
# Search for heavy di-muon resonances + b-jets

Chris McNicol

Warwick EPP Seminar – 7th March 2019

# Contents

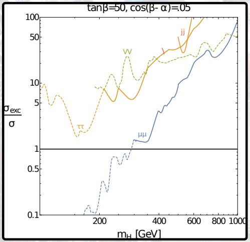

I will be describing the analysis I have been working on: arXiv:1901.08144

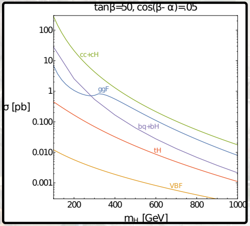
- Motivation
- Object & Event Selection
- Background & Signal Modeling
- Fit procedure
- Results
- New! Improved background rejection with BDT

# Motivation


- Many BSM models result in high-mass scalars
- E.g Flavourful Higgs Models
  - Different flavour-dependent yukawa structures
- Model-independent search, results applicable to many models

$$\lambda^{l} \sim \frac{\sqrt{2}}{v}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & m_\tau \end{pmatrix} \quad \lambda'^{l} \sim \frac{\sqrt{2}}{v'}\begin{pmatrix} m_e & m_e & m_e \\ m_e & m_\mu & m_\mu \\ m_e & m_\mu & m_\mu \end{pmatrix}$$



# Object & Event Selection


Muons:

- 2 muons with opposite charge
- mu20_iloose (2015) / mu26_ivarmedium (2016) OR mu50 Trigger
- Muon $p_T > 30\ GeV$

Jets:

- $p_T > 25\ GeV$
- b-tagged with MV2c10 85% working point

**Signal Regions**

High mass ($m_{\mu\mu} > 160\ GeV$):

**SRbVeto** $= 0$ b-tagged jets

**SRbTag** $\geq 1$ b-tagged jets

**Control Regions**

Low mass ($m_{\mu\mu} < 160\ GeV$):

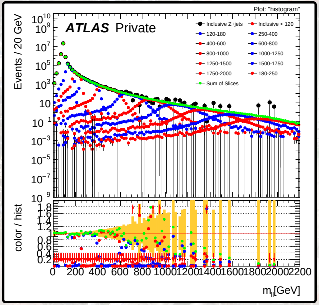
**CRbVeto** $= 0$ b-tagged jets

**CRbTag** b-tag AND $E^T_{miss} < 100\ GeV$

**CRttbar** b-tag AND $E^T_{miss} > 100\ GeV$

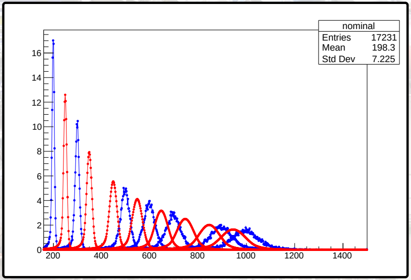
# Background MC

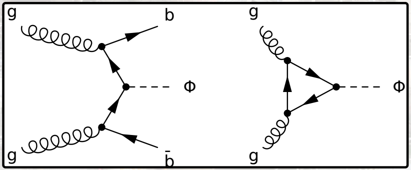
- Z+jets - Powheg+Pythia8 with AZNLO tune - sliced in $M_{\mu\mu}$ - applying mass-dependent k-Factor to match NNLO, QCD & EW
- Splitting Zjets into light-flavour (LF) and heavy-flavour (HF) components using HadronConeExclTruthLabelID.
- $t\bar{t}$ - Powheg+Pythia6 with Perugia tune.
- Other backgrounds: Diboson (Sherpa using mass-sliced samples) & Single-Top (Powheg+Pythia)



# Signal MC

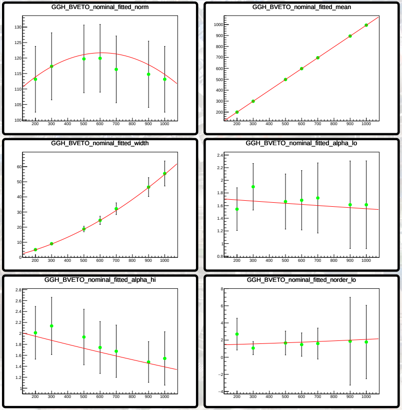
- Use heavy Higgs model as proxy for any narrow scalar resonance
- Looking at b-associated ($bb\Phi$) and gluon-fusion ($ggF$) production mechanisms
- $m_\Phi$ in the range $200 - 1000\ GeV$ in steps of $100\ GeV$
- Generators: $bb\Phi$ (aMc@NLO), ggF (Powheg)



# Signal Shape Interpolation

- Double-sided Crystal Ball models signal well across the mass range
- Parameterize each DCSB parameter with a polynomial
- Verify by removing templates and compare with interpolated shape




| DSCB Parameter | Polynomial |
|---|---|
| Normalisation | $a_N + b_N m_\Phi + c_N m_\Phi^2$ |
| Mean ($\bar{x}$) | $a_{\bar{x}} + b_{\bar{x}} m_\Phi$ |
| Width ($\sigma$) | $a_\sigma + b_\sigma m_\Phi + c_\sigma m_\Phi^2$ |
| $\alpha_L$ | $a_{\alpha_L} + b_{\alpha_L} m_\Phi$ |
| $n_L$ | $a_{n_L} + b_{n_L} m_\Phi$ |
| $\alpha_H$ | $a_{\alpha_H} + b_{\alpha_H} m_\Phi$ |
| $n_H$ | $a_{n_H}$ |

$$DSCB(M_{\mu\mu};\bar{x},\sigma,\alpha_L,\alpha_H,n_L,n_H) = \begin{cases} e^{-\left(\frac{(M_{\mu\mu}-\bar{x})^2}{2\sigma^2}\right)} & \text{for } \alpha_L < \frac{M_{\mu\mu}-\bar{x}}{\sigma} < \alpha_H \\ = \left(\frac{n_L}{|\alpha_L|}\right)^{n_L} \times \left(\frac{n_L}{|\alpha_L|} - |\alpha_L| - \frac{M_{\mu\mu}-\bar{x}}{\sigma}\right)^{-n_L} & \text{for } \alpha_L \geqslant \frac{M_{\mu\mu}-\bar{x}}{\sigma} \\ = \left(\frac{n_H}{|\alpha_H|}\right)^{n_H} \times \left(\frac{n_H}{|\alpha_H|} - |\alpha_H| - \frac{M_{\mu\mu}-\bar{x}}{\sigma}\right)^{-n_H} & \text{for } \alpha_H \leqslant \frac{M_{\mu\mu}-\bar{x}}{\sigma} \end{cases}$$

# Fit details

- Parameter of Interest: Signal strength
- 3 background normalisation factors - freely floating
- A lot of Nuisance parameters covering uncertainties
- Binned likelihood fit, test statistics:

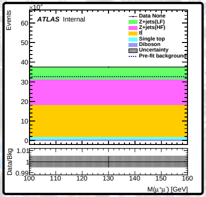
$$q_0 = \begin{cases} -2\log \frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(0,\hat{\hat{\theta}}(0))} & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

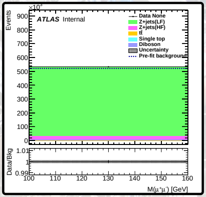
$$\tilde{q}_\mu = \begin{cases} -2\log \frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \\ -2\log \frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(\mu,\hat{\theta})} & 0 \leq \hat{\mu} \leq \mu \\ 0 & \hat{\mu} < \mu \end{cases}$$

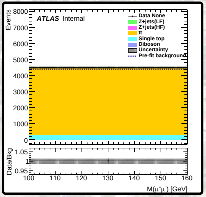
- Experimental uncertainties
  - Muons
  - Jets
  - Flavour tagging
  - $E_{miss}^T$
- Theoretical uncertainties
  - Special $m_{\mu\mu}$ shape uncertainties for Zjets+HF and $t\bar{t}$
  - Extrapolation uncertainties CR→SR
  - Parton shower etc.

# Results - Control Regions

- Post-fit Control Regions: each 1 bin
- $t\bar{t}$ and Zjets+LF normalisations adjusted by 2%
- Zjets+HF scaled by 46% to account for b multiplicity modeling deficiencies



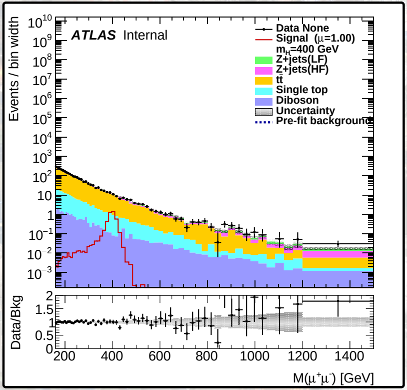



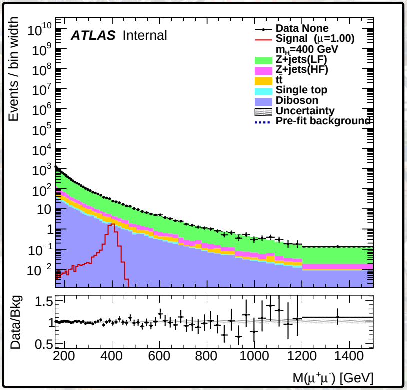




| Normalisation Factor | Asimov data | Real data |
|---|---|---|
| mu_Zjets_HF | $1.00 \pm 0.23$ | $1.46 \pm 0.26$ |
| mu_Zjets_LF | $1.00 \pm 0.02$ | $1.02 \pm 0.02$ |
| mu_ttbar | $1.00 \pm 0.04$ | $1.02 \pm 0.04$ |

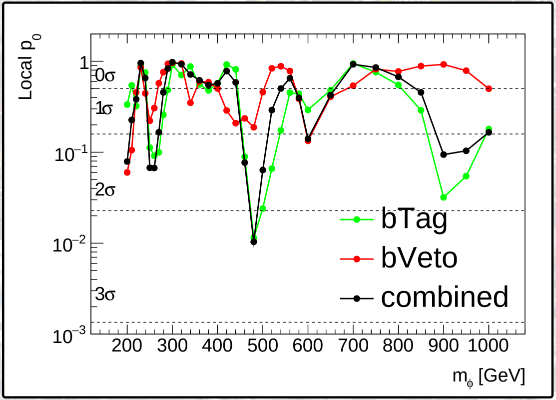
# Results - Signal Regions

Let's go bump hunting!



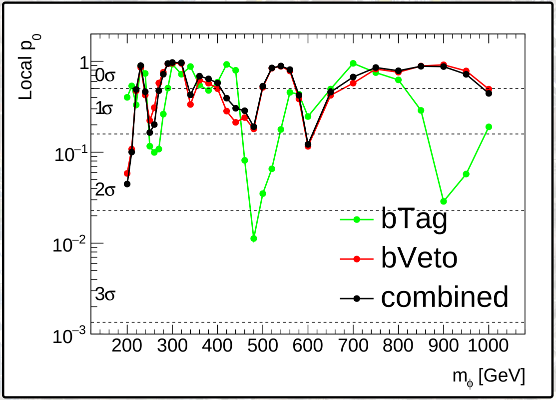



# Results - $p_0$ $bb\Phi$




- Most significant excess at approximately $m_\Phi = 480\ GeV$
- Driven by excess in b-Tag region
- Local significance $2.37\sigma$
- After Look-Elsewhere effect, global significance $0.68\sigma$

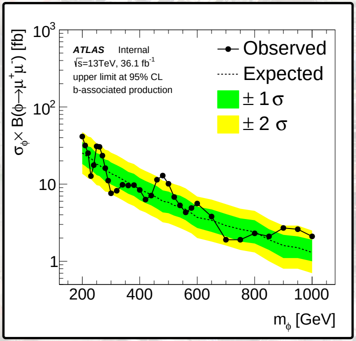
# Results - $p_0$ *ggF*



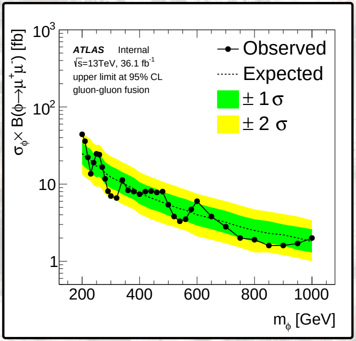

- No similarly significant excesses for *ggF*

# Results - upper limits

Observed Cross-section limits consistent with expectations in the range $30fb - 2fb$

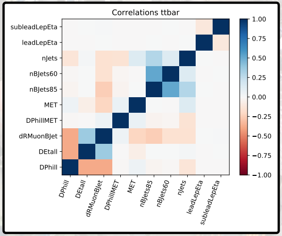


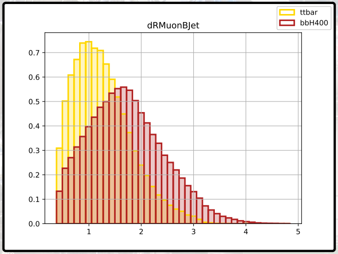

$bb\Phi$

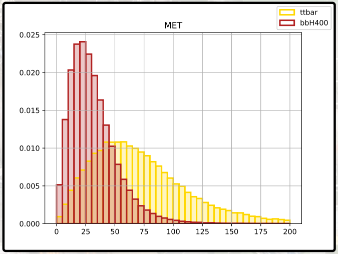



*ggF*


- Current Selection quite loose and “model independent”
- Can we gain sensitivity if we have a better selection?
- Several good, correlated discriminating variables
- Ideal for an MVA!



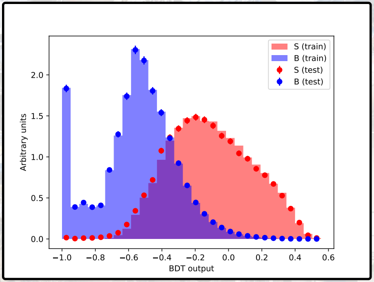



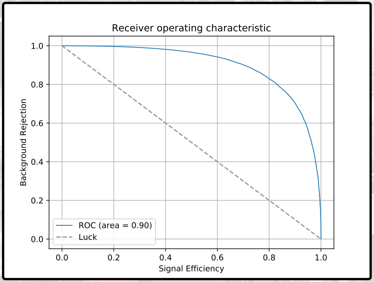



# BDT details


- Train a BDT with $t\bar{t}$ as "signal"
- Approximately 1M events for each sample
- Training on 30%
- Using `AdaBoost` classifier in `scikit-learn`
- Good discrimination
- Many possible optimisations, not a lot of time!

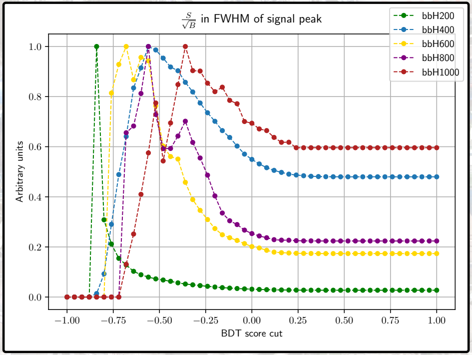
| HyperParameter | Value |
|---|---|
| Max_depth | 7 |
| Min_samples_leaf | 100 |
| n_estimators | 1500 |
| learning_rate | 0.05 |



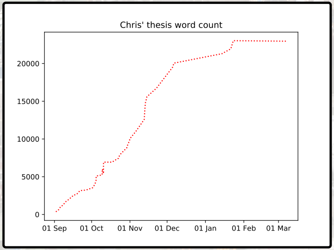
# Sensitivity Gains

- Scan through cuts on BDT score
- Check $\frac{S}{\sqrt{B}}$ in the signal peak
- Best gains in sensitivity at low mass but with a tighter cut
- Perhaps indicates choosing different cuts for different masses or even training separate BDTs for each mass - although there are lots of complications with this



# Conclusion

- Presented search for new physics in $\mu^+\mu^- + b$ final state
- No significant excesses above Standard Model expectations
- Cross-section times Branching Ratio limits set
- Use of BDT selection to improve sensitivity looks promising
- Thesis mostly ~~plagiarised~~ written



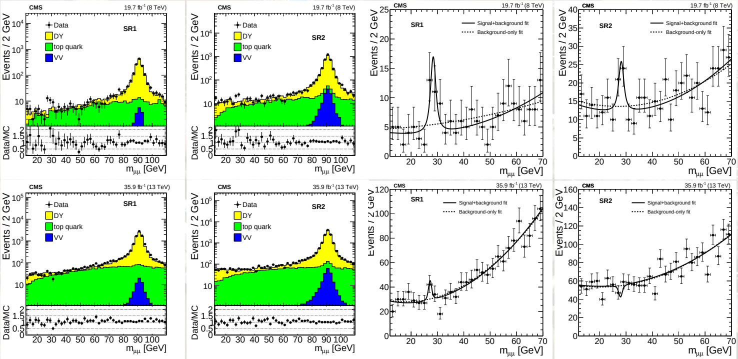

# Backup

Backup

How my family feel when I talk about CERN

# CMS 28GeV ‘bump’

- Recent CMS paper on arXiv looks for dimuon resonances at low mass, with associated b-jets, in 8 *TeV* and 13 *TeV* data.
- Excesses seen at $m_{\mu\mu} \approx 28\ GeV$, particularly in 8 *TeV* data.
- Cross-check with ATLAS 13 *TeV* data, since bHmumu ATLAS analysis has similar signature.



# POI uncertainty breaking

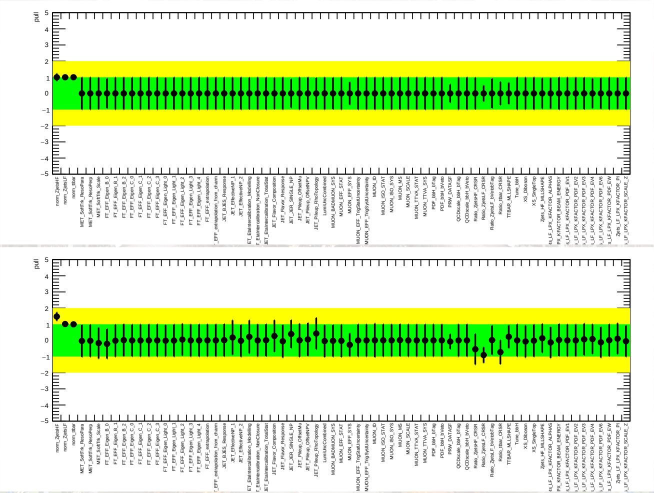
POI uncertainty breaking for different systematic sources from the combined fit with the 400 GeV bbH signal. Dominated by data statistics, then MC statistics.

| POI | Central Value |
|---|---|
| SigXsecOverSM | -0.0272564 |

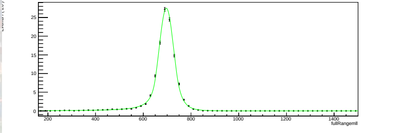
| Set of nuisance parameters | Impact on error |
|---|---|
| Total | $^{+0.417}_{-0.41}$, $\pm 0.414$ |
| DataStat | $^{+0.365}_{-0.357}$, $\pm 0.361$ |
| FullSyst | $^{+0.202}_{-0.2}$, $\pm 0.201$ |
| All normalizations | $^{+0.0517}_{-0.0392}$, $\pm 0.0455$ |
| Mll Shape | $^{+0.0712}_{-0.0768}$, $\pm 0.074$ |
| Muon | $^{+0.0628}_{-0.0427}$, $\pm 0.0528$ |
| Btag | $^{+0.0198}_{-0.0243}$, $\pm 0.0221$ |
| Jet/MET | $^{+0.00808}_{-0.00915}$, $\pm 0.00861$ |
| Ratio | $^{+0.031}_{-0.0267}$, $\pm 0.0289$ |
| Modeling | $^{+0.00971}_{-0.0196}$, $\pm 0.0146$ |
| Pileup reweighting | $^{+0.01}_{-0.0131}$, $\pm 0.0116$ |
| Signal theory | $^{+0.00489}_{-0.0064}$, $\pm 0.00565$ |
| Luminosity | $^{+0.0123}_{-0.0109}$, $\pm 0.0116$ |
| MC stat | $^{+0.139}_{-0.143}$, $\pm 0.141$ |

# Pulls

- top: asimov fit, bottom: data in control regions

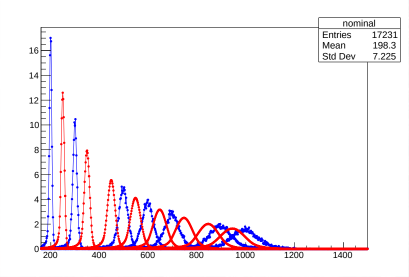


# Signal Interpolation - Quality of fit



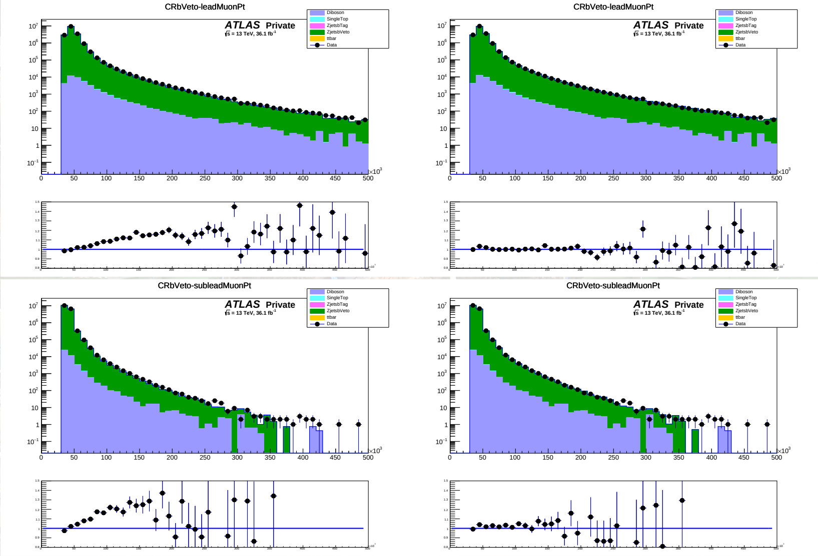

- The interpolated pdf (green ), calculated when each sample is removed from the fit, remains a good fit.

MC - Interpolated




- The shape and normalisation match the MC templates across the mass range.

# Ptll reweight - bVeto leadMuon pT

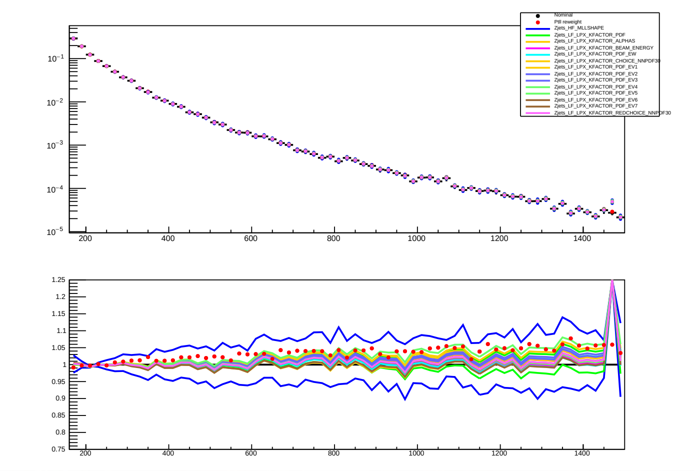



Before re-weight

After re-weight

# Ptll reweight - Signal Regions

Effect of Ptll reweight much smaller than Zjets modeling uncertainties currently applied. For b-Tag signal region (for b-Veto region, reweighting has smaller effect):



# Experimental Systematic Uncertainties

Include recomended systematics for following objectes are varied coherently between signal and backgrounds:

- muon (trigger, reco, momentum scale and resolution, muon TTVA, isolation and ID - (BadMuonSys for use with High pT working point)
- jet (use 21NP scheme)
- flavour tagging (use reduced scheme with 13 NPs)
- MET (SoftTrack Reso systs, and Scale up and Down)
- pile-up rescaling and pile-up rejection systematics

# Modelling and Theoretical Systematic Uncertainties

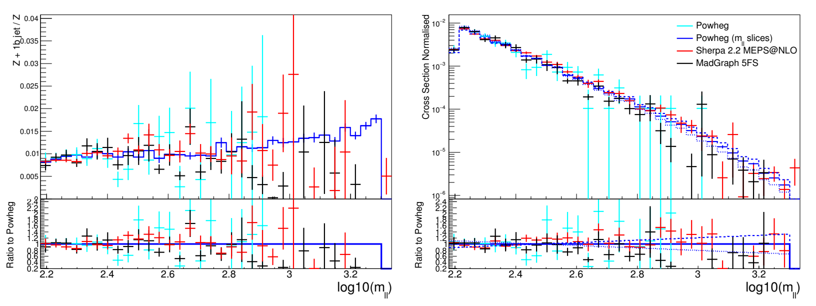
- Theory uncertainties due to PDFs, $\alpha_S$, QCD scale and EW corrections for Z+LF (both norm. and shape).
- Shape uncertainty on the $m_{ll}$ for Z+H.F.
- Shape uncertainty on the $m_{ll}$ for $t\bar{t}$ process, following recomendations from Top group
- Control Region to Signal Region extrapolation uncertainties

The theory uncertainties for Z+jets calculated by the $Z^{prime}$ analysis applied as shape-only to both the Z+LF and Z+HF components.

# Systematic uncertainty for Z+HF component

Z+jets modelling comparison at truth level of $m_{ll}$ dsitrubution in the range $m_{ll} > 160$ GeV between POWHEG, MEPSatNLO and MADGRAPH

- LEFT: the fraction of $Z+ \geq 1$ b-jet events
- RIGHT: log10($m_{ll}$) for $Z+ \geq 1$ b-jet events

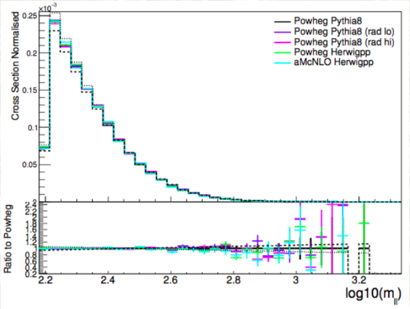



A shape systematic uncertainty as a function of truth $m_{ll}$ is defined to cover the differences: $\pm 0.4 \times \log_{10} \frac{m_{ll}}{300}$.

# Systematic uncertainty for *ttbar* background

Following reccomendations from Top group, a shape systematic uncertainty is estimated comparing $m_{ll}$ shape for different background samples.

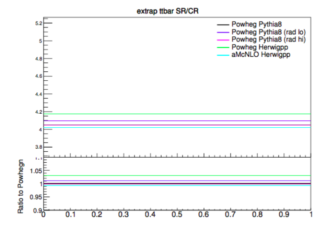
- Powheg+Pythia6 (DSID 410000)
- Powheg+Pythia6-radHi (DSID 410001)
- Powheg+Pythia6-radLo (DSID 410002)
- aMCatNLOHerwiggpp - ME/shower (DSID 410003)
- PowegHerwiggpp - shower (DSID 410004)




A functional form is chosen to encompass the differences: $\pm 0.2 \times \log_{10} \frac{m_{ll}}{300}$. Applied using reco-level $m_{ll}$.

# Extrapolation Uncertainty

Uncertainty on the relative normalisation for the $t\bar{t}$, Z+HF and Z+LF processes in the control and signal regions. Estimated for $t\bar{t}$ as acceptance ratio between the signal and the $t\bar{t}$ control region for different MC samples.




- Ratio_TTBAR_CRSR: 3.5 %

We use three extrapolation uncertainties for the Z+jets backgrounds.

- Ratio_ZjetsLF_bVetobTag - uncertainty on extrapolation of ZjetsLF from bVeto CR to bTag SR Value 27%. Applied in CRbTag and SRbTag
- Ratio_ZjetsLF_CRSR - uncertainty on extrapolation of ZjetsLF from bVeto CR to bVeto SR Value 2%. Applied in all SRs.
- Ratio_ZjetsHF_CRSR - uncertainty on extrapolation of ZjetsHF from bTag CR to bTag SR Value 5%. Applied in all SRs.