Validation of 2011 open data using measurements of top quark pair and $W + c$ production

Oleksandr Zenaiev

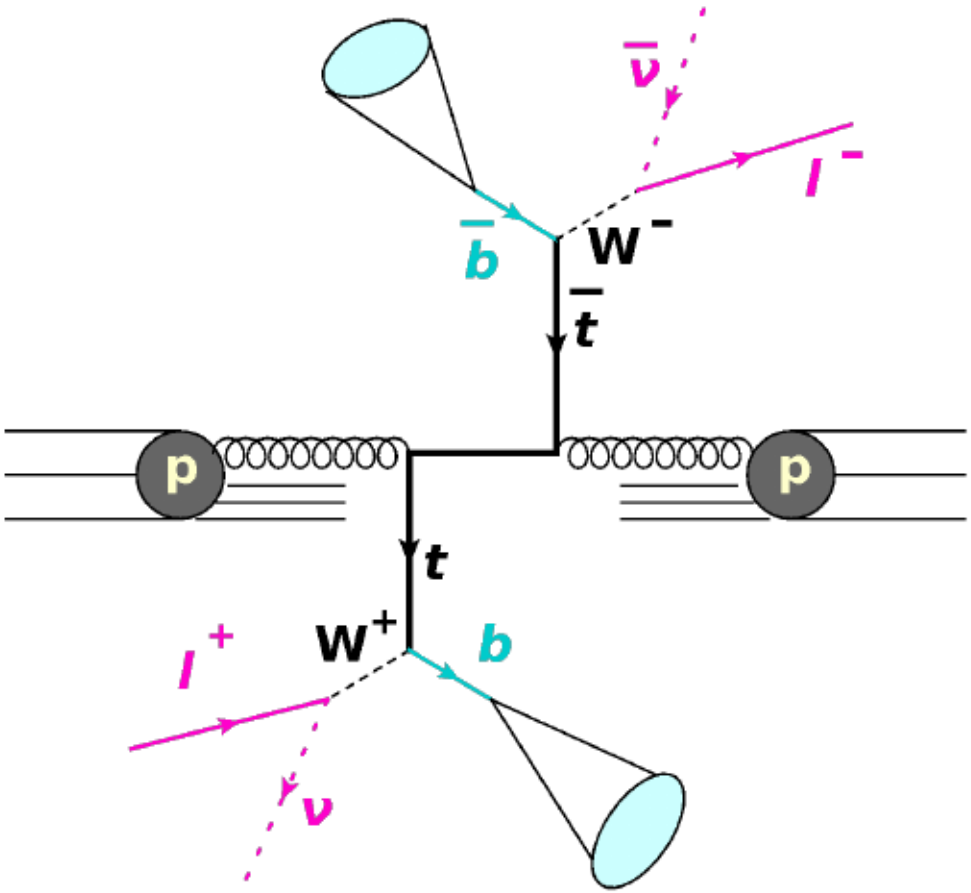
Overview:

- Measurements of $t\bar{t}$ production in dilepton channel at 7 TeV [TOP-11-013, TOP-13-004]
- Measurement of associated $W + c$ production [SMP-12-002]

Open data validation meeting
20.10.2016

- Using 2011 open data, 2.5fb^{-1} + corresponding MC samples
- **Goal: reproduce published CMS results ('research' mode):**
  - "Measurement of differential top-quark pair production cross sections in pp colisions at $\sqrt{s} = 7$ TeV" [TOP-11-013, arXiv:1211.2220, EPJ C73 (2013) 2339] + total $t\bar{t}$ cross section [TOP-13-004, arXiv:1603.02303, JHEP 1608 (2016) 029]
  - "Measurement of associated $W + c$ production in pp collisions at $\sqrt{s} = 7$ TeV" [SMP-12-002, arXiv:1310.1138, JHEP 02 (2014) 013]
- Running VM on private or office laptop, no usage of CMS, CERN and other resources
- Using information mainly from CMS papers, theses, public twiki pages, sometimes needed to consult analysis notes
- All analysis code has been written completely from scratch: no usage of CMS code, except for CMSSW provided with VM

Measurement of differential top-quark pair production cross sections in pp colisions at $\sqrt{s} = 7$ TeV



Dilepton channel: 3 possible final states:

- $e^{\pm}e^{\mp}$ + 2 b-jets + 2 neutrinos
- $\mu^{\pm}\mu^{\mp}$ + 2 b-jets + 2 neutrinos
- $e^{\pm}\mu^{\mp}$ + 2 b-jets + 2 neutrinos

$\Rightarrow$ selecting events with two leptons, two jets, also expect missing transverse energy (MET)

t and $\bar{t}$ are then reconstructed from measured final state particles and assumptions on m_W, m_t

Data samples:

- $e^{\pm}e^{\mp}$: '`DoubleElectron`'
- $\mu^{\pm}\mu^{\mp}$: '`DoubleMu`'
- $e^{\pm}\mu^{\mp}$: '`MuEG`'

Primary vertex:

- dof > 4
- impact parameter < 2 cm in transverse plane and < 24 cm in z

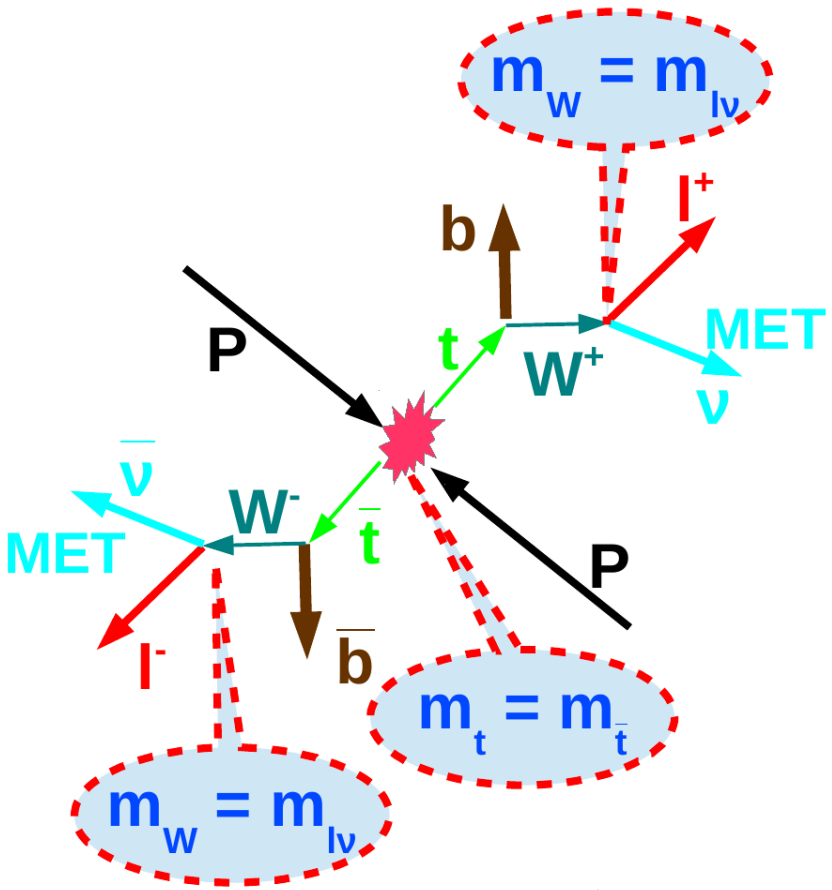
Leptons: two leading p_T opposite signed $p_T > 20$ GeV, $|\eta| < 2.4$

- Electrons ('`gsfElectrons`')
  - isolated with $I_{\text{rel}}^{\Delta R<0.3} < 0.17$ ($\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$)
  - no missing hits in the silicon tracker
- Muons ('`muons`')
  - isolated with $I_{\text{rel}}^{\Delta R<0.3} < 0.20$
  - required to be global muons
  - at least 10 valid tracker hits and 2 pixel hits
  - global track fit χ^2/dof < 10
  - impact parameter to PV $<$ 0.02 cm in transverse plane and $<$ 0.5 cm in z
- $M(ll) > 12$ GeV
- in ee and $\mu\mu$ channels exclude $16 < M(ll) < 106$ GeV
- in ee and $\mu\mu$ channels require MET > 30 GeV

Jets ('`ak5PFJets`'): at least two with $p_T > 30$ GeV, $|\eta| < 2.4$

- anti-k_T with clustering parameter 0.5
- at least one b-tagged using CSVL (discriminant $>$ 0.244)
- CSV information stored only for '`ak5CaloJets`': perform matching choosing closest jet in ΔR

Goal: obtain $\vec{p_t}$ and $\vec{p_{\bar{t}}}$



Efficiency determined in signal MC:
$\approx 70\%$ vs 90% in the paper

- Measured input: 2 leptons, 2 jets, MET
- Unknowns: $\bar{p}_\nu$, $\bar{p}_{\bar{\nu}}$ (6)
- Constraints:
  - m_t, $m_{\bar{t}}$ (2)
  - m_{W+}, m_{W-} (2)
  - $(\bar{p}_\nu + \bar{p}_{\bar{\nu}})_T = \mathrm{MET}$ (2)
- For each pair of jets, solve this using the method from [Phys. Rev. D 73 (2006)]
- If there are several solutions in event (either because of many jet combinations, or several solutions for one configuration), prefer:
  - with 2 b-tagged jets
  - with 1 b-tagged jets
  - with highest weight, weight is determined according to the MC neutrino energy spectrum
- Difference from the paper: no m_t scan $\Rightarrow$ worse efficiency due to detector effects

MC samples

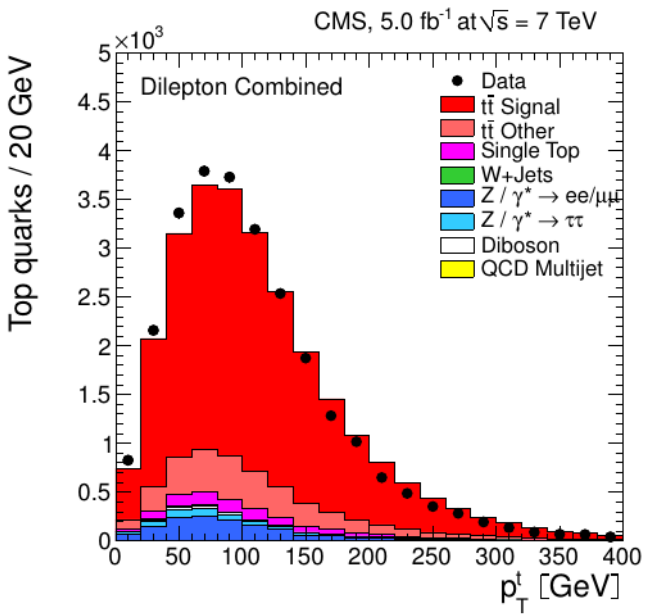
Signal:

- MadGraph + Pythia6, 55M

Background:

- $t\bar{t}$ 'other', mainly via τ decays (MadGraph + Pythia6)
- single top (POWHEG + Pythia6), 1.5M
- Drell-Yan (DY) (MadGraph + Pythia6), 44M
- W + jets (MadGraph + Pythia6), 55M
- Diboson and QCD multijet considered in the paper, but contribute negligibly: not used

TOP-11-013



In total processed:

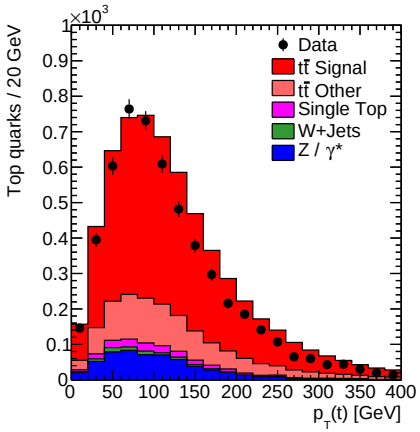
- Data: 33M ($e\mu$) + 50M (ee) + 40M ($\mu\mu$) $\approx$ 123M
- MC: $\approx$ 155M (processing MC took $\times 5$ more time than data: busy events, larger fraction selected)

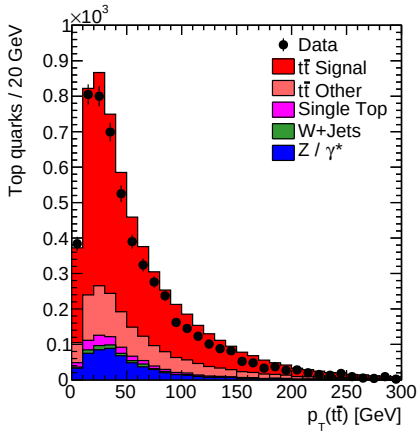
Overall $\sim$ 2 weeks (not CPU time!), running several jobs in parallel on one machine, but also gaps between running jobs. Some jobs needed to be resubmitted.

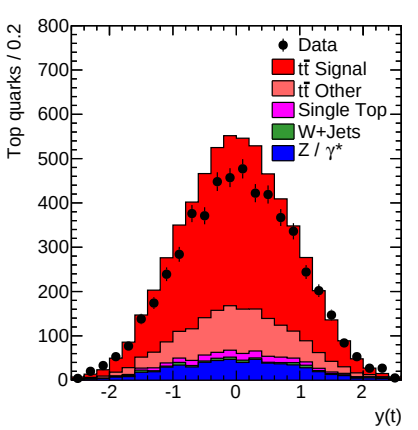
Any improvement here is very desirable: how will it be feasible for more complicated analyses, or 2012 data?..

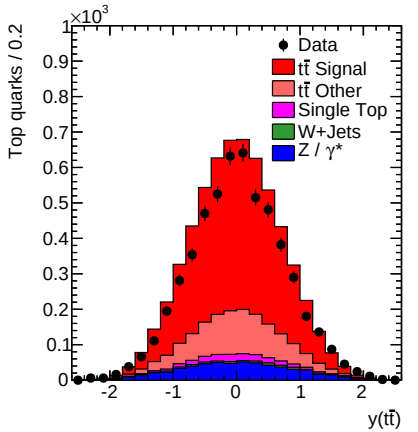
Control distributions

Open Data

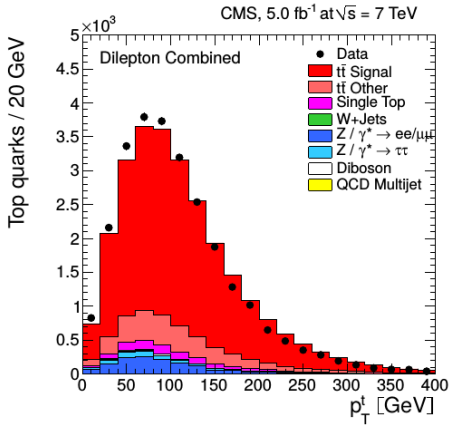


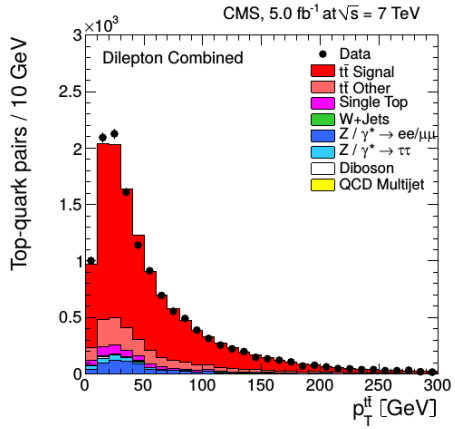


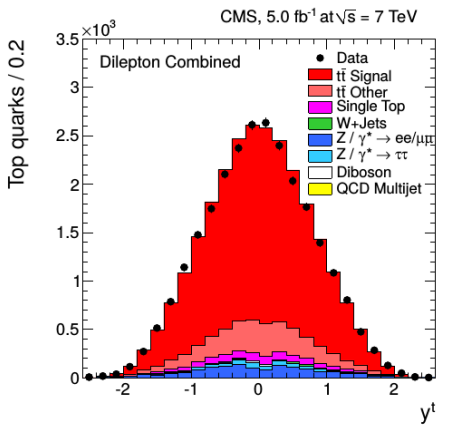


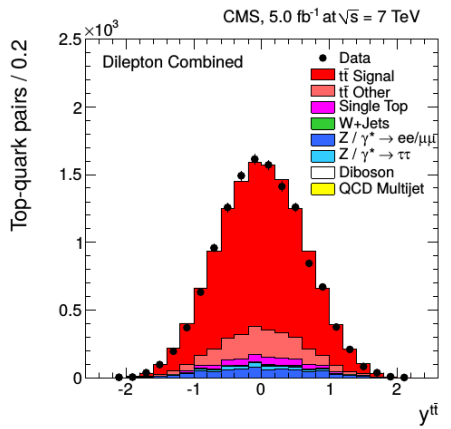


TOP-11-013









- $\approx$ 25% less events than expected from TOP-11-013: consistent with smaller kin. reco efficiency
- Larger MC / data: consistent with missing scale factors, corrections etc.
- Slightly larger background fraction
- Shapes very similar

* $p_T(t)$, $y(t)$ vs $p_T(t), p_T(\bar{t})$ and $y(t), y(\bar{t})$ in paper

Cross section measured at parton level in full phase space:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}Y} = \frac{N_{Sig}}{ALB\Delta Y},\ N_{Sig} = N_{DATA} - N_{MCbackgr},\ E = \frac{N_{MCreco}}{N_{MCgen}}$$

$$\sigma = \int \frac{\mathrm{d}\sigma}{\mathrm{d}Y}$$

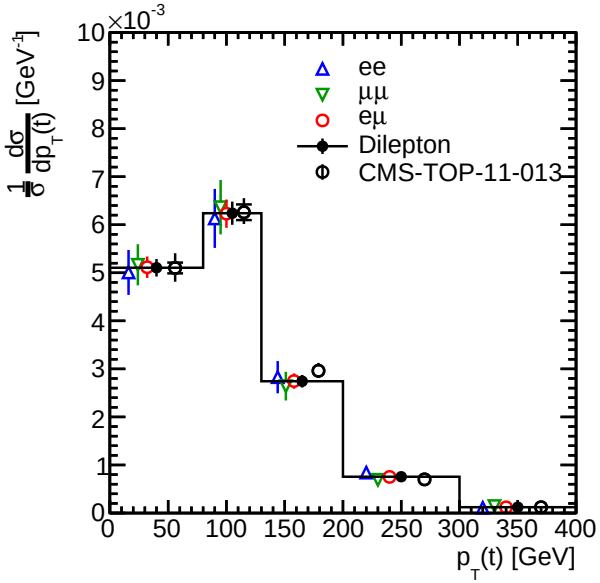
$L = 2.5\ \mathrm{fb}^{-1},\ B = 4.6\%$

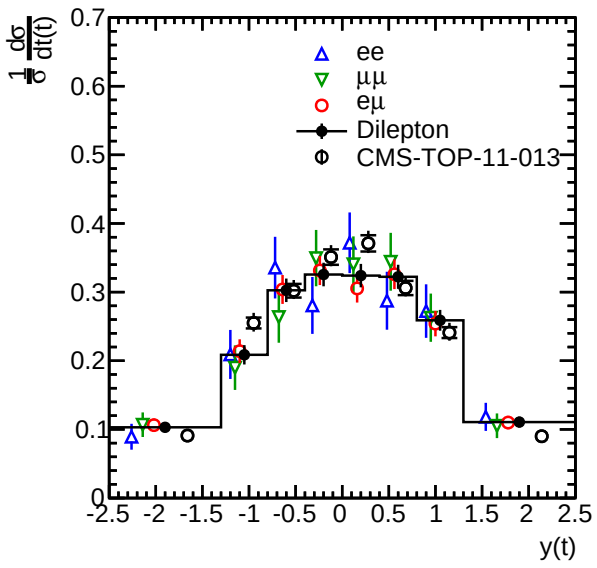
Efficiency E determined as bin-to-bin corrections: no 'unfolding'.
This should give underestimated stat. uncertainties and might bias central values in bins with small purity/stability and poor MC description.

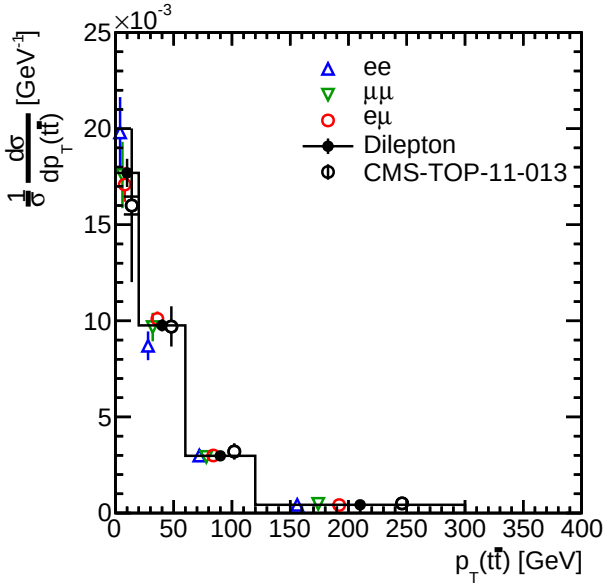
Measure:

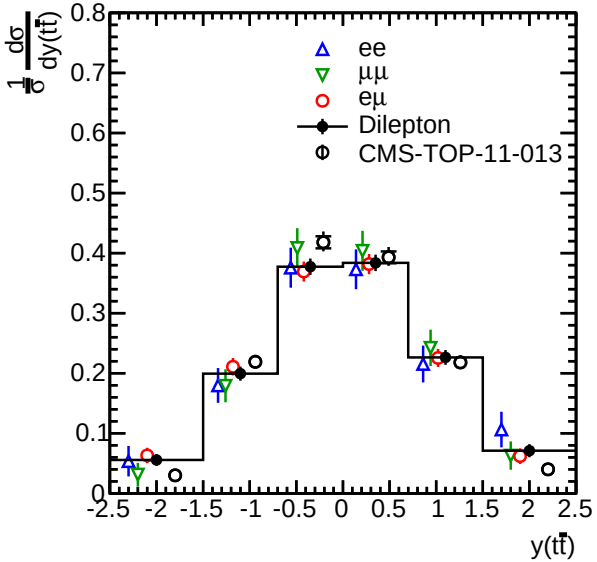
- normalised differential x-section $\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}Y}$ (published in TOP-11-013)
- total x-section (published in TOP-13-004)

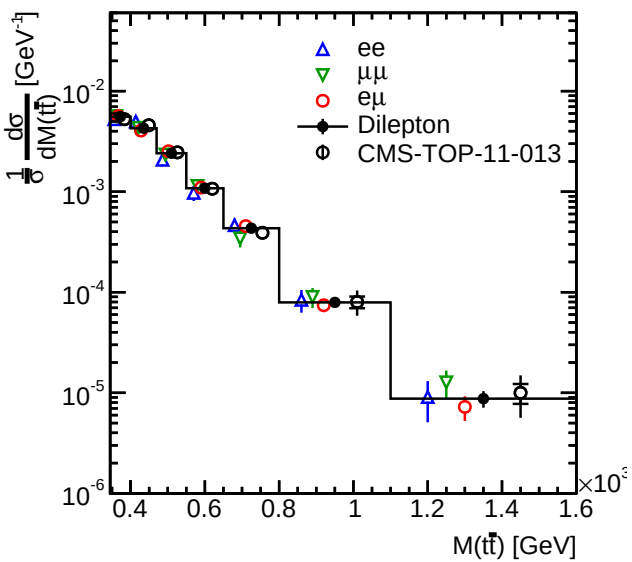
Normalised differential cross sections











- Consistent results in all three channels
- **Consistent with the paper!** (no difference larger than 3 σ)
- Trends at large $|y(t)|$, $|y(t\bar{t})|$: unfolding might be important here

Total cross section

How it is determined:

- In this analysis obtained by integrating over differential x-section
- More sophisticated procedure in the paper

Open data

e.g. by integrating $p_T(t)$ diff. x-section:

ee:	149.7 $\pm$ 8.2 (stat) pb
$\mu\mu$:	136.0 $\pm$ 6.8 (stat) pb
$e\mu$:	155.4 $\pm$ 4.1 (stat) pb
dilepton:	150.4 $\pm$ 3.2 (stat) pb

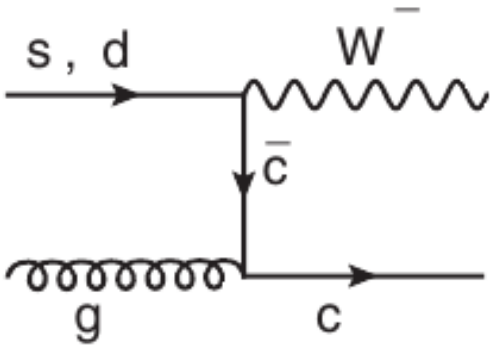
additionally spread between integrations over different variables ≈ 10 fb

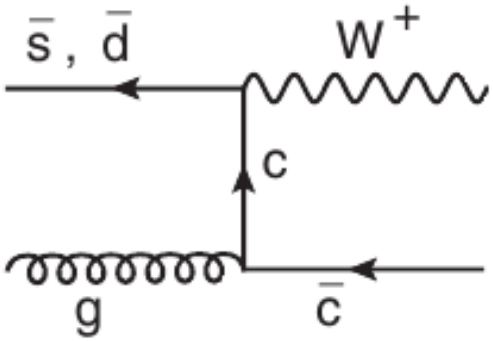
TOP-13-004

$e\mu$ channel only:
173.6 ± 2.1 (stat) $^{+4.5}_{-4.0}$ (sys) ± 3.8 (lum) pb

- Reasonable consistency. In agreement with larger MC / data rate, missing corrections etc.
- Total x-section sensitive to (in)efficiency, scale factor issues (cancel to large extend for normalised x-section).

Measurement of associated $W + c$ production





$W \to e$ or μ
$c \to D^*$, or D^+, or μ in jet

In total 6 final states
$\Rightarrow$ selecting events with isolated lepton and jet with charm 'signature'

Signal: oppositely signed (OS) W and c
Background: expect equal fraction of OS and same signed (SS) W and c
$\Rightarrow$**: suppressed background for OS-SS distributions**

Data samples:

- μ: '`SingleMu`' dataset
- e: '`SingleEl`' dataset

MC:

- W + jets (MadGraph + Pythia6) [signal]
- Drell-Yan (MadGraph + Pythia6) [background] — not included

Event selection

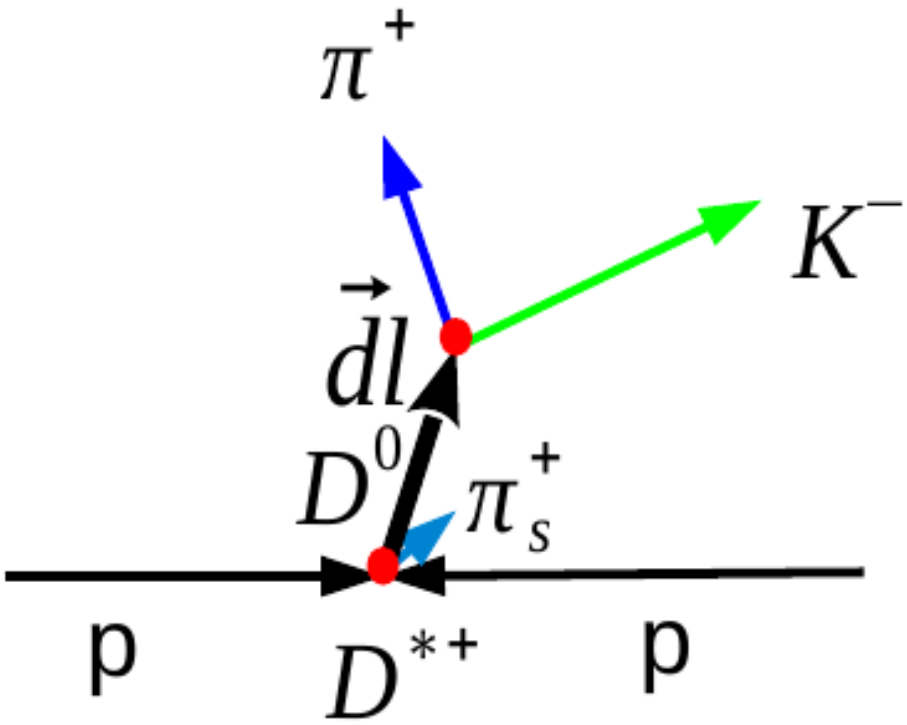
Lepton: leading p_T muon or electron:

- Muons ('`muons`')
  - $p_T > 25$ GeV
  - $|\eta| < 2.1$
  - isolated with $I_{\text{rel}}^{\Delta R<0.4} < 0.12$ ($\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$)
  - $M_T > 40$ GeV, $M_T = \sqrt{2 p_T E_T^{\text{miss}} (1 - \cos(\phi_l - \phi_{E_T^{\text{miss}}}))}$
  - required to be global muons
  - at least 10 valid tracker hits and 2 pixel hits
  - global track fit χ^2/dof < 10
  - impact parameter to PV < 0.02 cm in transverse plane and < 0.5 cm in z
  - no oppoiste sign muon with $p_T > 25$ GeV, $|\eta| < 2.4$
- Electrons ('`gsfElectrons`')
  - $p_T > 35$ GeV
  - $|\eta| < 2.1$, excluding $1.44 < |\eta| < 1.57$
  - $M_T > 55$ GeV, $M_T = \sqrt{2 p_T E_T^{\text{miss}} (1 - \cos(\phi_l - \phi_{E_T^{\text{miss}}}))}$
  - isolated with $I_{\text{rel}}^{\Delta R<0.3} < 0.05$
  - no missing hits in the silicon tracker
  - no oppoiste sign electron with $p_T > 20$ GeV, $|\eta| < 2.5$

Jet ('`ak5PFJets`'): leading p_T with $p_T > 25$ GeV, $|\eta| < 2.5$ and charm 'signature'

- $D^+ \to K^- \pi^+ \pi^+$ candidate
- $D^{*+} \to D^0 \pi_s^+, D^0 \to K^- \pi^+$ candidate
- muon

- **Crucial to suppress background for D^+, D^{*+}**
- Differs from the paper:
  - open data: all track combinations
  - paper: 'exclusive' veritces (no shared tracks)
- Expecting larger selection efficiency, but possibly larger background in open data



$$l_{proj} = \frac{(\vec{dl} \cdot \vec{p}_D)_{xy}}{|\vec{P}_D|_{xy}}, \quad \vec{p}_D = \vec{p}_{K^-} + \vec{p}_{\pi^+}$$

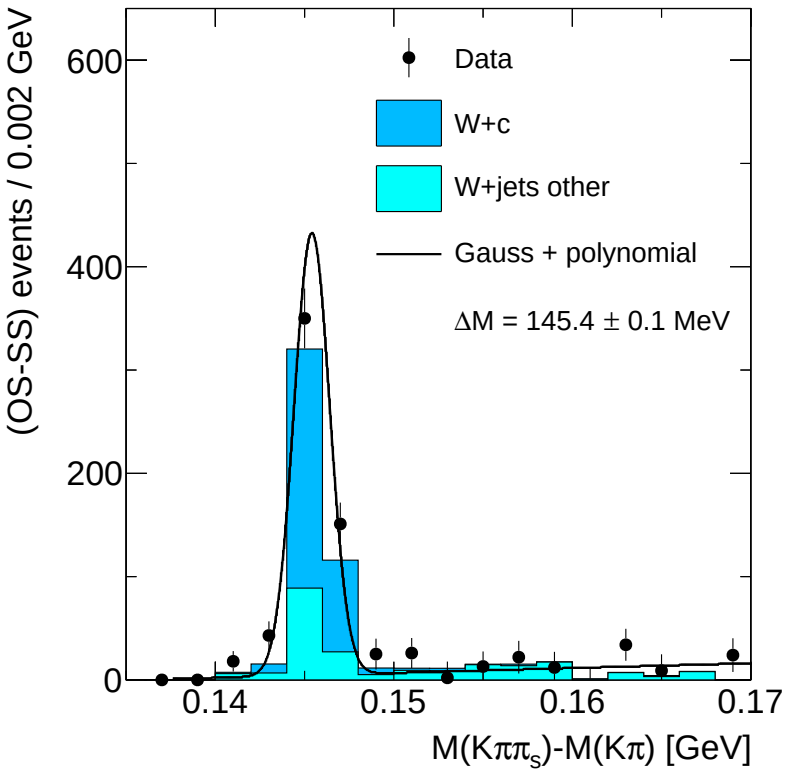
$$s = \frac{l_{proj}}{\Delta l_{proj}}$$

Require:

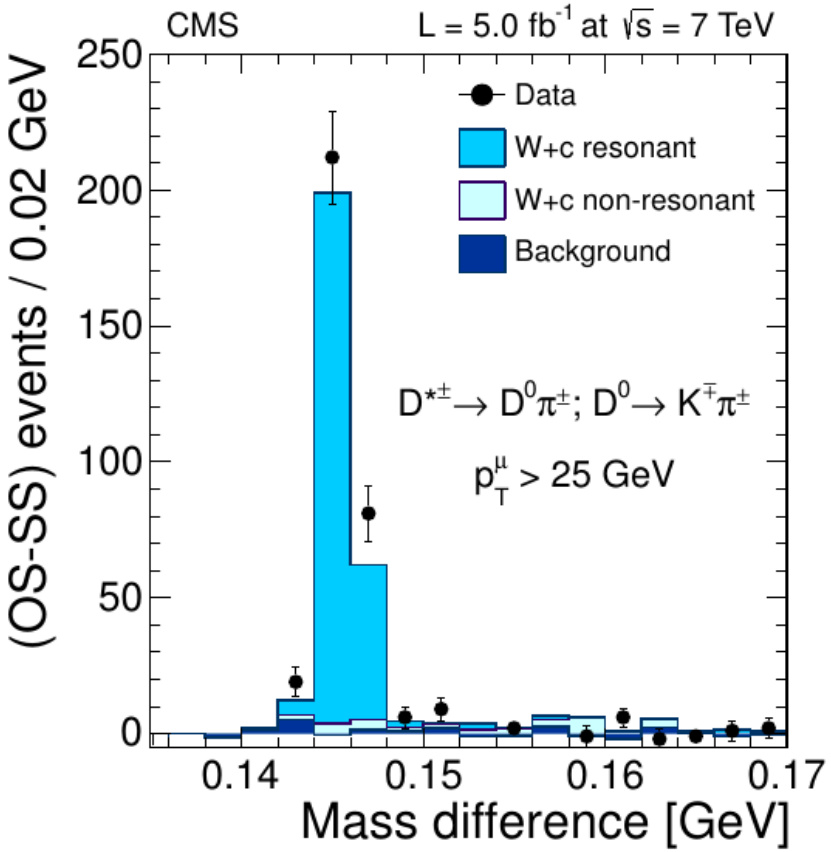
- $s > 3$
- $(dl)_{xy} < 2$ cm, $(dl)_z < 2$ cm

Signal extraction: D^*

Open Data



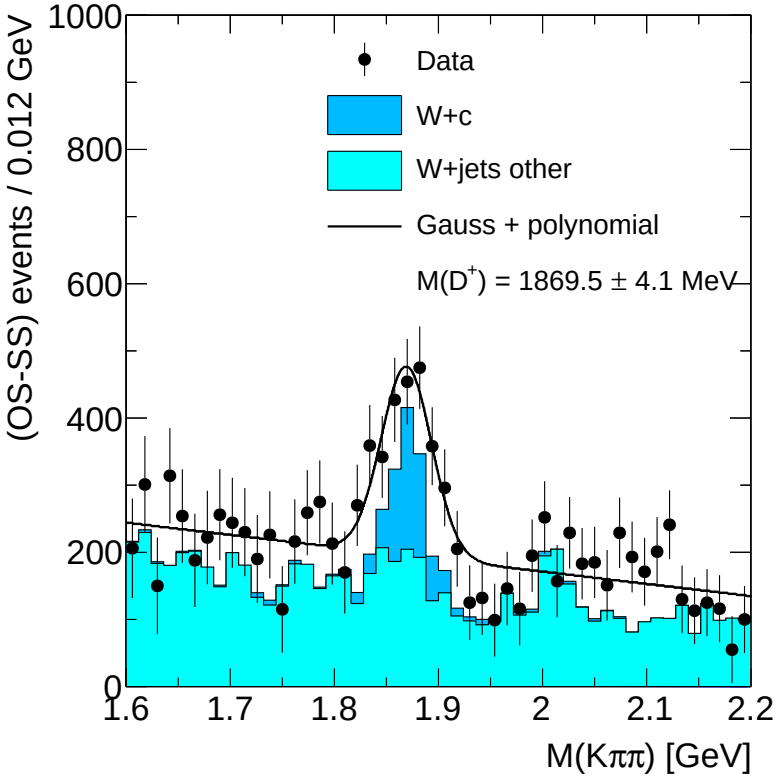
TOP-12-002



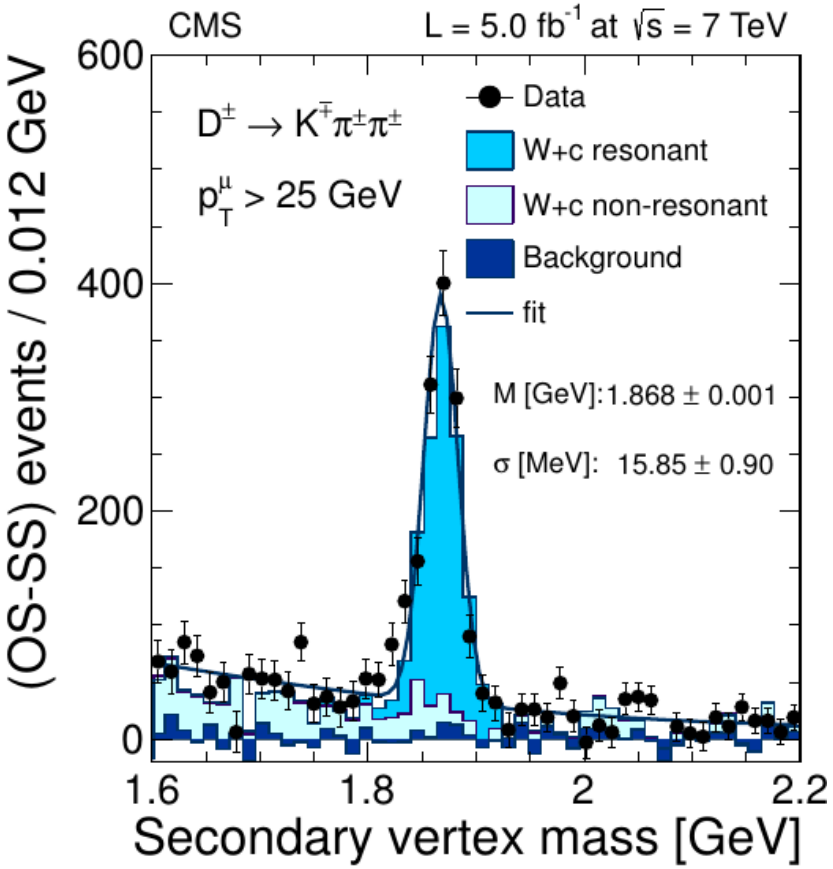
- Fitted to Gaus (signal) + 1st order polynomial (background)
- More signal events in open data
- Similar signal / background ratio

Signal extraction: D^+

Open Data



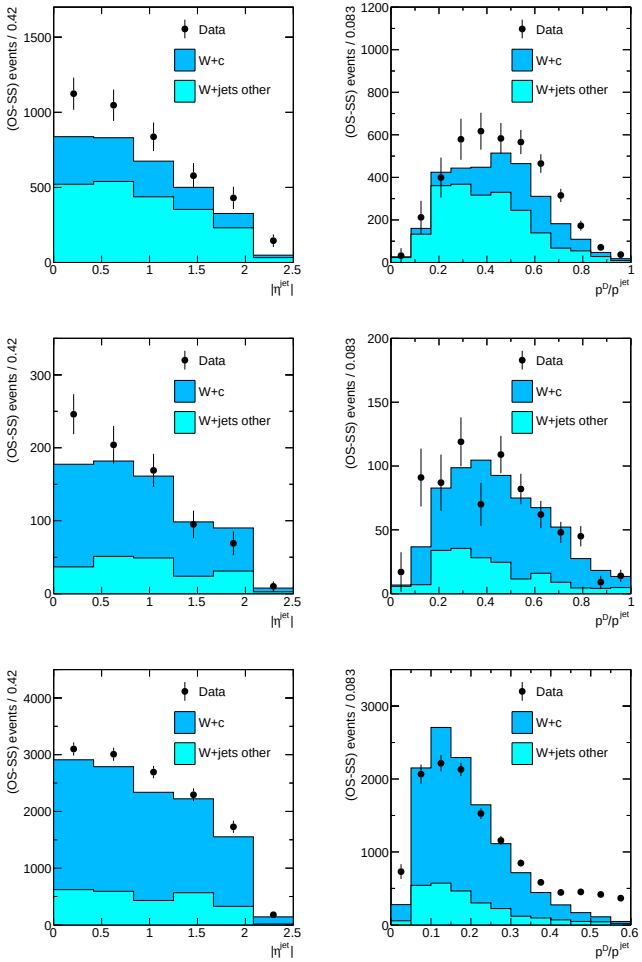
TOP-12-002



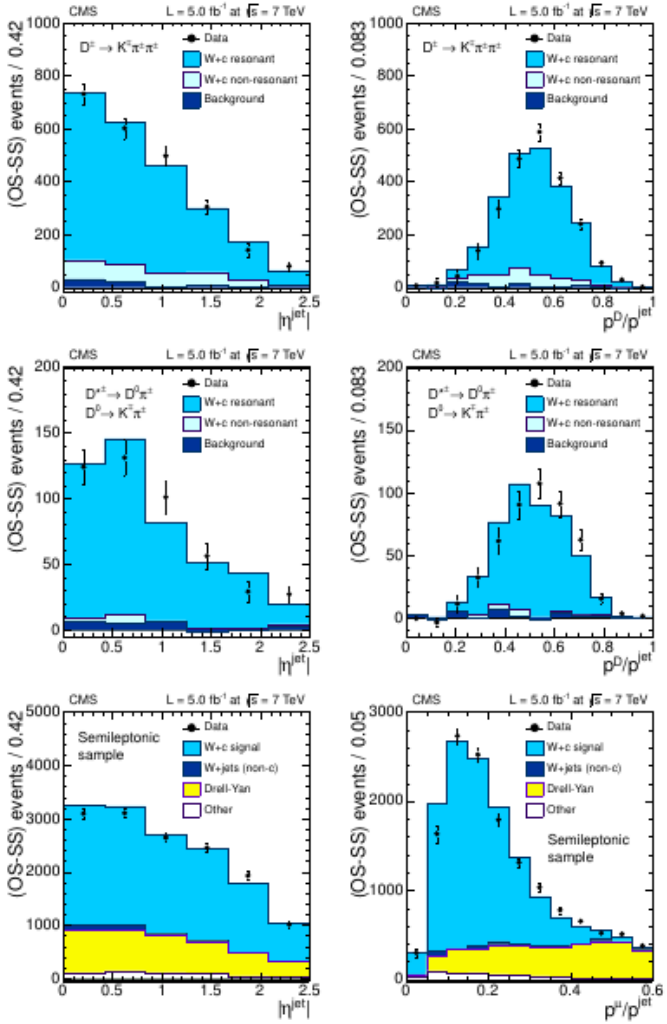
- Fitted to Gaus (signal) + 1st order polynomial (background)
- More signal events in open data
- Worse signal / background ratio
- *Drawback in open data vertexing: usage of Adaptive Vertex Fitter with track weighting*

Control distributions

Open Data



TOP-12-002



- Shapes very similar for D mesons
- MC $<$ data for D^+, μ: missing background samples
- Prevents from measuring cross section in semileptonic channel

Cross section measured at parton level:

$p_T^c > 25$ GeV, $|\eta^c| < 2.5|$, $p_T^\mu > 25$ GeV, $|\eta^\mu| < 2.1|$:

$$\sigma = \frac{N_{Sig}}{ALB},\ E = \frac{N_{MCreco}}{N_{MCgen}}$$

$$L = 2.5\ \text{fb}^{-1}$$

$$D^*: \quad B = f(c \to D^* \to D^0(K\pi)\pi) = 0.622\%$$
$$D^+: \quad B = f(c \to D^+ \to K\pi)\pi) = 2.08\%$$

N_{Sig} extracted from fit of invariant mass spectrum (differs from the paper)

Open data

D^+: 142.9 $\pm$ 20.4 (stat) pb

D^*: 139.2 $\pm$ 14.5 (stat) pb

TOP-12-002

D^+: 103.6 $\pm$ 7.8 (stat) + 8.1 (syst) pb

D^*: 116.9 $\pm$ 8.7 (stat) + 10.0 (syst) pb

$\Rightarrow$ **reasonable consistency with the paper**

Summary

- **Successfull validation of 2011 open data by re-doing two CMS measurements**
- Only similar event selection/reconstruction, no sophisticated corrections
- ... and sometimes even completely different reconstruction methods
- All key features reproduced, cross sections in reasonable agreement with published values
- Results have been (or will be) shown and discussed in physics groups: received interest from ongoing analyses
- **Planning to document $t\bar{t}$ measurement as analysis/validation example**