Opportunities and Challenges of Standard Model Production Cross Section Measurements at 8 TeV using CMS Open Data

Aram Apyan, Markus Klute, Matthias Schott
- Typical cross section measurements select a signal process such that it maximizes statistics and minimizes backgrounds
  - E.g. Z boson selection: 2 muons with $p_T > 20 \text{ GeV}$, $|\eta| < 2.5$, $60 < m_\ell \ell < 120 \text{ GeV}$

- Measurements are important to test perturbative QCD predictions at (N)NNLO
  - Experimental precisions at 1% level, however, inclusive measurements dominated by luminosity uncertainties
What is CMS Open Data?
- CMS published the full 2011 and 2012 data-sets under
  - Webpage: http://opendata.cern.ch
  - Newer data will be added successively
- Including Data and SM MC Samples
  - No BSM Samples available

The full analysis software (for reconstruction and simulation) is available
- Basic information on how to run the software and read the data/MC
- Use public analysis framework (Bacon), which is available under
  - https://github.com/ksung25/BaconProduct/tree/Run1
Repeating SM Cross Section Measurements

- Measurement of inclusive Cross-Sections straight forward

\[ \sigma_{V}^{\text{incl}} = \frac{N_{\text{signal}}}{\epsilon \cdot BR \cdot \int Ldt} \]

- Need to know
  - the number of selected signal candidates (i.e. need data)
  - the number of background events (i.e. need MC samples)
  - the detector response (i.e. MC Truth and Reco-level information of signal sample)
  - Integrated luminosity of the data used for the analysis

<table>
<thead>
<tr>
<th>Process</th>
<th>Definition of fid. phase-space</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Z/\gamma^* \to e^+e^-)</td>
<td>((1e^+1e^-), 60 &lt; m_{ee} &lt; 120 \text{ GeV}, p_T^e &gt; 25 \text{ GeV},</td>
</tr>
<tr>
<td>(Z/\gamma^* \to \mu^+\mu^-)</td>
<td>((1\mu^+1\mu^-), 60 &lt; m_{\mu\mu} &lt; 120 \text{ GeV}, p_T^\mu &gt; 25 \text{ GeV},</td>
</tr>
<tr>
<td>(Z/\gamma^* \to \mu^+\mu^-) + (\geq 1\text{ jet})</td>
<td>((1\mu^+1\mu^-), 70 \leq m_{\mu\mu} &lt; 110 \text{ GeV}, p_T^\mu &gt; 20 \text{ GeV},</td>
</tr>
<tr>
<td>(W^+ \to \mu^+\nu)</td>
<td>((1\mu^+), p_T^\mu &gt; 25 \text{ GeV},</td>
</tr>
<tr>
<td>(W^- \to \mu^-\nu)</td>
<td>((1\mu^-), p_T^\mu &gt; 25 \text{ GeV},</td>
</tr>
<tr>
<td>(t\bar{t} \to \mu^+e^+\nu\bar{\nu}b\bar{b})</td>
<td>(1\mu^\pm, 1e^\mp, p_T^{\text{vis}} &gt; 20 \text{ GeV},</td>
</tr>
<tr>
<td>(W^\pm Z \to l^\pm l^\mp l^+l^-)</td>
<td>((e^\pm e^\mp \mu^\pm), (\mu^\pm \mu^\mp e^\pm), (\mu^\pm \mu^\mp \mu^\pm), p_T &gt; 25 \text{ GeV},</td>
</tr>
</tbody>
</table>

- Decided to re-measure
  - \(Z\to\ell\ell\) (to test lepton performance)
  - \(Z\to\mu\mu+\text{jets}\) (to test jet performance)
  - \(W\to\mu\nu\) (to test \(E_T^{\text{Miss}}\) and trigger)
  - Top-pair production (to test b-jets)
  - WZ (to test leptons)
### Data and MC Samples

- In total 1.7-1.8 fb\(^{-1}\) of data analyzed
- Mainly used muon-triggered data-stream
- Sufficient fraction of MC skimmed, so that MC stat. uncertainties are not dominant

<table>
<thead>
<tr>
<th>Process</th>
<th>Dataset Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>(pp \rightarrow Z/\gamma^* + X \rightarrow e^+e^- + X)</td>
<td>DYToEE_M-20_CT10_TuneZ2star_v2_8TeV [35]</td>
</tr>
<tr>
<td>(pp \rightarrow Z/\gamma^* + X \rightarrow \mu^+\mu^- + X)</td>
<td>DYToMuMu_M-20_CT10_TuneZ2star_v2_8TeV [36]</td>
</tr>
<tr>
<td>(pp \rightarrow Z/\gamma^* + X \rightarrow l^+l^- + X)</td>
<td>DYJetsToLL_M-50_TuneZ2Star_8TeV [37]</td>
</tr>
<tr>
<td>(pp \rightarrow W^+ + X \rightarrow \mu^+\nu + X)</td>
<td>WplusToMuNu_CT10_8TeV [38]</td>
</tr>
<tr>
<td>(pp \rightarrow W^- + X \rightarrow \mu^-\nu + X)</td>
<td>WminusToMuNu_CT10_8TeV [39]</td>
</tr>
<tr>
<td>(pp \rightarrow W^+ + X \rightarrow \tau^+\nu + X)</td>
<td>WplusToTauNu_CT10_8TeV [40]</td>
</tr>
<tr>
<td>(pp \rightarrow W^- + X \rightarrow \tau^-\nu + X)</td>
<td>WminusToTauNu_CT10_8TeV [41]</td>
</tr>
<tr>
<td>(pp \rightarrow tt + X \rightarrow 2l2\nu 2b + X)</td>
<td>TTJets_FullLeptMGDecays_TuneP11TeV_8TeV [42]</td>
</tr>
<tr>
<td>(pp \rightarrow tt + X \rightarrow 1l1\nu 2q2b + X)</td>
<td>TTJets_SemiLeptMGDecays_8TeV [43]</td>
</tr>
<tr>
<td>(pp \rightarrow tt + X \rightarrow 4q2b + X)</td>
<td>TTJets_HadronicMGDecays_TuneP11mpiHi_8TeV [44]</td>
</tr>
<tr>
<td>(pp \rightarrow WW + X \rightarrow 2l2\nu + X)</td>
<td>WWJetsTo2L2Nu_TuneZ2star_8TeV [45]</td>
</tr>
<tr>
<td>(pp \rightarrow WZ + X \rightarrow 3l1\nu + X)</td>
<td>WZJetsTo3LNu_8TeV_TuneZ2Star [46]</td>
</tr>
<tr>
<td>(pp \rightarrow ZZ + X \rightarrow 4\mu + X)</td>
<td>ZZTo4mu_8TeV [47]</td>
</tr>
</tbody>
</table>
Object Calibration (1/2)

- Lepton momentum scale calibrated using the Z boson invariant mass spectrum
  - Scale, offset, resolution parameters in three bins of eta

- Data/MC correction factors as well as uncertainties on lepton efficiencies (reco-, ID-, trigger) taken from CMS publications
Jet calibration is already included in *Bacon*-framework
- Note: CMS Open Data also provides recipe for these corrections
- Uncertainties taken from CMS publications: Tested in Z+jet events

$E_T^{\text{Miss}}$ calibration also available in *Bacon*-framework
- Uncertainties on soft-component taken from publications
- Uncertainties on hard-components transferred from jets
Electroweak and Top-Quark related backgrounds are estimated using MC samples and their corresponding cross-sections
- Cross-Section and event-filter information had to be “googled” in thesis and public Twikis

Multi-jet background contributions are estimated using an ABCD method using
- Isolation of leptons
- Charge of leptons
- Background enhanced regions, such as small $M_T$ or small $E_T^{\text{Miss}}$
Results and Precision

- In total 7 fiducial cross-section measured with systematic uncertainties between 1.6% and 6.7%
  - Dominant uncertainties due to $E_T^{\text{Miss}}$ and Jet calibration
  - Background uncertainties not dominant
  - Add. Luminosity of 2.5%

<table>
<thead>
<tr>
<th>Process</th>
<th>cross section $[\text{pb}]$ (stat. ± sys. ± lumi.)</th>
<th>Prediction $[\text{pb}]$ (signal MC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z/\gamma^* \rightarrow e^+e^-$</td>
<td>$\sigma^{\text{fid}} = 461 \pm 17$ $(1 \pm 13 \pm 11)$</td>
<td>$\sigma^{\text{fid.}} = 450 \pm 0.02$</td>
</tr>
<tr>
<td>$Z/\gamma^* \rightarrow \mu^+\mu^-$</td>
<td>$\sigma^{\text{fid}} = 406 \pm 12$ $(1 \pm 6 \pm 10)$</td>
<td>$\sigma^{\text{fid.}} = 400 \pm 0.01$</td>
</tr>
<tr>
<td>$Z/\gamma^* \rightarrow \mu^+\mu^- + \geq 1 \text{ jet}$</td>
<td>$\sigma^{\text{fid}} = 77.1 \pm 5.5$ $(0.4 \pm 5.1 \pm 1.9)$</td>
<td>$\sigma^{\text{fid.}} = 76.3 \pm 5.0$</td>
</tr>
<tr>
<td>$W^+ \rightarrow \mu^+\nu$</td>
<td>$\sigma^{\text{fid}} = 3052 \pm 124$ $(1 \pm 98 \pm 76)$</td>
<td>$\sigma^{\text{fid.}} = 3015 \pm 100$</td>
</tr>
<tr>
<td>$W^- \rightarrow \mu^-\nu$</td>
<td>$\sigma^{\text{fid}} = 2103 \pm 86$ $(1 \pm 69 \pm 52)$</td>
<td>$\sigma^{\text{fid.}} = 2105 \pm 60$</td>
</tr>
<tr>
<td>$t\bar{t} \rightarrow \mu^+e^+\nu\bar{\nu}bb$</td>
<td>$\sigma^{\text{fid}} = 4.54 \pm 0.35$ $(0.14 \pm 0.30 \pm 0.11)$</td>
<td>$\sigma^{\text{fid.}} = 4.37 \pm 0.35$</td>
</tr>
<tr>
<td>$W^\pm Z \rightarrow l^\pm\nu l^+l^-$</td>
<td>$\sigma^{\text{fid}} = 28.1 \pm 3.3$ $(3.1 \pm 0.9 \pm 0.7)$</td>
<td>$\sigma^{\text{fid.}} = 23.7 \pm 0.4$</td>
</tr>
</tbody>
</table>
Results and Comparisons

- All results are consistent with
  - theory prediction at NNLO
  - official measurements results of ATLAS and CMS

- Only WZ cross-section stat. limited
  - Smaller experimental uncertainties on leptons compared to official measurement of CMS

- $W^+ / W^-$ results are compared to ATLAS due to similar fiducial volume definition

- Note: cross-section ratios are difficult to extract, as we don’t have knowledge on correlations
Opportunities and Challenges of CMS Open Data

- Cross section measurements at ≈5% precision level possible
  - Certainly difficult to perform precision measurements

- Possible improvements
  - Providing simplified Data/MC corrections as well as uncertainties
  - Providing several public standard analyses as a guide
    - NTuple-Maker already exists: https://github.com/cms-opendata-analyses
  - Tutorials (if there is enough interest)
  - Add (information on) MC samples
    - Cross-sections / generator information / ….
  - Some basic samples, e.g. W→ev
But have not all interesting Cross-Sections already measured?

- Only rarely cross-sections in extreme phase-space regions are measured
  - Typical example: Control-regions in searches, where we observe large differences between predictions and reality
    - Here we would learn a lot of QCD
  - Nice example: Z/W/Top Control regions of the latest ATLAS LQ Analysis
    - Signal selection: 2 leptons ($p_T>$40, 65 GeV), 2 jets ($p_T>$60 GeV)+BDT(LQ$_{\text{Mass'...}}$)
How to obtain QCD cross-sections in other regions?

- Until to very recently: Get in contact with experimentalists from ATLAS or CMS
  - Either convince them to do this measurements for you
  - Get (short-time) member of the experiment yourself and guide the experimentalists what they should measure

- With the CMS Open Data initiative, you can (in principle) measure cross-sections yourself
  - Certainly CMS Open Data has also many other interesting applications - this is just an example :)
Summary

- First re-measurement of SM cross sections in proton-proton collisions at 8 TeV using CMS Open Data presented
  - Precision of 2-6% reached
  - Consistent with SM predictions as well as official measurements

- CMS Open Data offers the unique opportunity for cross section measurements (and further low-precision analyses) by (trained) physicists outside of the LHC collaborations

- Some suggestions discussed for the future improvement of the Open Data initiative.