-PU_S10_START53_V7A`

Z+light jets statistics in MC
usage of exclusive samples

- The events reduction for Z+light jets background is very large when b-tagging is required. Statistical problem in MC
- We can use the exclusive `DY`**N**`Jets with` **N**=`1,2,3,4` instead of the inclusive `DYJets`.
- In addition, add the `DY`**0**`Jets` from `DYJets`, selected at parton level from hard process.
  - The contribution is non negligible even after requiring 2 jets $p_T > 20, 20\ GeV$.
- Use overlapping samples by adding a weight $\frac{1}{2}$ to the events.
  - *e.g.* **Z+1 jet** from inclusive `DYJets` sample and from exclusive dataset `DY1Jets`
- We are adding also the exclusive p_T^Z-bin datasets in the same way. (not yet in this presentation)
- $Z + bb$ is reweighted $\frac{1}{2}$ from all these samples and from the exclusive dataset `ZbbToLL`

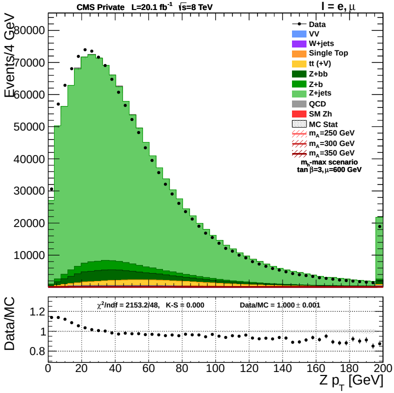




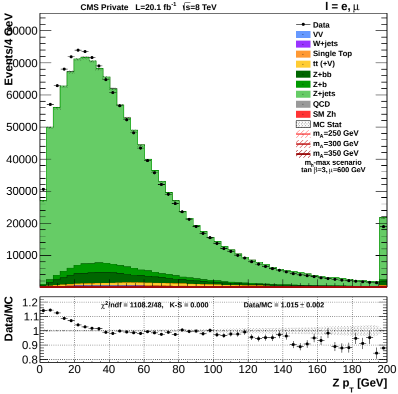
All events

Only with pre-selection cuts: discrepancy. ($bb == jj$)

With mixed exclusive dataset



With DYJets inclusive dataset only



Problem Z spectrum in DY* samples. Need re-weight as in VHbb analysis.

Analysis overview

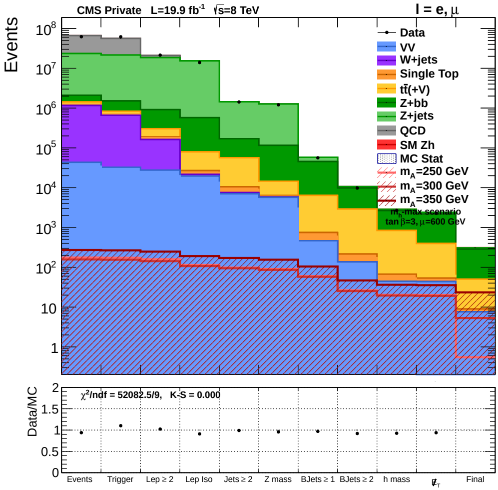
Preselection

- either HLT_Mu17_Mu8 or HLT_Ele17[...]Ele8[...] trigger fired;
- $N_\ell \geq 2$: $p_T > 20(10)\ GeV$, $\pm$, same flavour, isolated ($PF_{iso}^{rel} < 0.15$);
- $M_{\ell\ell} > 50\ GeV$
- $N_{jets} \geq 2$: $p_T > 20\ GeV$, $\Delta R_{jet,\ell} > 0.5$;

Analysis cuts

- Z selection: $80 < m_{\ell\ell} < 100\ GeV$;
- b-tagging (CSV): jet_1 is CSVT, jet_2 CSVM;
- h selection: $90 < m_{bb} < 140\ GeV$;
- top rejection: $MET < 60\ GeV$
- Final selection is m_A dependent.

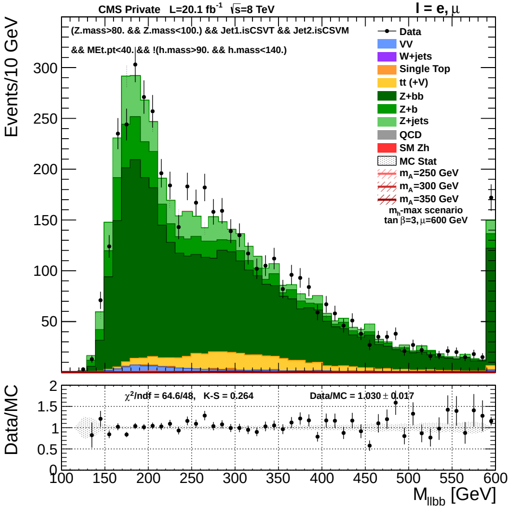
Events reductions



After preselection dominating backgrounds are:

1. $Z + bb$
2. $t\bar{t}$
3. $Z +$ light jets reducible asking b-jets
4. other: singleTop, VV

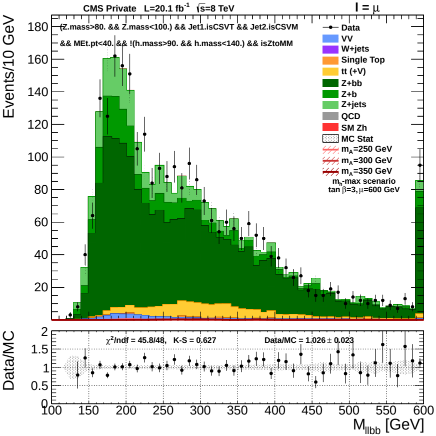
Zbb control region



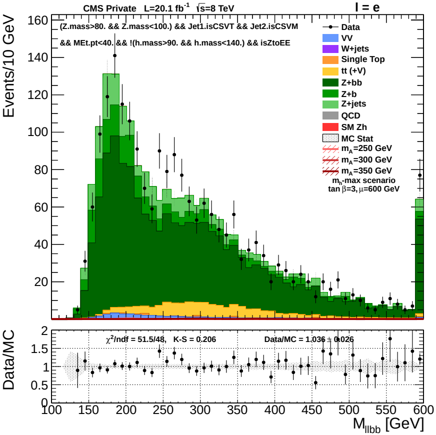
Zbb CR cuts

- Preselection
- Z selection: $80 < m_{\ell\ell} < 100\ GeV$;
- B-tag: Jet_1 CSVT, Jet_2 CSVM
- h veto $m_{bb} < 90$ $m_{bb} > 140\ GeV$
- top veto: $MET < 40\ GeV$
- Data/Bkg= 1.030±0.017

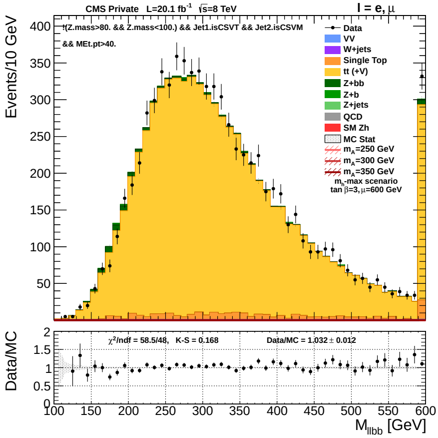
$\ell = \mu$



$\ell = \mathbf{e}$



TTbar control region

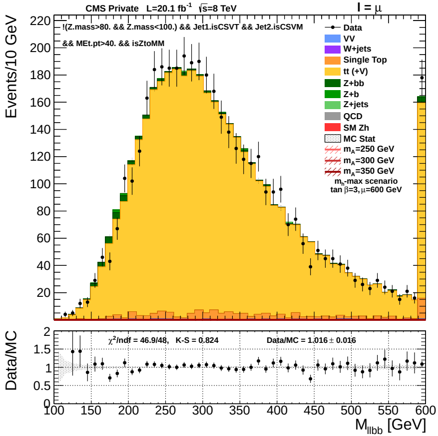


TTbar CR cuts

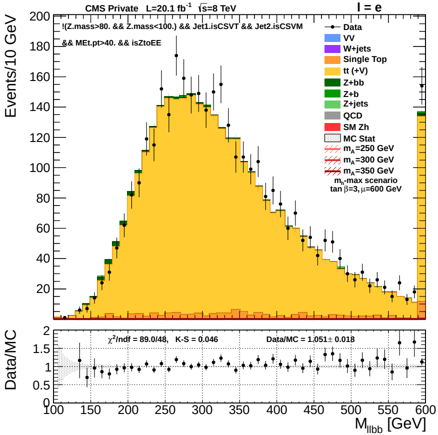
- Preselection
- Z veto $m_{\ell\ell} < 80$ OR $m_{\ell\ell} > 100$ GeV;
- B-tag: Jet_1 CSVT, Jet_2 CSVM
- top selection $MET > 40$ GeV
- Data/Bkg$= 1.032 \pm 0.012$

TTbar control region

$\ell = \mu$

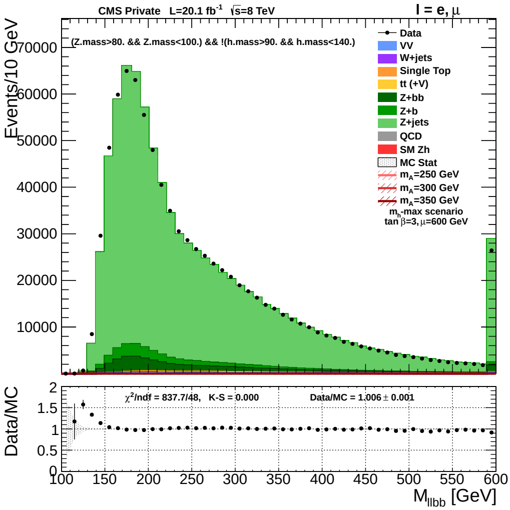


$\ell = \mathbf{e}$



ZJets control region

$m_A = m_{\ell\ell bb}$



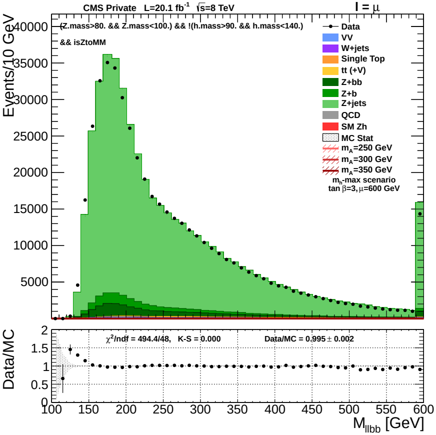
ZJets CR cuts

- Preselection
- Z selection: $80 < m_{\ell\ell} < 100\ GeV$;
- h veto: $m_h < 80$ or $m_h > 140\ GeV$
- top veto: $MET < 40\ GeV$
- Data/Bkg$= 1.006 \pm 0.002$
- Some problem for $M_{\ell\ell bb} < 150\ GeV$

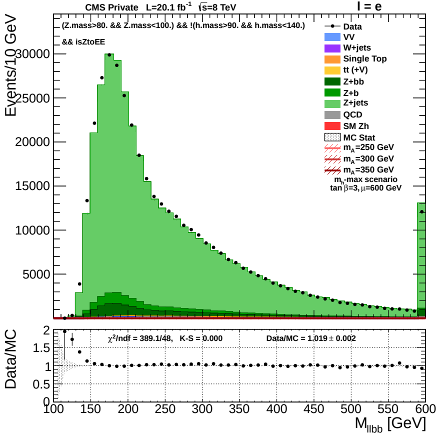
ZJets control region

$m_A = m_{\ell\ell bb}$ for μ and e separately

$\ell = \mu$

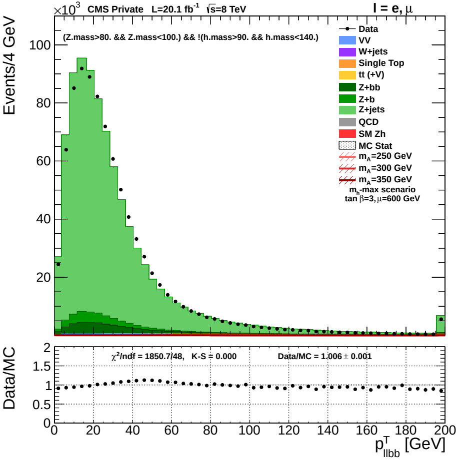


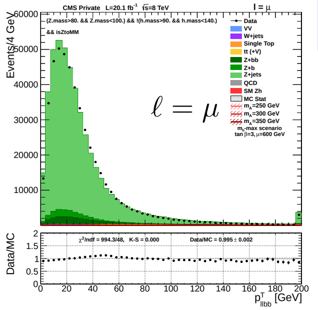
$\ell = \mathbf{e}$

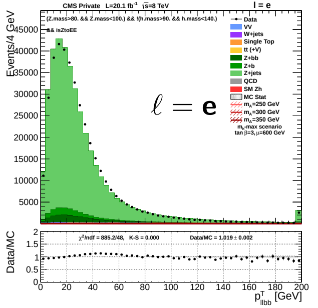


ZJets control region

$p_T^A = p_T^{\ell\ell bb}$





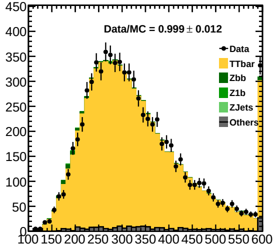


Control Region simultaneous fit

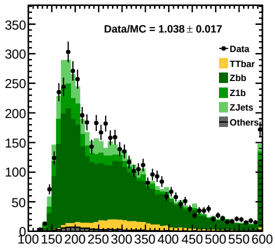
- We have 3 control regions for the 3 major background sources.
- Do a simultaneous likelihood fit of the 3 CRs to get the MC scale factor.
- The normalization of the other minor backgrounds are kept fixed.

Scale factors

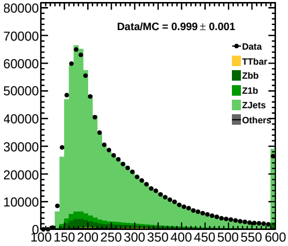
Control Region	Scale Factor simultaneous fit	Scale Factor single CR ratio
Zbb	1.035 ± 0.012	1.030 ± 0.017
TTbar	0.989 ± 0.021	1.032 ± 0.012
ZJets	1.007 ± 0.002	1.006 ± 0.002



TTbar SF: 1.035 ± 0.012



Zbb SF: 0.989 ± 0.021



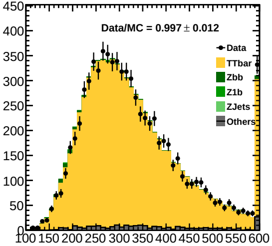
ZJets SF: 1.007 ± 0.002

Control Region simultaneous fit

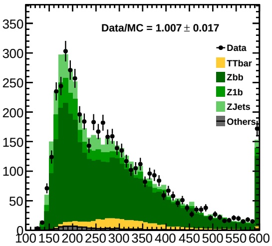
- Same but excluding $M_{\ell\ell bb} < 150\ GeV$, where ZJets has problem;
- Data/MC Agreement even better

Scale factors

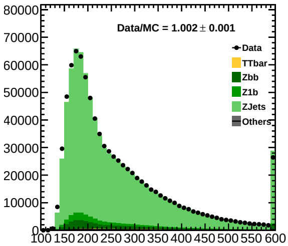
Control Region	Scale Factor simultaneous fit	Scale Factor single CR ratio
Zbb	1.030 ± 0.022	1.028 ± 0.017
TTbar	1.037 ± 0.012	1.032 ± 0.012
ZJets	0.997 ± 0.002	0.999 ± 0.001



TTbar SF: 1.037 ± 0.012



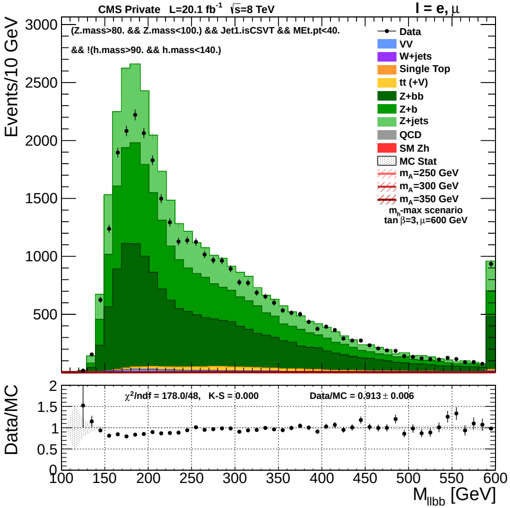
Zbb SF: 1.030 ± 0.022



ZJets SF: 0.997 ± 0.002

Zb control region

$m_A = m_{\ell\ell bb}$



Zb CR cuts

- Preselection
- Z selection: $80 < m_{\ell\ell} < 100\ GeV$;
- h veto: $m_h < 80$ or $m_h > 140\ GeV$
- B-tag: Jet_1 CSVT
- top veto: $MET < 40\ GeV$
- Data/Bkg$= 0.913 \pm 0.006$

Same as Zbb but just 1 b-tag.

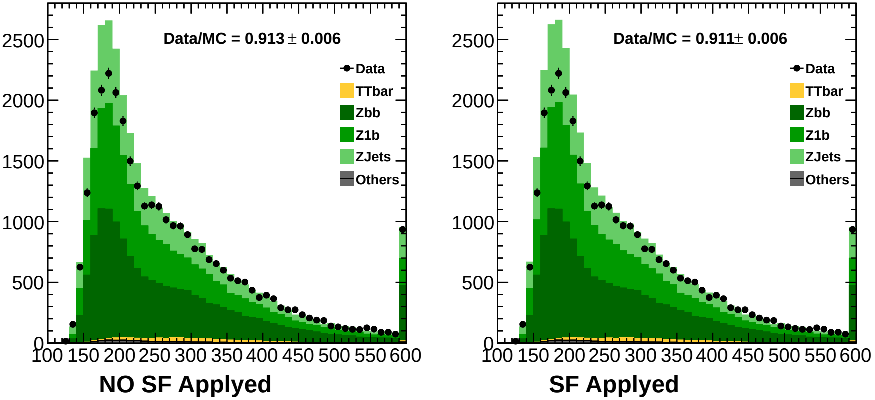
No SF applied (yet)

Statistical correlation with Zbb sample!

Will try anti b-tag on jet_2.

What about Zb Scale Factor?

Try to use the Zjets scale factor for Z1b



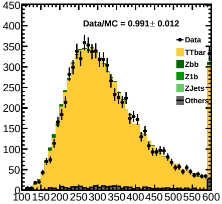
Not really great . . .

What if we use 4 CR (Zbb, TTbar, ZJets and Z1b)

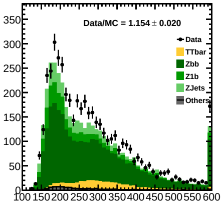
- Use 4 control regions for the 4 major background sources.
- Zbb CR is significantly worse
- Similar results excluding $M_{\ell\ell bb} < 150\ GeV$ (in backup)
- Correlation between Zbb and Zb plays a significant role!

Scale factors

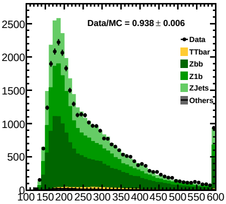
Control Region	Scale Factor simultaneous fit	Scale Factor single CR ratio
Zbb	0.842 ± 0.011	1.030 ± 0.017
TTbar	1.044 ± 0.021	1.052 ± 0.012
ZJets	1.014 ± 0.001	1.032 ± 0.002
Z1b	0.913 ± 0.006	0.913 ± 0.002



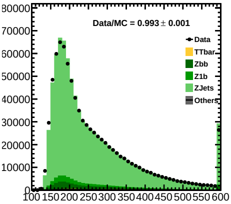
TTbar SF: 1.044 ± 0.012



Zbb SF: 0.842 ± 0.011



Z1b SF: 0.913 ± 0.006



ZJets SF: 1.014 ± 0.001

Use scale factors from 3 (or 4 CR) fit to get systematics uncertainties on background normalization

Summary and TODO

Summary

- Data and MC processing up-to-date
  - Need to re-weight p_T^Z spectrum
  - `DYJets` statistics improved with exclusive samples.
  - Add also exclusive DY with p_T^Z bin
  - More mass points for m_A
  - Is PYTHIA enough for signal kinematics?
- Control region consolidated
  - Work on correlation between CR
    - ★ try anti b-tag on Zjets and/or Z1b
  - Still some problem with Z+1b
- Finalize cut and counts analysis
  - already in good shape, no time to show it today: maybe next meeting.
- Start with MVA approach

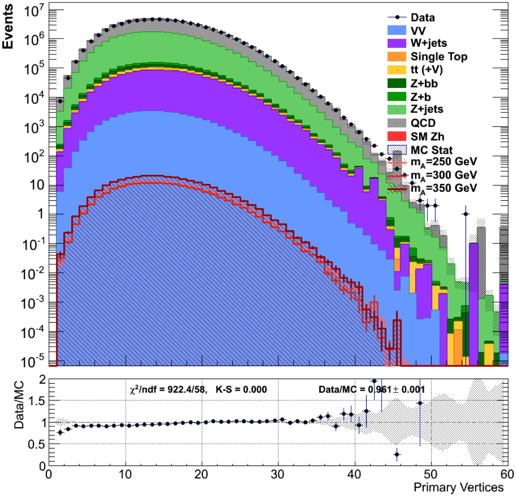
Backup

Guess what?

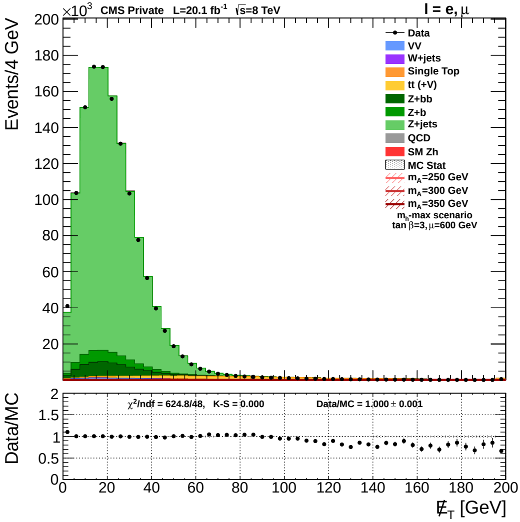
Yes!

Backup slides.

PileUp reweight

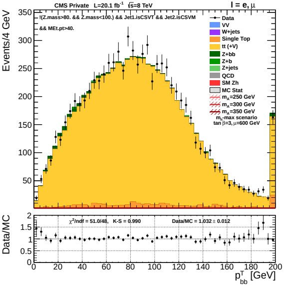


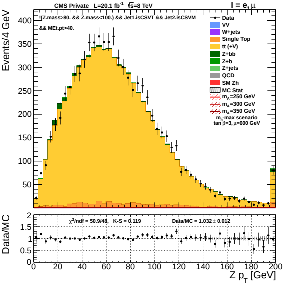
MET



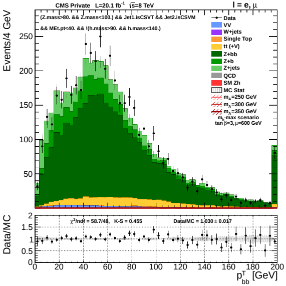
×10^3 CMS Private L=20.1 fb^-1 √s=8 TeV
l = e, μ
Events/4 GeV
Data
VV
W+jets
Single Top
tt (+V)
Z+bb
Z+b
Z+jets
QCD
SM Zh
MC Stat
mA=250 GeV
mA=300 GeV
mA=350 GeV
mh-max scenario
tan β=3, μ=600 GeV
Data/MC
χ2/ndf = 624.8/48, K-S = 0.000
Data/MC = 1.000 ± 0.001
E̸T [GeV]

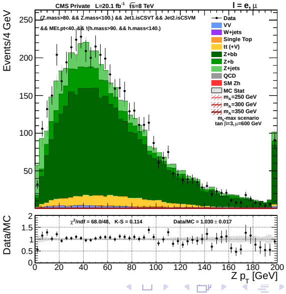
TTbar CR





Zbb CR



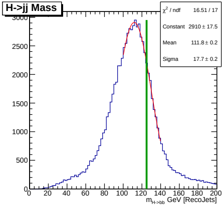


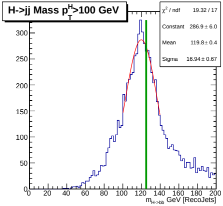
Jets energy scale issues (solved)

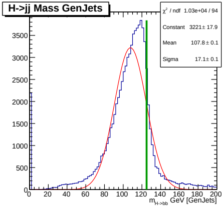
- With the last PAT productions (based on 52X) we had a problem in the Jet energy scale.
  - $< m_{125}^{h\to bb} >= 103\ GeV$, $< m_{Z\to jj} >= 78\ GeV$
- Problem tracked down to incorrect PU treatment
  - many thanks to Michele for his suggestion!
- Furthermore, there is a very clear correlation between the reconstructed invariant mass and the pt of the H/W/Z the jets come from
- given our cuts, the heavy object is less boosted than the analogous in the standard VH analysis, so care must be taken in comparing the results.
- Following slides: jets used comes from H/W/Z (MC truth)

$H_{125} \to bb$ mass distribution

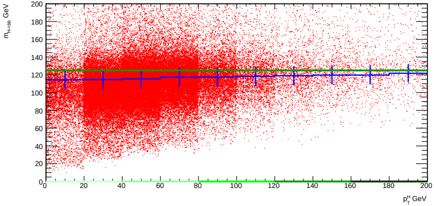
Invariant mass bb: all events, $p_T^h > 100\ GeV$, using `GenJets`

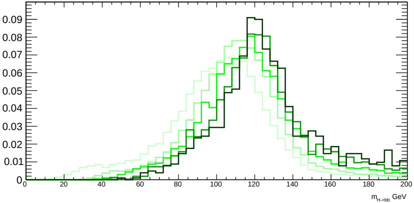






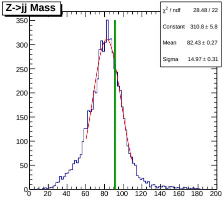
Invariant mass bb vs p_T^h: error bar are σ of Gaussian fit

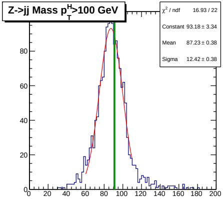


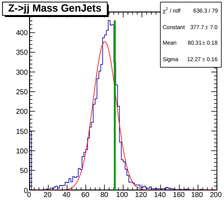


$Z \to jj$ mass distribution

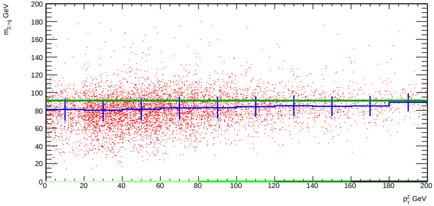
Invariant mass jj: all events, $p_T^Z > 100$ GeV, using `GenJets`

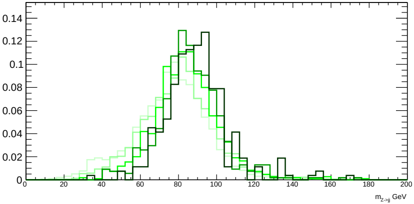






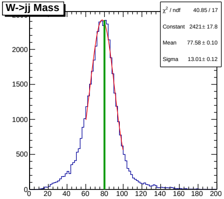
Invariant mass bb vs p_T^Z: error bar are σ of Gaussian fit

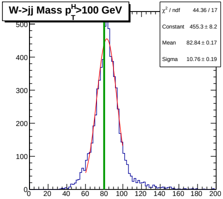


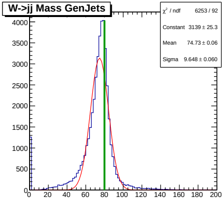


$W \to jj$ mass distribution

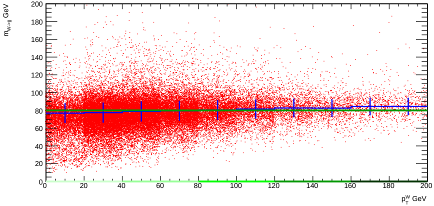
Invariant mass jj: all events, $p_T^W > 100$ GeV, using `GenJets`

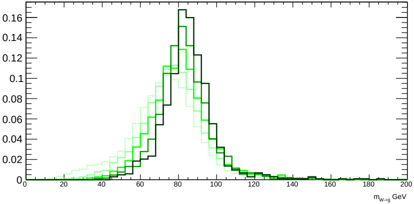






Invariant mass bb vs p_T^W: error bar are σ of Gaussian fit



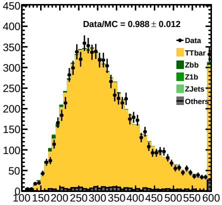


4 CR (Zbb, TTbar, ZJets and Z1b) and $M_{\ell\ell bb} > 150\ GeV$

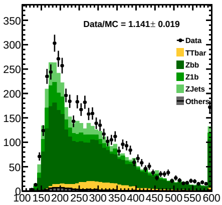
- likewise excluding $M_{\ell\ell bb} < 150\ GeV$
- Zbb CR is significantly worse

Scale factors

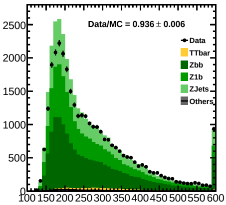
Control Region	Scale Factor simultaneous fit	Scale Factor single CR ratio
Zbb	0.855 ± 0.019	1.028 ± 0.017
TTbar	1.047 ± 0.012	1.032 ± 0.012
ZJets	0.994 ± 0.001	0.999 ± 0.001
Z1b	0.913 ± 0.006	0.911 ± 0.006



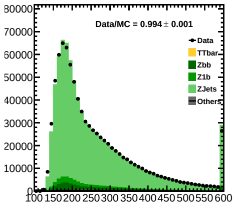
TTbar SF: 1.047 ± 0.012



Zbb SF: 0.855 ± 0.011



Z1b SF: 0.913 ± 0.006



ZJets SF: 1.006 ± 0.001