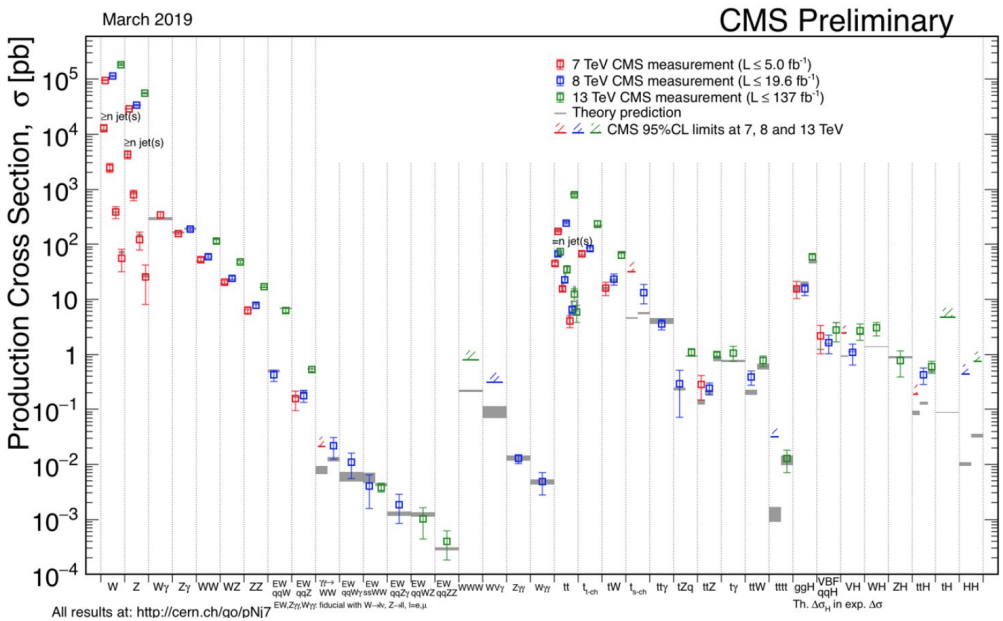
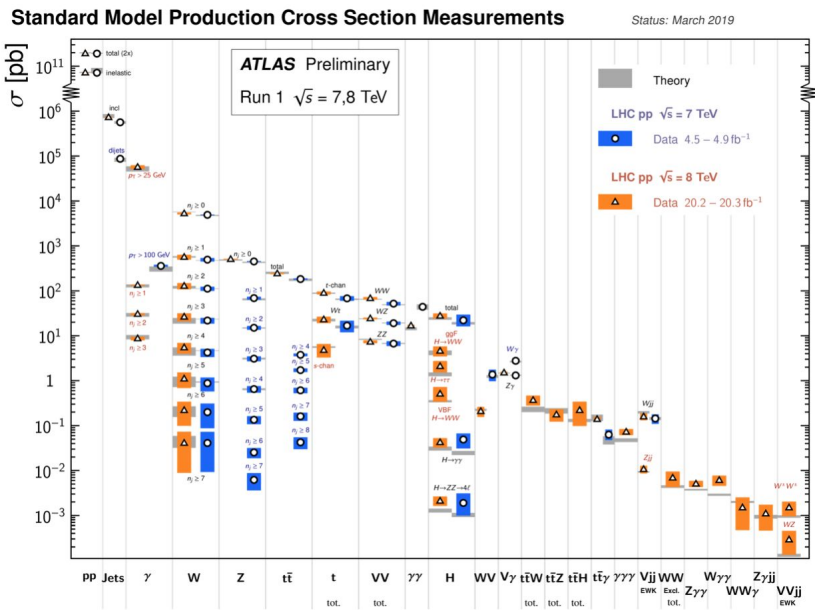
JOHANNES GUTENBERG UNIVERSITÄT MAINZ

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

Aram Apyan, Markus Klute, Matthias Schott

Standard Model Cross Section Measurements



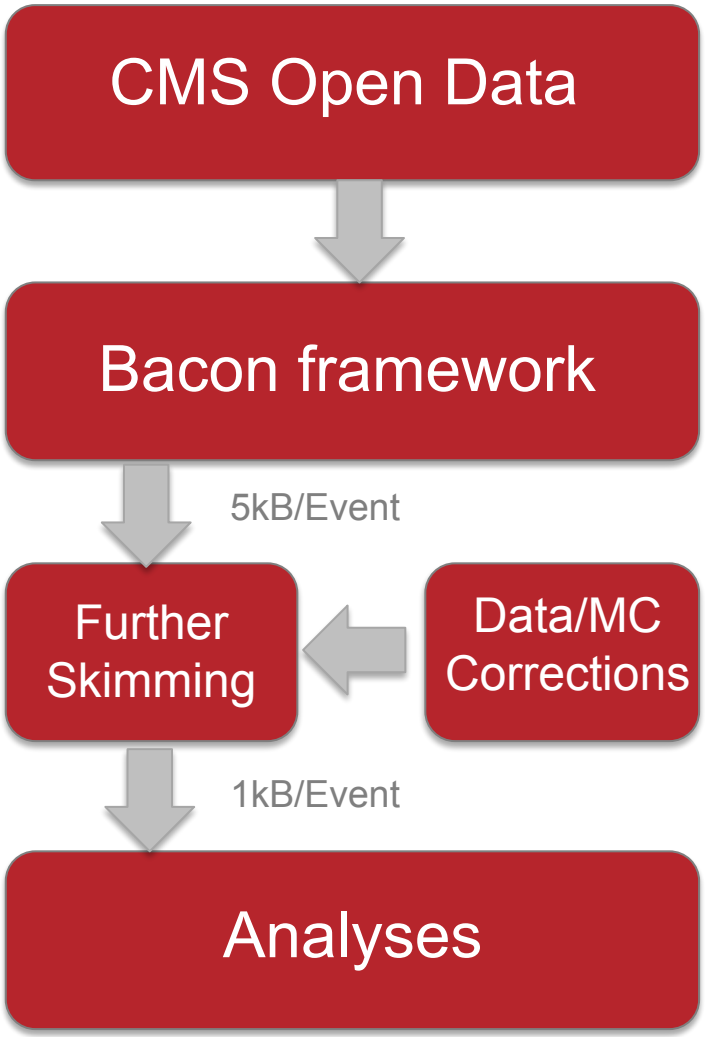


- Typical cross section measurements select a signal process such, that it maximizes statistics and minimizes backgrounds
  - E.g. Z boson selection: 2 muons with p_T>20 GeV, |η|<2.5, 60<m_{ll}<120 GeV
- Measurements are important to test perturbative QCD predictions at (N)NNLO
  - Experimental precisions at 1% level, however, inclusive measurements dominated by luminosity uncertainties

The CMS Open Data Project

- 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/BaconProd/tree/Run1



Repeating SM Cross Section Measurements

- Measurement of inclusive Cross-Sections straight forward

$$\sigma_V^{incl} = \frac{N_{signal}}{\epsilon \cdot BR \cdot \int Ldt}$$

- 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

Process	Definition of fid. phase-space
$Z/\gamma^* \to e^+e^-$	$(1e^+1e^-)$, $60 < m_{ee} < 120$ GeV, $p_T^e > 25$ GeV, $\lvert\eta^e\rvert < 2.1$
$Z/\gamma^* \to \mu^+\mu^-$	$(1\mu^+1\mu^-)$, $60 < m_{\mu\mu} < 120$ GeV, $p_T^\mu > 25$ GeV, $\lvert\eta^\mu\rvert < 2.1$
$Z/\gamma^* \to \mu^+\mu^-$ $+ \geq 1jet$	$(1\mu^+1\mu^-)$, $70 < m_{\mu\mu} < 110$ GeV, $p_T^\mu > 20$ GeV, $\lvert\eta^\mu\rvert < 2.4$, $p_T^{jet} > 30$ GeV, $\lvert y^{jet}\rvert < 2.4$ $\Delta R(j,l) > 0.5$
$W^+ \to \mu^+\nu$	$(1\mu^+)$, $p_T^\mu > 25$ GeV, $\lvert\eta^\mu\rvert < 2.4$, $p_T^\nu > 25GeV$, $m_T > 40$ GeV
$W^- \to \mu^-\nu$	$(1\mu^-)$, $p_T^\mu > 25$ GeV, $\lvert\eta^\mu\rvert < 2.4$, $p_T^\nu > 25GeV$, $m_T > 40$ GeV
$t\bar{t} \to \mu^\mp e^\pm \nu\bar{\nu}b\bar{b}$	$1\mu^\pm, 1e^\mp$, $p_T^l > 20GeV$, $\lvert\eta^\mu\rvert < 2.4$,
$W^\pm Z \to l^\pm\nu l^+l^-$ $(l = e.\mu)$	$(e^\pm e^\mp \mu^\pm), (\mu^\pm\mu^\mp e^\pm), (\mu^\pm\mu^\mp\mu^\pm)$, $p_T^l > 25$ GeV, $\lvert\eta^l\rvert < 2.5$, $80 < m_{ll} < 100$ GeV, $m_T > 40$ GeV

- Decided to re-measure
  - Z→ll (to test lepton performance)
  - Z→μμ+jets (to test jet performance)
  - W→μν (to test E_T^{Miss} and trigger)
  - Top-pair production (to test b-jets)
  - WZ (to test leptons)

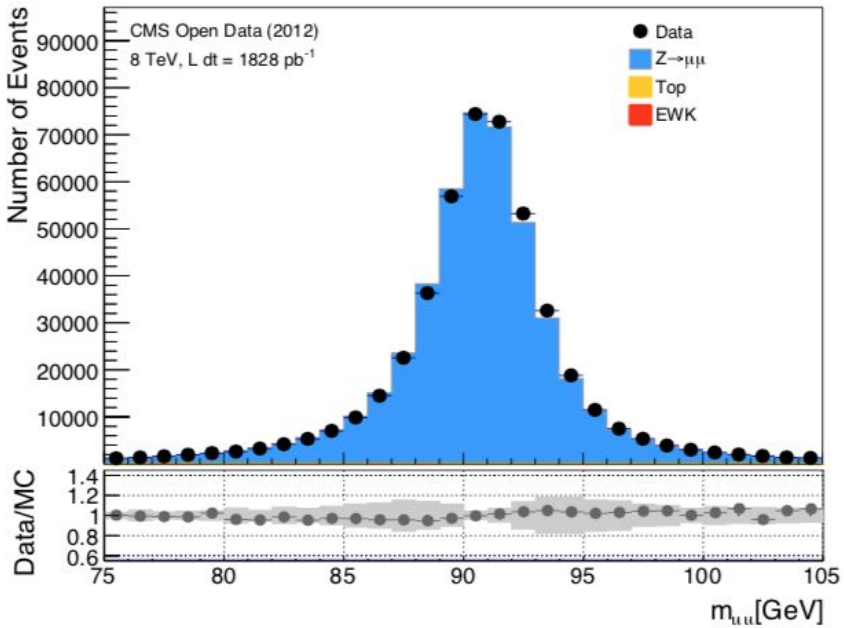
Data and MC Samples

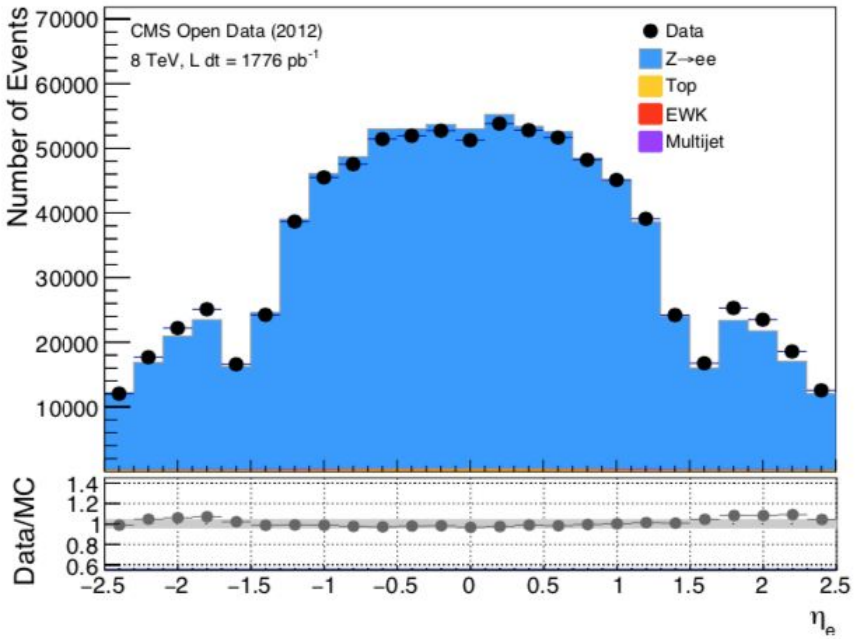
- In total 1.7-1.8 fb^{-1} of data analyzed
  - Mainly used muon-triggered data-stream
- Sufficient fraction of MC skimmed, so that MC stat. uncertainties are not dominant

Process	Dataset Name
$pp \to Z/\gamma^* + X \to e^+e^- + X$	DYToEE_M-20_CT10_TuneZ2star_v2_8TeV [35]
$pp \to Z/\gamma^* + X \to \mu^+\mu^- + X$	DYToMuMu_M-20_CT10_TuneZ2star_v2_8TeV [36]
$pp \to Z/\gamma^* + X \to l^+l^- + X$	DYJetsToLL_M-50_TuneZ2Star_8TeV [37]
$pp \to W^+ + X \to \mu^+\nu + X$	WplusToMuNu_CT10_8TeV [38]
$pp \to W^- + X \to \mu^-\nu + X$	WminusToMuNu_CT10_8TeV [39]
$pp \to W^+ + X \to \tau^+\nu + X$	WplusToTauNu_CT10_8TeV [40]
$pp \to W^- + X \to \tau^-\nu + X$	WminusToTauNu_CT10_8TeV [41]
$pp \to t\bar{t} + X \to 2l2\nu2b + X$	TTJets_FullLeptMGDecays_TuneP11TeV_8TeV [42]
$pp \to t\bar{t} + X \to 1l1\nu2q2b + X$	TTJets_SemiLeptMGDecays_8TeV [43]
$pp \to t\bar{t} + X \to 4q2b + X$	TTJets_HadronicMGDecays_TuneP11mpiHi_8TeV [44]
$pp \to WW + X \to 2l2\nu + X$	WWJetsTo2L2Nu_TuneZ2star_8TeV [45]
$pp \to WZ + X \to 3l1\nu + X$	WZJetsTo3LNu_8TeV_TuneZ2Star [46]
$pp \to ZZ + X \to 4\mu + X$	ZZTo4mu_8TeV [47]

Data stream / Trigger name	Dataset Name	$\int Ldt$ [pb^{-1}]
single muon trigger (HLT_IsoMu24, HLT_IsoMu24_eta2p1)	/SingleMu/Run2012C-22Jan2013-v1 [20] /SingleMu/Run2012B-22Jan2013-v1 [21]	1,828
single electron trigger (HLT_Ele27_WP80)	/SingleElectron/Run2012B-22Jan2013-v1 [22] /SingleElectron/Run2012C-22Jan2013-v1 [23]	1,776

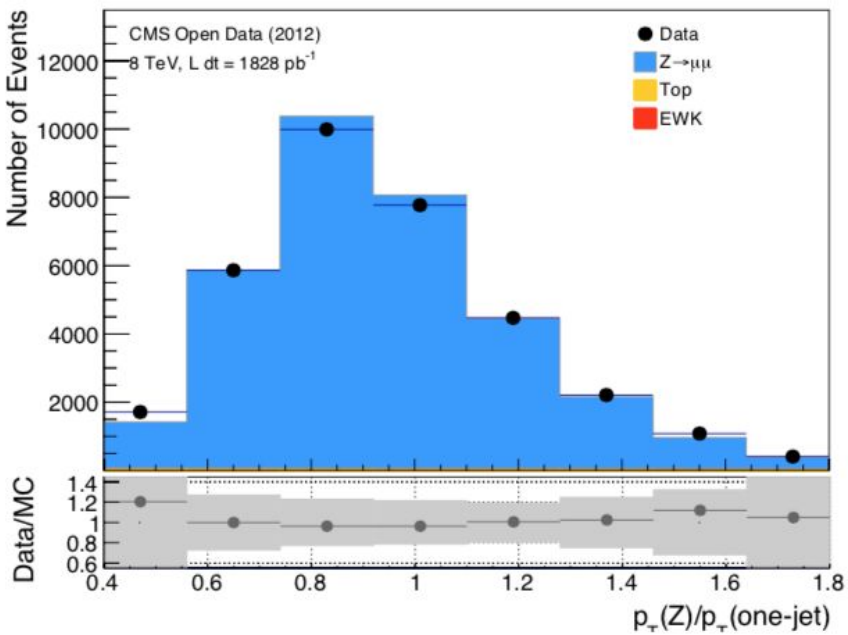
Object Calibration (1/2)

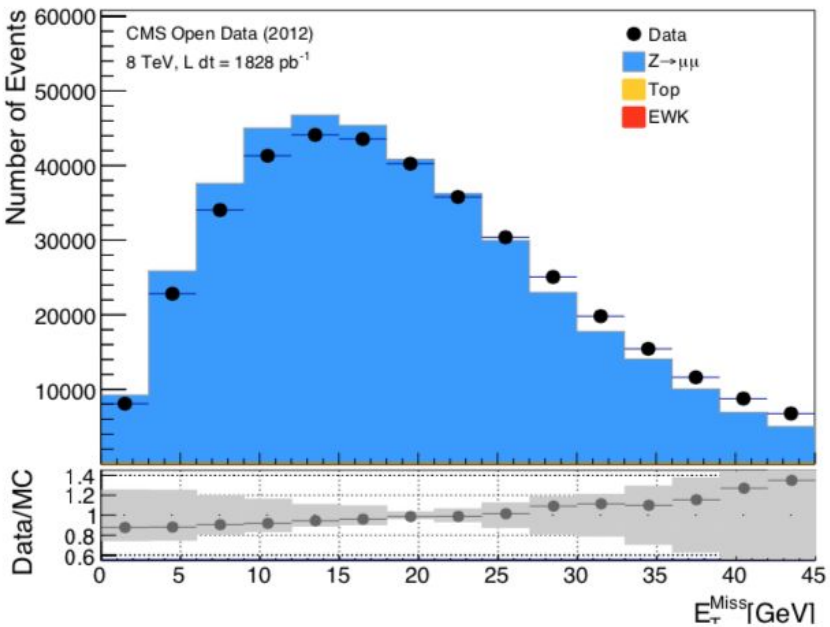




- Lepton momentum scale calibrated using the Z boson invariant mass spectrum
  - Scale, offset, resolution parameters in three bins of eta
- Data/MC correction factors as well as uncertainties on lepton efficiencies (reco-, ID-, trigger) taken from CMS publications

Object Calibration (2/2)





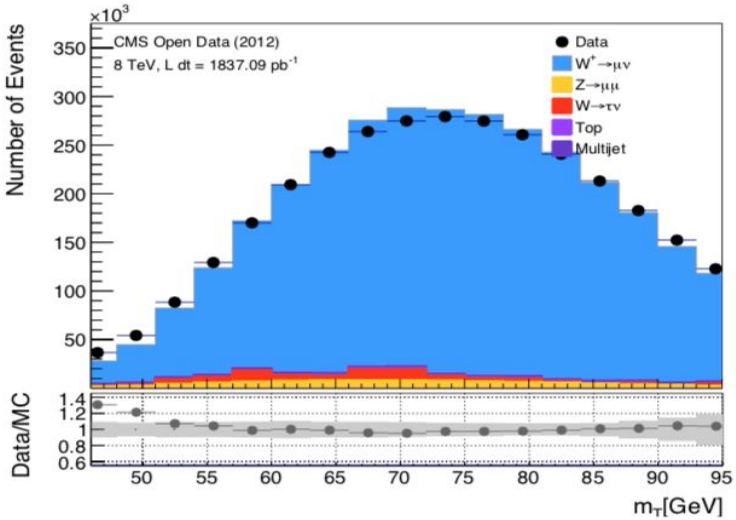
- Jet calibration is already included in *Bacon*-framework
  - Note: CMS Open Data also provides recipe for these corrections
  - Uncertainties taken from CMS publications: Tested in Z+jet events

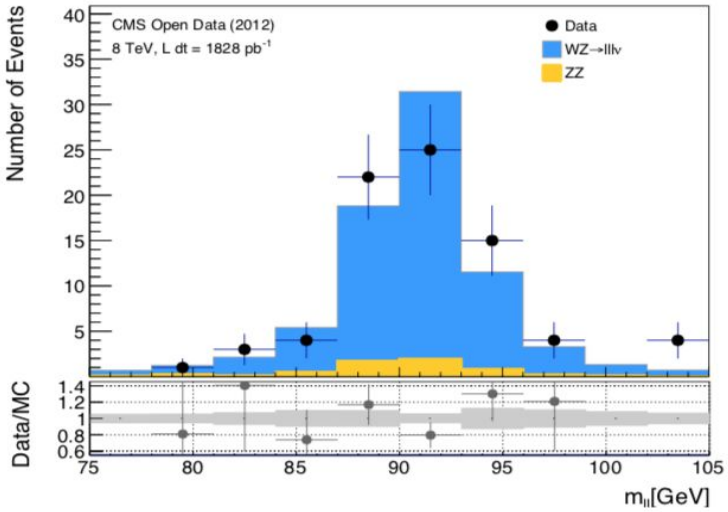
- E_T^{Miss} calibration also available in *Bacon*-framework
  - Uncertainties on soft-component taken from publications
  - Uncertainties on hard-components transferred from jets

Background Estimations

- 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^{Miss}





Results and Precision

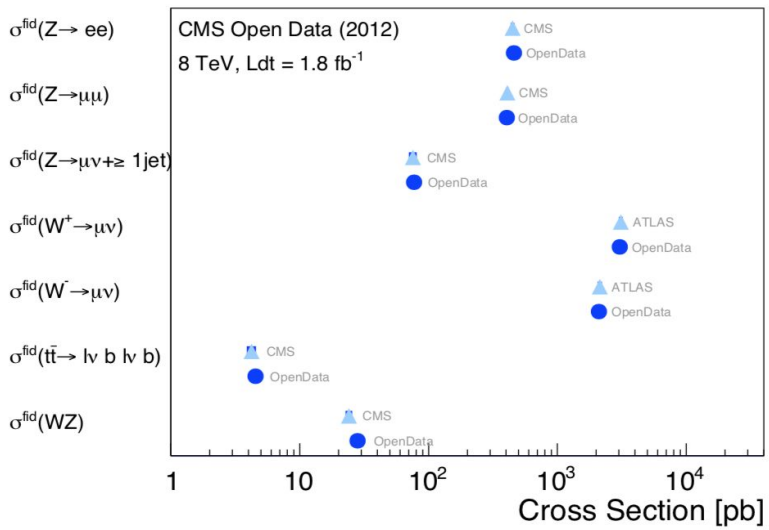
- In total 7 fiducial cross-section measured with systematic uncertainties between 1.6% and 6.7%
  - Dominant uncertainties due to E_T^{Miss} and Jet calibration
  - Background uncertainties not dominant
  - Add. Luminosity of 2.5%

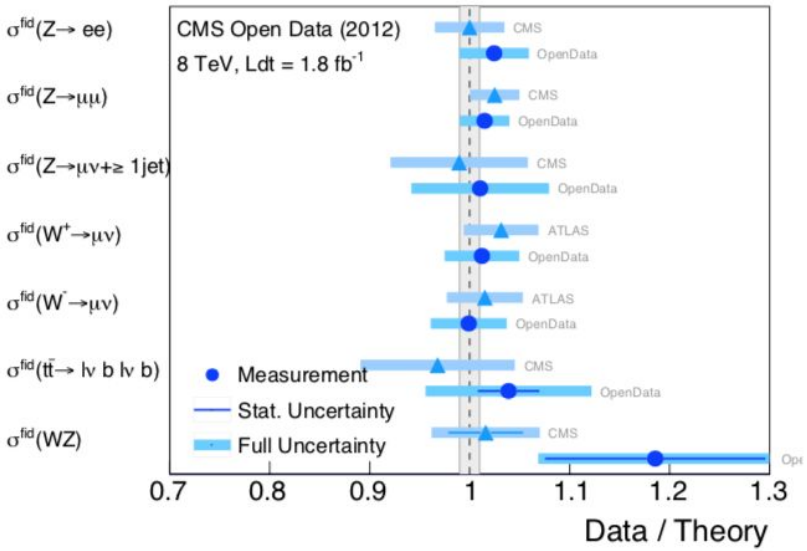
Process	cross section [pb] ($stat. \pm sys. \pm lumi.$)	Prediction [pb] (signal MC)
$Z/\gamma^* \to e^+e^-$	$\sigma^{fid} = 461 \pm 17$ $(1 \pm 13 \pm 11)$	$\sigma^{fid.} = 450 \pm 0.02$
$Z/\gamma^* \to \mu^+\mu^-$	$\sigma^{fid} = 406 \pm 12$ $(1 \pm 6 \pm 10)$	$\sigma^{fid.} = 400 \pm 0.01$
$Z/\gamma^*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

Process	Elec. Eff.	Elec. Scale/ Res.	Muon Eff.	Muon Scale/ Res.	JES/ JER	$\not{E}_T$	b-tag-ging	pile-up	Total
$Z/\gamma^* \to e^+e^-$	2.9%	0.2%	-	-	-	-	-	0.1%	2.9%
$Z/\gamma^* \to \mu^+\mu^-$	-	-	1.5%	0.3%	-	-	-	0.1%	1.6%
$Z/\gamma^* \to \mu^+\mu^- + \geq 1$ jet	-	-	1.5%	0.4%	6.5%	-	-	0.3%	6.7%
$W^+ \to \mu^+\nu$	-	-	0.8%	0.2%	-	2.2%	-	1.8%	2.9%
$W^- \to \mu^-\nu$	-	-	0.8%	0.2%	-	2.2%	-	1.8%	2.9%
$t\bar{t} \to \mu^\mp e^\pm \nu\bar{\nu}b\bar{b}$	1.5%	0.2%	1.0%	0.2%	5.2%	1.4%	3%	1.9%	6.7%
$W^\pm Z \to l^\pm \nu l^+ l^-$ $(l = e.\mu)$	1.3%	0.2%	1.5%	0.3%	-	1.3%	-	1.8%	3.0%

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

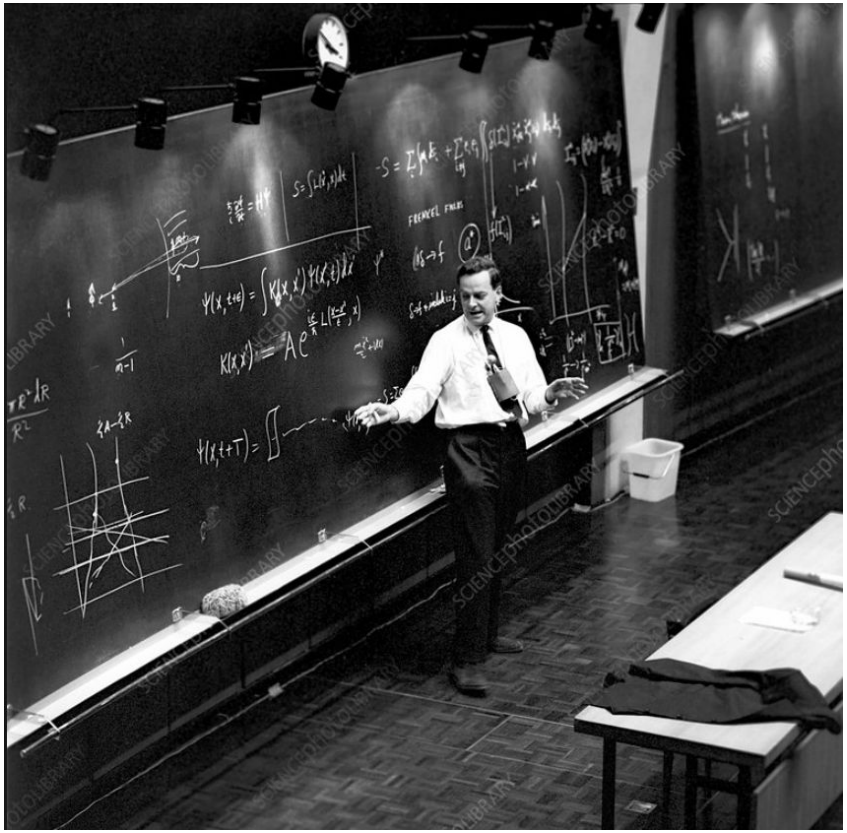




Opportunities and Challenges of CMS Open Data

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

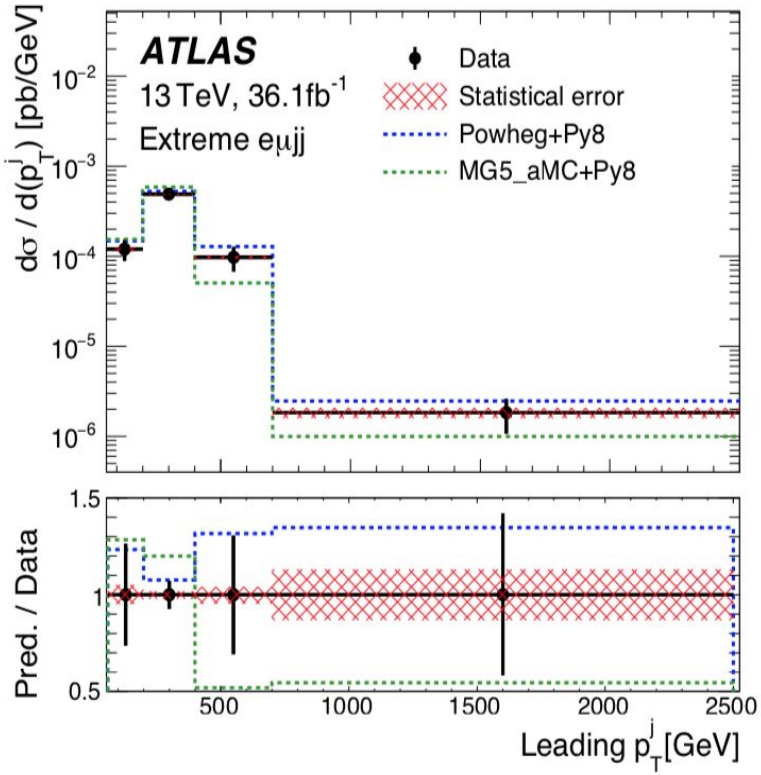
- Possible improvements
  - Providing simplified Data/MC corrections as well as uncertainties
  - Providing several public standard analyses as a guide
    - NTuple-Maker already exists: https://github.com/cms-opendata-analyses
  - Tutorials (if there is enough interest)
  - Add (information on) MC samples
    - Cross-sections / generator information / ….
    - Some basic samples, e.g. W→eν



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

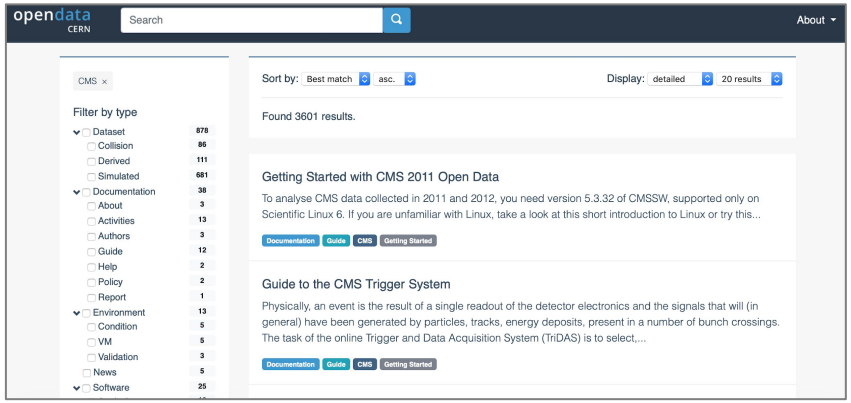
- Nice example: Z/W/Top Control regions of the latest ATLAS LQ Analysis
  - Signal selection: 2 leptons (p_T>40, 65 GeV), 2 jets (p_T>60 GeV)+BDT(LQ_{Mass}, ...)
  - Paper: https://arxiv.org/abs/1902.00377



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 :)



JOHANNES GUTENBERG UNIVERSITÄT MAINZ

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.