xFitter performance: CMS report and wishes

Oleksandr Zenaiev (DESY)
on behalf of the CMS Collaboration

xFitter External Meeting, Oxford
20.03.2017

New recent CMS results for PDF studies

- "Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functions"
  [arXiv:1703.01630, submitted to EPJ C] *using xFitter!*
  similar measurement at 13 TeV arXiv:1610.04191

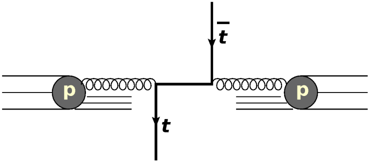
- "Measurement of Triple-Differential Dijet Cross Sections at $\sqrt{s} = 8$ TeV with the CMS Detector and Constraints on Parton Distribution Functions"
  [CMS-PAS-SMP-16-011] *using xFitter!*

- "Measurement and QCD analysis of double-differential inclusive jet cross-sections in pp collisions at $\sqrt{s} = 8$ TeV and ratios to 2.76 and 7 TeV"
  [arXiv:1609.05331, submitted to JHEP] *using xFitter!*

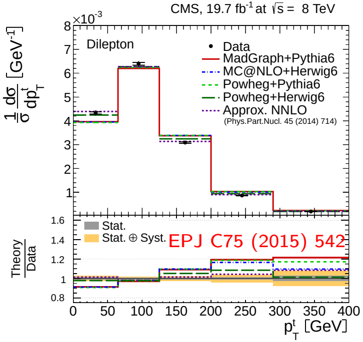
- "Determination of the strong coupling constant from the measurement of inclusive multijet event cross sections in pp collisions at $\sqrt{s} = 8$ TeV"
  [CMS-PAS-SMP-16-008]

- "Measurement of associated Z + charm production in pp collisions at $\sqrt{s} = 8$ TeV"
  [CMS-PAS-SMP-15-009]

Why measure $t\bar{t}$ production?

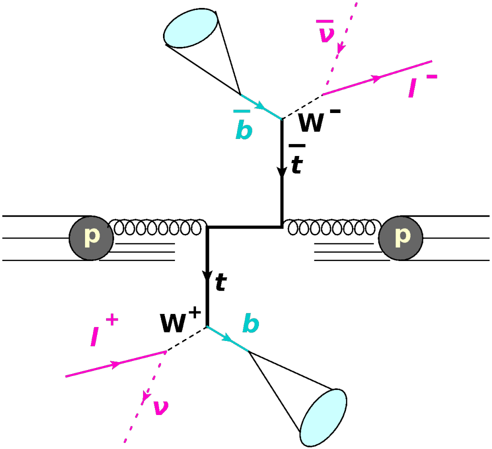


- m_t provides a hard scale
  $\Rightarrow$ ultimate probe of pQCD
  (NLO, aNNLO, NNLO, ...)
- Produced mainly via gg
  $\Rightarrow$ constrain gluon PDF
- **Inclusive and 1D $t\bar{t}$ already in PDF fits**
- Production sensitive to α_s and m_t
- May provide insight into possible new physics



Why measure 2D?

- Previous 1D measurements revealed some trends
- **2D measurement: especially better PDF sensitivity**

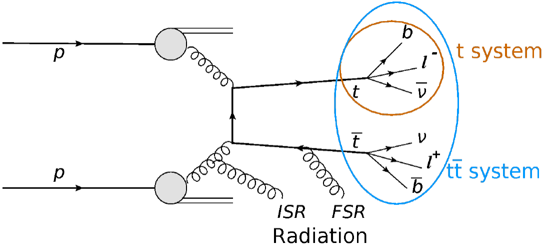


Measurement in $e\mu$ channel:

- **Leptons:**
  - at least 2 oppositely signed e, μ
  - $p_T > 20$ GeV
  - $|\eta| < 2.4$
- **Jets:**
  - at least 2
  - $p_T > 30$ GeV
  - $|\eta| < 2.4$
  - at least 1 b-tagged

(similar to previous 1D measurement)

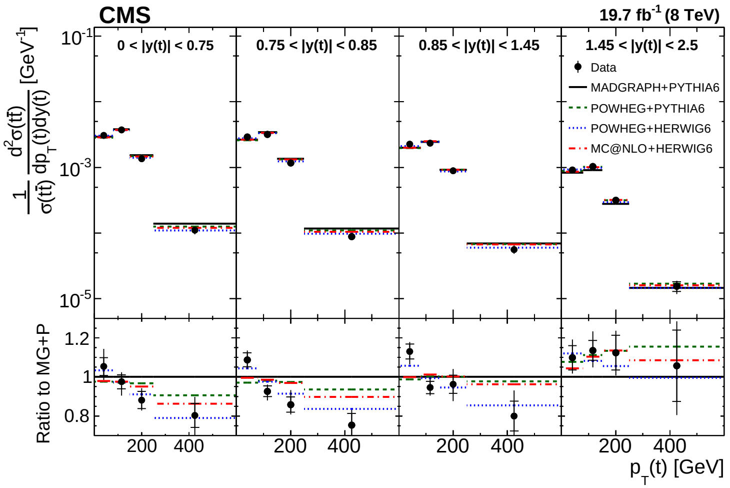
Kinematic reconstructions, 2D unfolding $\Rightarrow$ cross sections measured **at parton level**



- **t production:**
  - $y(t)$-$p_T(t)$: most simple, aNNLO publicly available

- **$t\bar{t}$ production:**
  - $M(t\bar{t})$-$y(t\bar{t})$: **most sensitive to PDFs (at LO $x_{1,2} = \sqrt{\frac{M(t\bar{t})}{s}} e^{\pm y(t\bar{t})}$)**
  - $M(t\bar{t})$-$p_T(t\bar{t})$: sensitive to radiation (at LO $p_T(t\bar{t}) \equiv 0$)

- **t, $t\bar{t}$ mixed:**
  - $M(t\bar{t})$-$y(t)$: sensitive to PDFs (at LO $y(t\bar{t}) = (y(t) + y(\bar{t}))/2$)
  - $M(t\bar{t})$-$\Delta\phi(t,\bar{t})$: sensitive to radiation (at LO $\Delta\phi(t\bar{t}) \equiv \pi$)
  - $M(t\bar{t})$-$\Delta\eta(t,\bar{t})$: correlated with $p_T(t)$ as well as sensitive to radiation

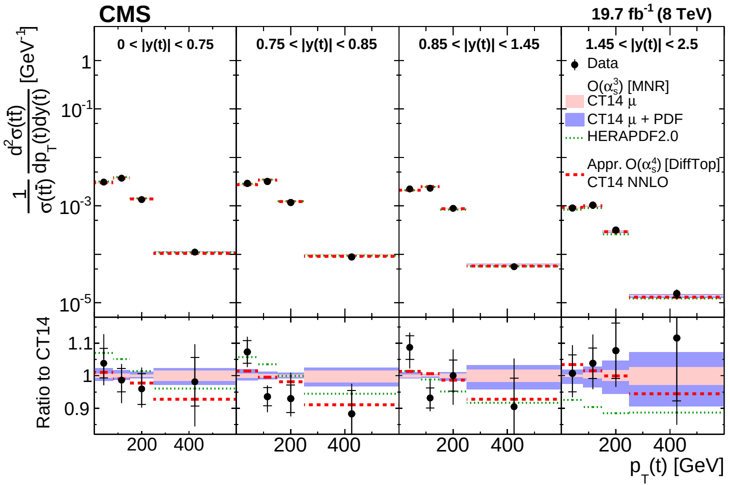


- Harder $p_T(t)$ distribution in all MC
- This trend is in wide $y(t)$ range

	MP	PP	PH	MH
χ^2	96	58	14	46

(dof = 15)

calculated using xFitter!

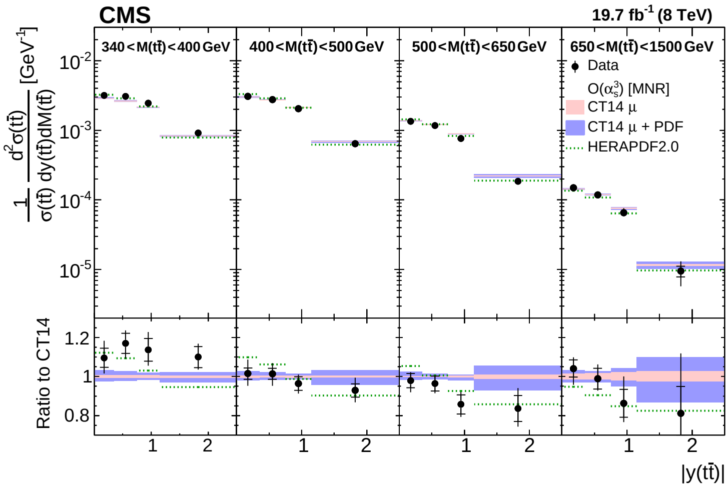


- aNNLO provide better data description
- Substantial PDF sensitivity
- Moderate scale uncertainties at NLO (normalised distribution)

	HERA2	CT14	CT14 NNLO
χ^2	46	24	13

(dof = 15)

calculated using xFitter!



	HERA2	CT14
χ^2	29	16

(dof = 15)

calculated using xFitter!

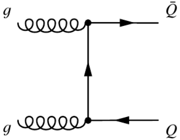
- PDF sensitivity exceeds scale uncertainties
- ⇒ can use these data for PDF fits

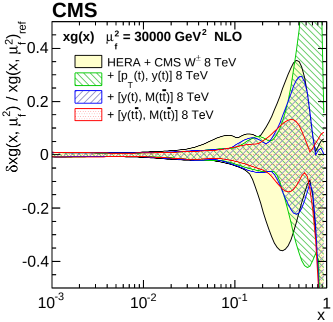
QCD factorisation:

$$\sigma_X = \sum_{a,b} \int \mathrm{d}x_1 \mathrm{d}x_2 \underline{f_a(x_1,\mu_f^2) f_b(x_2,\mu_f^2)} \hat{\sigma}_{ab\to X}(x_1,x_2,\mu_f^2,\ldots)$$

$$\mu_f \sim m_t$$

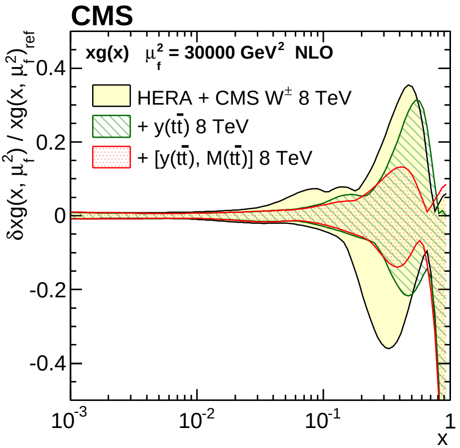
$$x_{1,2} = \sqrt{\frac{M(t\bar{t})}{s}} \mathrm{e}^{\pm y(t\bar{t})} \quad \Rightarrow \quad 0.01 \lesssim x_{1,2} \lesssim 0.25$$





- Using `xFitter` (former `HERAFitter`): open-source QCD fit framework [EPJ C75 (2015) 394, `xfitter.org`]
- Input data:
  - HERA inclusive DIS data [EPJ C75 (2015) 580]
  - CMS W asymmetry [EPJ C76 (2016) 469]
  - these 2D $t\bar{t}$
- NLO calculations for $t\bar{t}$ [NPB373 (1992) 295] using `MCFM` $\oplus$ `ApplGrid`
- Significant improvement of g at high x
- Best improvement comes from $M(t\bar{t})$-$y(t\bar{t})$

- Repeated analysis for 1D distribution $p_T(\mathrm{t})$, $y(\mathrm{t})$, $M(t\bar{t})$, $y(t\bar{t})$
- Checked their impact on PDFs



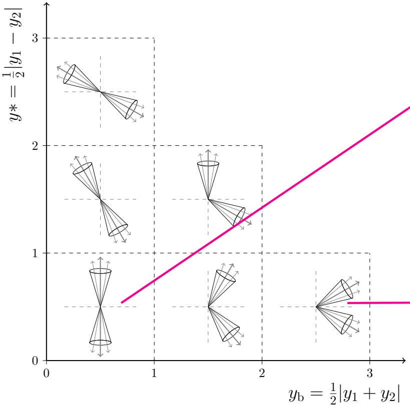
- **2D impact exceeds 1D**
- First study of such kind
- Strongly suggests to use these data in global PDF fits

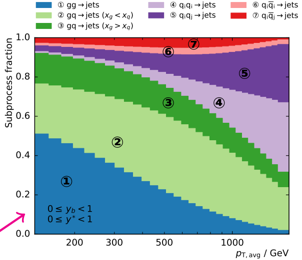
Triple-differential dijets [CMS-PAS-SMP-16-011]

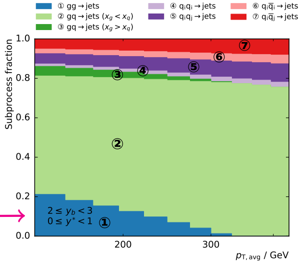
Measurement at 8 TeV, $L = 19.7\ \text{fb}^{-1}$, anti-k_T $R = 0.7$

3D cross sections

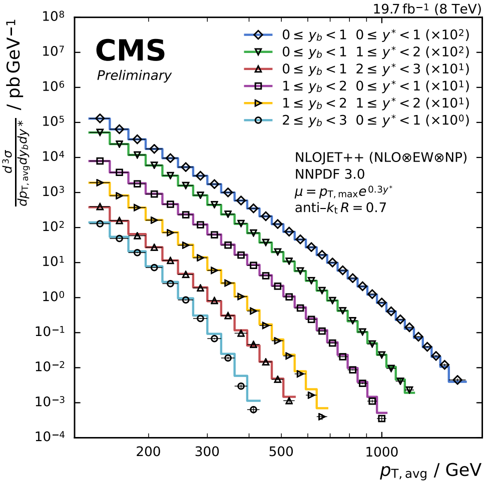
- $p_{\mathrm{T,avg}} = (p_{T,1} + p_{T,2})/2$,
- rapidity separation $y^* = \frac{1}{2}|y_1 - y_2|$
- dijet boost $y_b = \frac{1}{2}|y_1 + y_2|$







$> 80\%$ with at least one g at large y_b



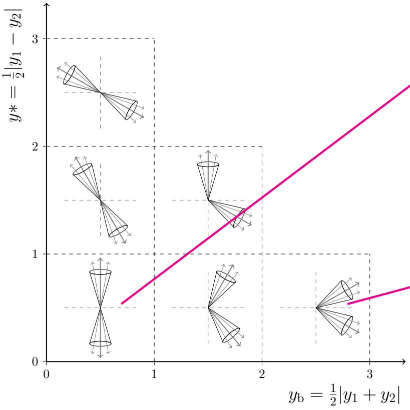
Data compatible with theory over a wide range of phase space

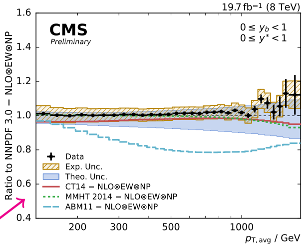
Triple-differential dijets [CMS-PAS-SMP-16-011]

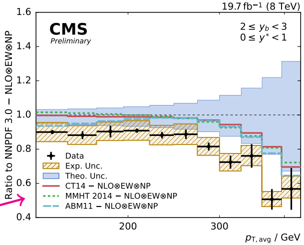
Data uncertainties:

- jet energy scale (2.5–12%)
- statistical uncertainties $\sim 1\%$, up to 20% at high $p_{\mathrm{T,avg}}$

Theory uncertainties larger than experimental

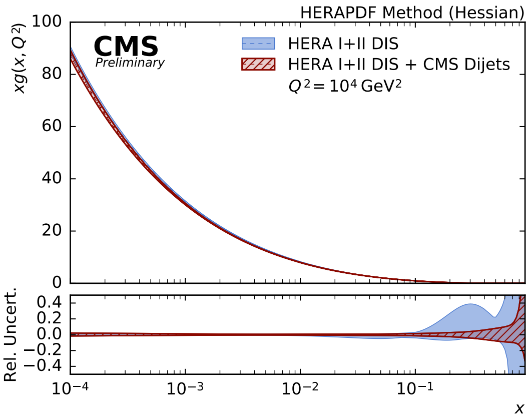






Data are well described in most of the phase space, but some differences at high $p_{\mathrm{T,avg}}$, y_b

$\Rightarrow$ **data can constrain PDFs**

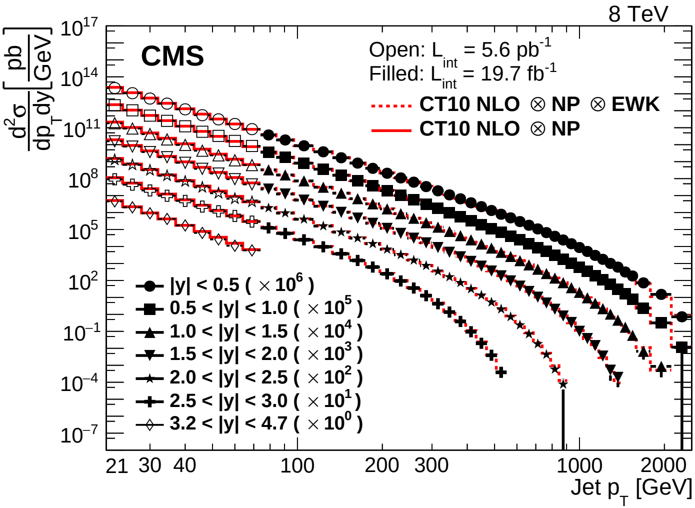


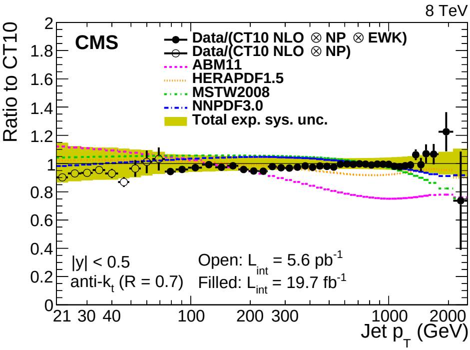
- **3D cross sections** ultimately probe Q^2, x_1, x_2 in PDFs
- PDF fit using HERA DIS + these 3D dijet data
- **strong improvement of gluon distribution, precise α_s extraction:**

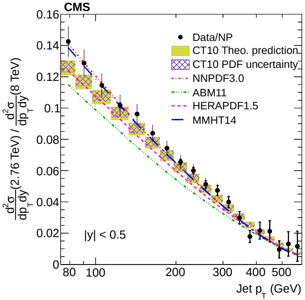
$$\alpha_S(M_Z) = 0.1199 \pm 0.0015\,(\mathrm{exp}) \pm 0.0002\,(\mathrm{mod})\,^{+0.0002}_{-0.0004}\,(\mathrm{par})\,^{+0.0031}_{-0.0019}\,(\mathrm{scale})$$

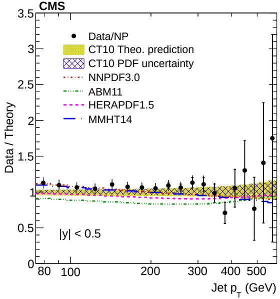
Inclusive jets [arXiv:1609.05331]

- Measurement at 8 TeV, anti-k_T $R = 0.7$:
  - $74 < p_T < 2500$ GeV ($L = 19.7$ fb^{-1})
  - $21 < p_T < 74$ GeV ($L = 5.6$ pb^{-1}) **NEW!**
- **2D cross sections** as function of p_T and y

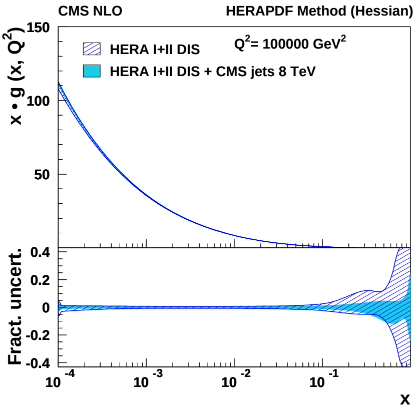








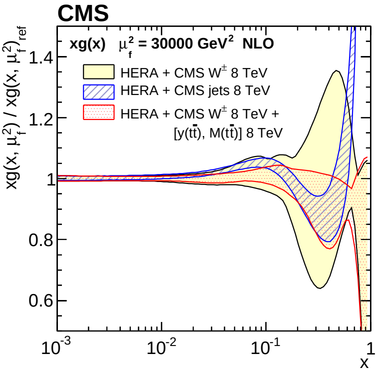
- Data uncertainties: jet energy scale (1–45% depending on y), lumi (2.6%)
- NLO uncertainties: scales (5–40%), PDFs (10–100%)
- Data well distinguish between PDFs
- **Ratios 2.76/8, 7/8 available: partial reduction of uncertainties**



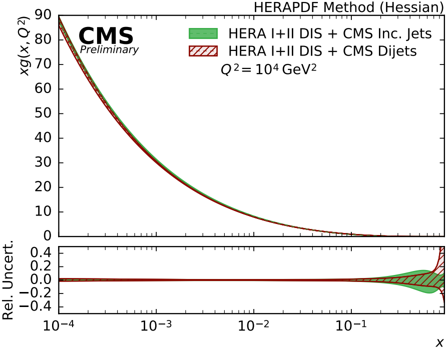
- Data: combined HERA DIS + these jet data
- Theory (for all jet results): NLOJET++ interfaced to fastNLO + EWK, if available
- **Strong improvement of gluon distribution, precise α_s extraction:**

$$\alpha_S(M_Z) = 0.1185^{+0.0019}_{-0.0021}\,(\text{exp})^{+0.0002}_{-0.0015}\,(\text{model})^{+0.0000}_{-0.0004}(\text{param})^{+0.0022}_{-0.0018}\,(\text{scale})$$

arXiv:1703.01630



CMS-PAS-SMP-16-011



Consistent and competitive PDF constraints using different CMS data!

Complementary sensitivity to α_s and m_t

$\Rightarrow$ one should use all the data sets to better test/constrain different QCD aspects

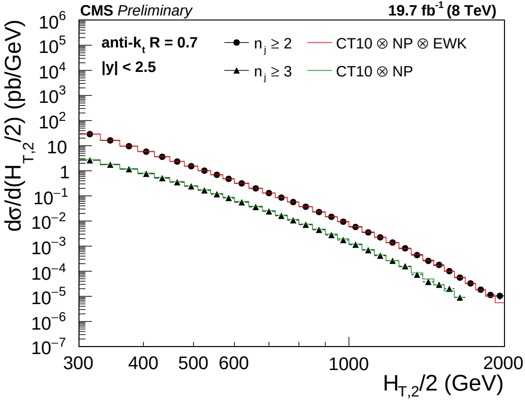
Many new results from CMS using xFitter:

- **2D $t\bar{t}$ [arXiv:1703.01630]:**
  - first PDF fit of 2D $t\bar{t}$: demonstrated better impact on PDFs w.r.t 1D
  - results competitive to those from jets
  - WISH: fully-differential NNLO $\Rightarrow$ would allow to study PDFs, α_s, m_t by fitting e.g. total + 2(3..)D normalised $t\bar{t}$ x-sections
- **3D dijets [CMS-PAS-SMP-16-011]:**
  - ultimate probe PDFs
  - improved gluon distribtuion, accurate α_s determination
  - WISH: NNLO needed to match data precision
- **Inclusive jets [arXiv:1609.05331]:**
  - improved gluon distribtuion, accurate α_s determination
  - ratios $7/8$, $2.76/8$ available: partial cancelation of theory and exp. unc.
  - WISH: NNLO needed to match data precision

For many of these measurements data are typically more precise than theory: need higher order calculations to fully reveal data precision

BACKUP

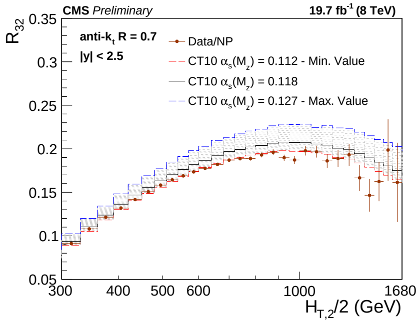
- Measurement at 8 TeV, $L = 19.7$ fb^{-1}, anti-k_T $R = 0.7$
- 2-jet and 3-jet event cross sections as function of average transverse momentum $H_{\mathrm{T},2}/2 = \frac{1}{2}(p_{\mathrm{T},1} + p_{\mathrm{T},2})$ of two leading jets

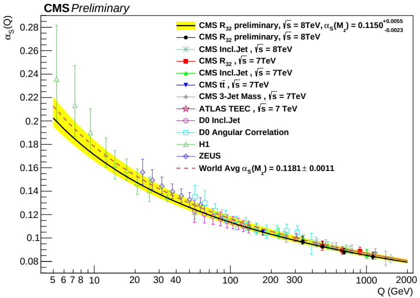


Data are described by theory predictions

3-jet to 2-jet cross section ratio R_{32}:

- many uncertainties cancel
- sensitive α_s
- $\Rightarrow \alpha_s$ extracted, also in ranges of $H_{T,2}/2$



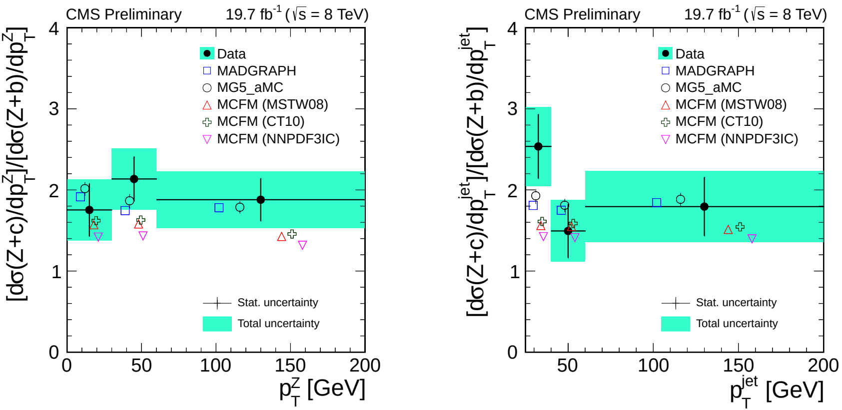


Precise α_s extraction, consistent with other CMS results:

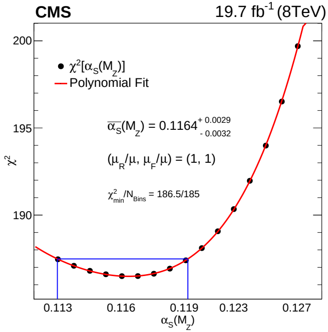
$$\alpha_s(M_Z) = 0.1150 \pm 0.0010\,(\text{exp}) \pm 0.0013\,(\text{PDF}) \pm 0.0015\,(\text{NP}) \,^{+0.0050}_{-0.0000}\,(\text{scale})$$

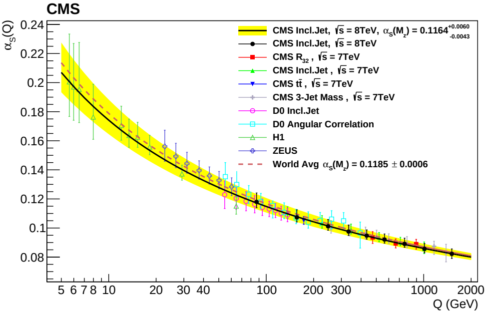
BACKUP. Measurement of associated Z + charm production [CMS-PAS-SMP-15-009]

- Measurement at 8 TeV, $L = 19.7$ fb^{-1}
- Cross section of $Z + c$ and ratio $Z + c/Z + b$ as function of p_T
- **Important for searches beyond SM, sensitive to possible intrinsic charm**



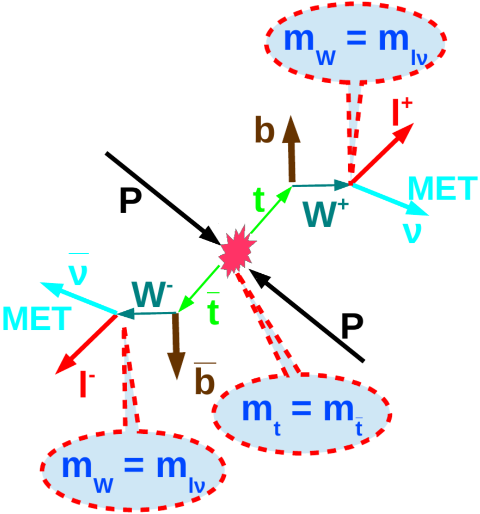
Data are compared to NLO calculations with different PDFs (including those with IC)





$$\alpha_S(M_Z)(\text{NLO}) = 0.1164^{+0.0025}_{-0.0029}(\text{PDF})^{+0.0053}_{-0.0028}(\text{scale}) \pm 0.0001(\text{NP})^{+0.0014}_{-0.0015}(\text{exp}) = 0.1164^{+0.0060}_{-0.0043}$$

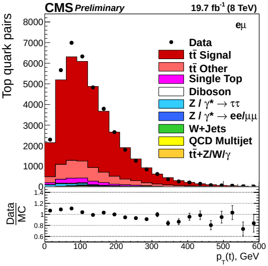
- Using CT10NLO PDFs
- Result consistent with other determinations
- Performed in separate p_T bins: determine running $\alpha = \alpha(Q)$

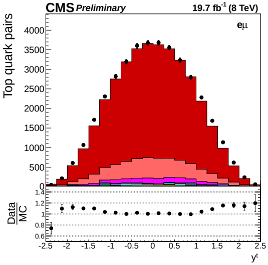


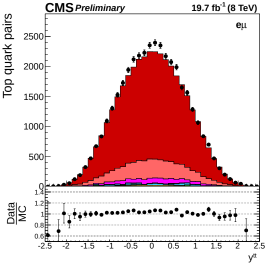
- Measured input: leptons, jets, MET
- Unknowns: $\bar{p}_\nu$, $\bar{p}_{\bar{\nu}}$ (6)
- Constraints:
  - m_t, $m_{\bar{t}}$ (2)
  - m_{W+}, m_{W-} (2)
  - $(\bar{p}_\nu + \bar{p}_{\bar{\nu}})_T = \mathrm{MET}$ (2)
- Take weighted average of 100 reconstructions with inputs smeared by their resolution
- This helps to recover events with no solution because of detector effects

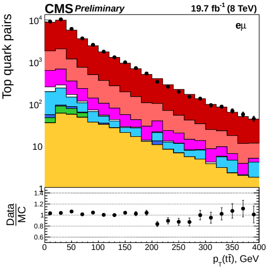
[Phys. Rev. D73 (2006) 054015]

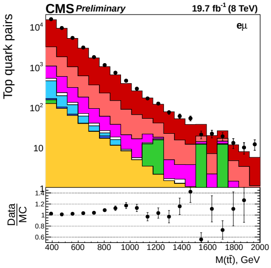
BACKUP. Kinematic distributions





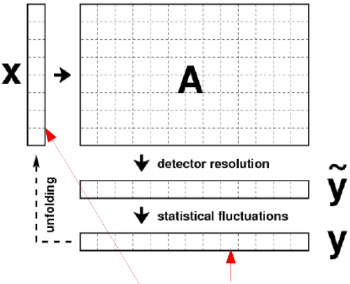






MadGraph + Pythia6 provides overall good description, but:

- $p_T(\mathrm{t})$ harder
- $y(\mathrm{t})$ more central
- $y(t\bar{t})$ less central



2d distributions are mapped to 1d arrays

`TUnfold` [JINST 7 (2012) T10003]

χ^2 minimisation with regularisation ($\approx 1\%$)

reco. data — unfolded distribution — regularization strength — regularization conditions (second derivative)

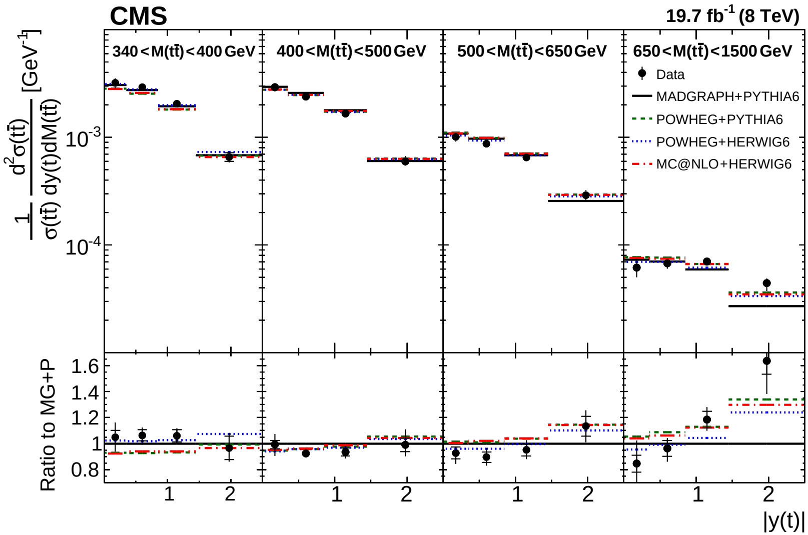
$$\chi^2=(Y-AX)^T V_Y^{-1}(Y-AX)+\tau^2(X-X_0)^T L^T L(X-X_0)$$

migration probability matrix — stat. errors of reco. — **gen.** distribution

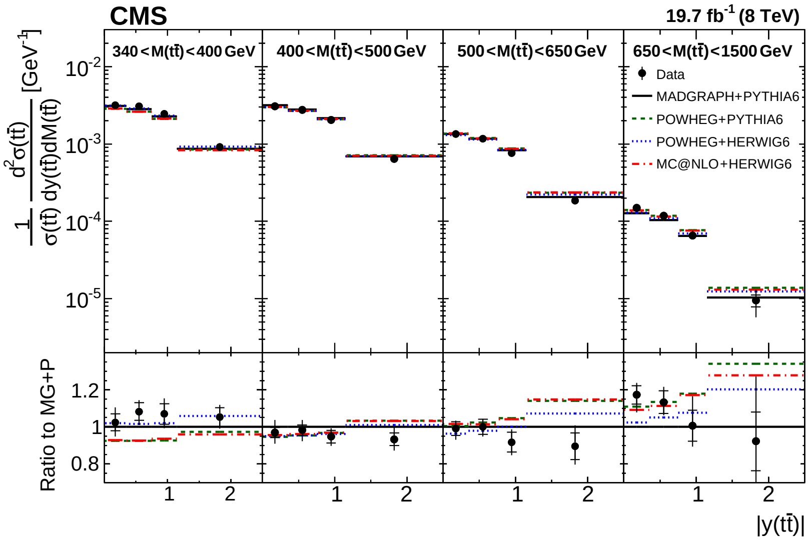
$$Y = N_{measured} - N_{Background}$$

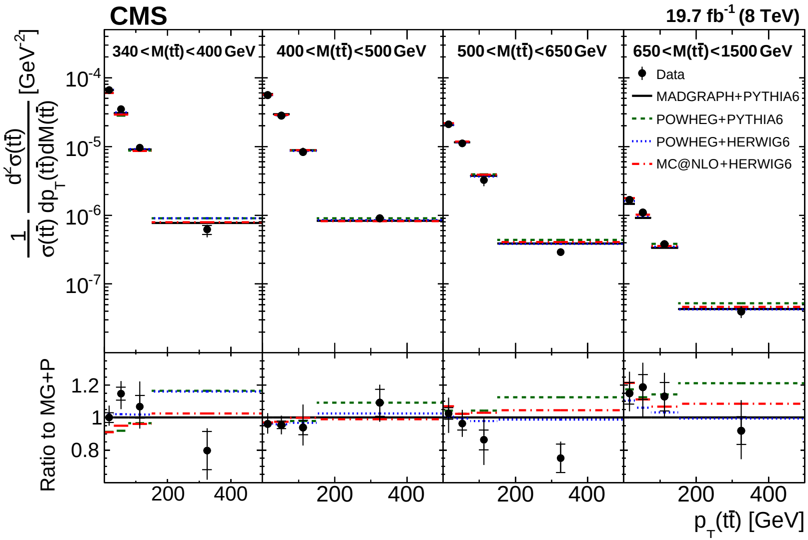
For each Δ**a**^{i}:

$$\left(\frac{1}{\sigma}\frac{d\sigma}{db}\right)^{ij}=\frac{1}{\sigma}\cdot\frac{X^{ij}}{BR\cdot L\cdot\Delta b^j}$$

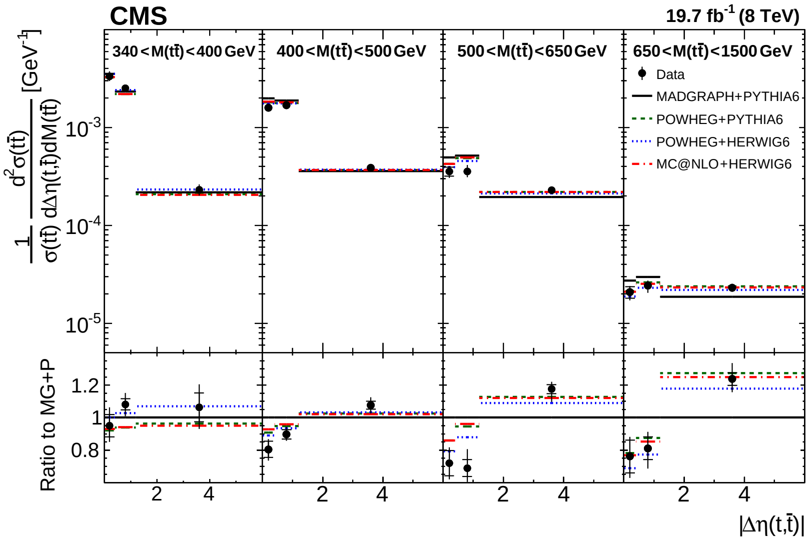

CMS
19.7 fb^{-1} (8 TeV)
340<M(tt̄)<400 GeV
400<M(tt̄)<500 GeV
500<M(tt̄)<650 GeV
650<M(tt̄)<1500 GeV
Data
MADGRAPH+PYTHIA6
POWHEG+PYTHIA6
POWHEG+HERWIG6
MC@NLO+HERWIG6
$\frac{1}{\sigma(t\bar{t})} \frac{d^2\sigma(t\bar{t})}{dy(t)dM(t\bar{t})}$ [GeV^{-1}]
10^{-3}
10^{-4}
Ratio to MG+P
1.6
1.4
1.2
1
0.8
1
2
|y(t)|

BACKUP. Comparison to MC

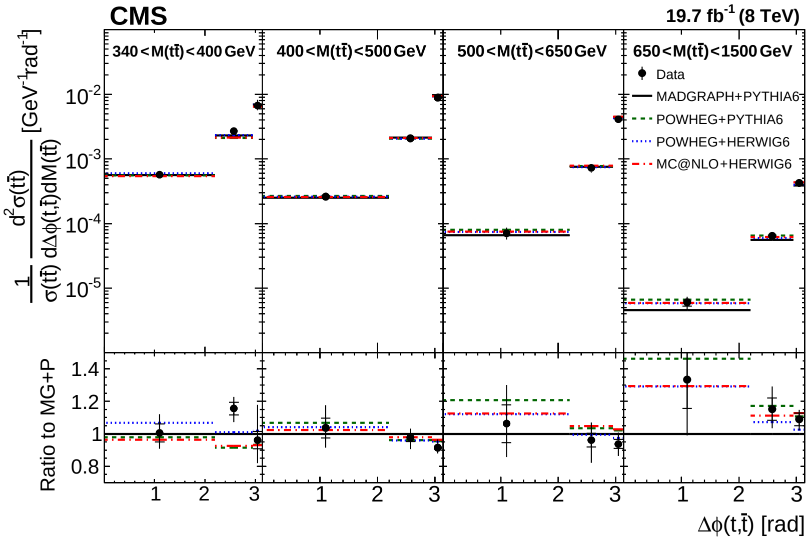


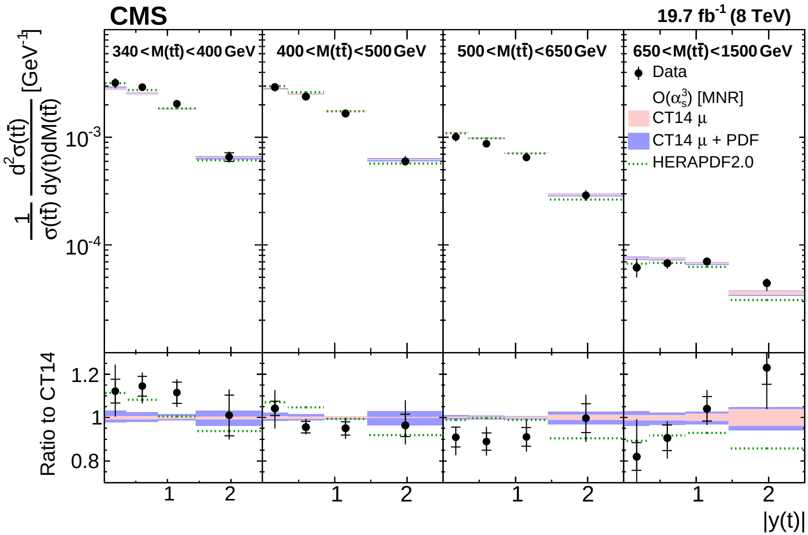

CMS
19.7 fb^{-1} (8 TeV)
340<M(tt̄)<400 GeV
400<M(tt̄)<500 GeV
500<M(tt̄)<650 GeV
650<M(tt̄)<1500 GeV
Data
MADGRAPH+PYTHIA6
POWHEG+PYTHIA6
POWHEG+HERWIG6
MC@NLO+HERWIG6
$\frac{1}{\sigma(t\bar{t})} \frac{d^2\sigma(t\bar{t})}{dp_T(t\bar{t})dM(t\bar{t})}$ [GeV^{-2}]
Ratio to MG+P
$p_T(t\bar{t})$ [GeV]

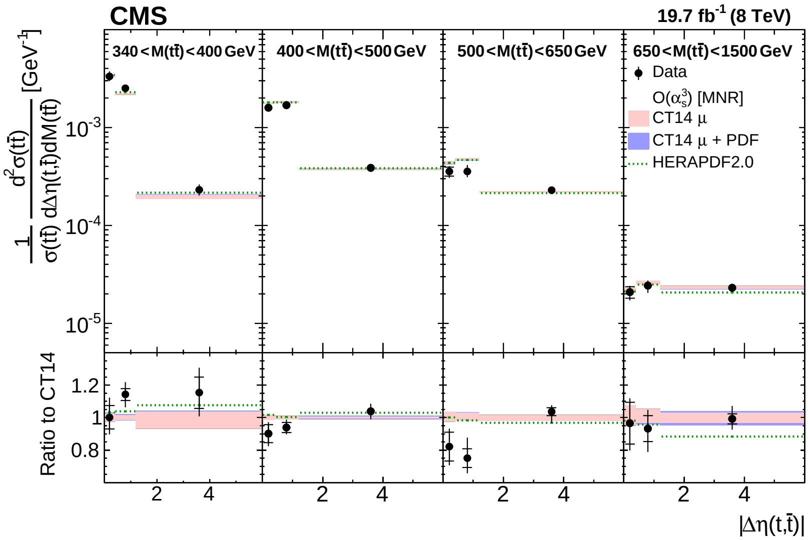
BACKUP. Comparison to MC



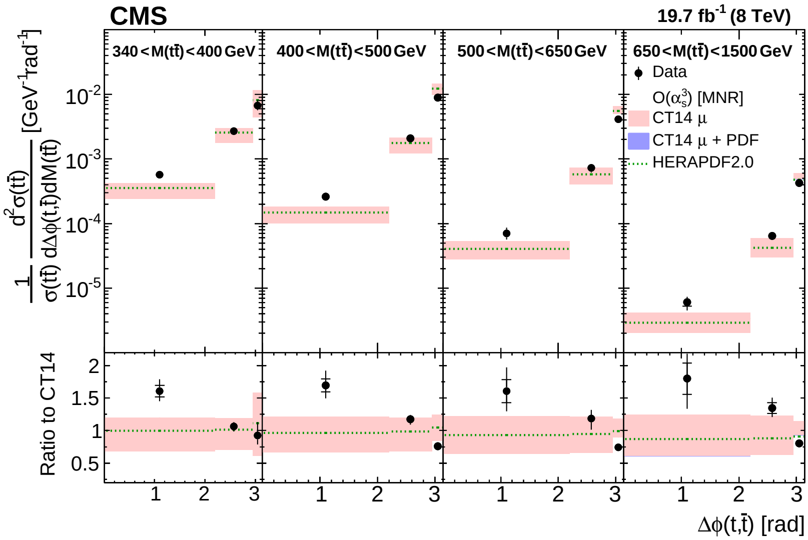
BACKUP. Comparison to MC

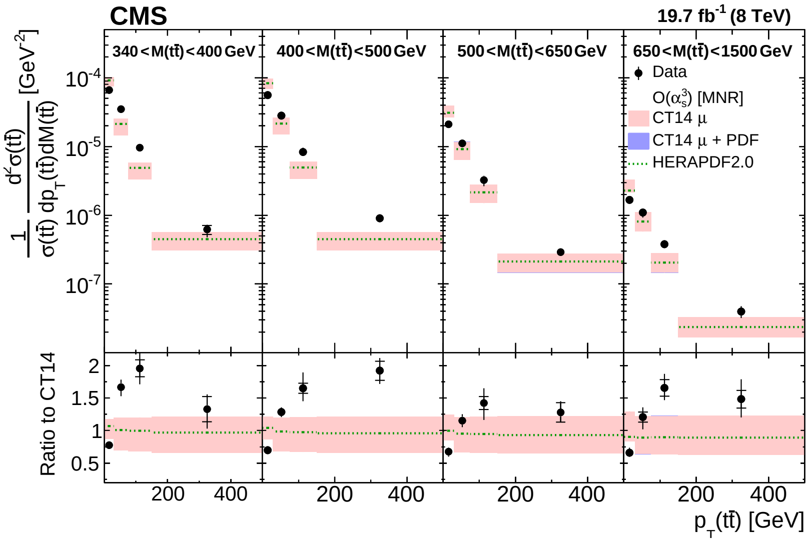



CMS
19.7 fb^{-1} (8 TeV)
$340 < M(t\bar{t}) < 400$ GeV
$400 < M(t\bar{t}) < 500$ GeV
$500 < M(t\bar{t}) < 650$ GeV
$650 < M(t\bar{t}) < 1500$ GeV
Data
$O(\alpha_s^3)$ [MNR]
CT14 μ
CT14 μ + PDF
HERAPDF2.0
$\frac{1}{\sigma(t\bar{t})} \frac{d^2\sigma(t\bar{t})}{dy(t)dM(t\bar{t})}$ [GeV^{-1}]
10^{-3}
10^{-4}
Ratio to CT14
1.2
1
0.8
1
2
|y(t)|


CMS
19.7 fb^{-1} (8 TeV)
340<M(tt̄)<400 GeV
400<M(tt̄)<500 GeV
500<M(tt̄)<650 GeV
650<M(tt̄)<1500 GeV
Data
O(α_s^3) [MNR]
CT14 μ
CT14 μ + PDF
HERAPDF2.0
$\frac{1}{\sigma(t\bar{t})} \frac{d^2\sigma(t\bar{t})}{d\Delta\eta(t,\bar{t})dM(t\bar{t})}$ [GeV^{-1}]
10^{-3}
10^{-4}
10^{-5}
Ratio to CT14
1.2
1
0.8
2
4
|Δη(t,t̄)|

BACKUP. Comparison to NLO




CMS
19.7 fb^{-1} (8 TeV)
340<M(tt̄)<400 GeV
400<M(tt̄)<500 GeV
500<M(tt̄)<650 GeV
650<M(tt̄)<1500 GeV
Data
O(α_s^3) [MNR]
CT14 μ
CT14 μ + PDF
HERAPDF2.0
$\frac{1}{\sigma(t\bar{t})} \frac{d^2\sigma(t\bar{t})}{dp_T(t\bar{t})dM(t\bar{t})}$ [GeV^{-2}]
Ratio to CT14
$p_T(t\bar{t})$ [GeV]

Following 1D cross section measurement (TOP-12-028):

- pile-up
- lepton selection
- trigger efficiency
- **jet energy scale and resolution ($\lesssim 2\%$)**
- b-tagging efficiency
- kinematic reconstruction efficiency
- background variation:
  - DY varied separately by $\pm 30\%$
  - other backgrounds varied simultaneously by $\pm 30\%$
- model uncertainties ($\lesssim 10\%$):
  - **perturbative scale variation**
  - **matching scale variation**
  - m_t variation
  - PDFs
  - **Hadronization (PowhegHerwig - PowhegPythia)**
  - **Hard scattering (PowhegPythia - MadgraphPythia)**

BACKUP. χ^2 comparison w.r.t NLO

Distribution	NDoF	MC				NLO nominal (including PDF unc.)						
		MP	PP	PH	MH	HERAPDF2	MMHT14	CT14	NNPDF30	ABM11nf5	JR14	CJ12
$y(t) - p_T(t)$	15	96	58	14	46	46 (40)	26 (24)	24 (21)	28 (25)	62 (51)	47 (47)	24 (23)
$M(t\bar{t}) - y(t)$	15	53	20	13	21	52 (44)	22 (20)	19 (18)	14 (14)	71 (55)	44 (44)	20 (20)
$M(t\bar{t}) - y(t\bar{t})$	15	19	21	15	22	29 (25)	15 (15)	16 (15)	10 (10)	42 (31)	25 (25)	15 (15)
$M(t\bar{t}) - \Delta\eta(t\bar{t})$	11	163	33	20	39	46 (43)	31 (31)	32 (31)	45 (42)	48 (44)	39 (39)	32 (32)
$M(t\bar{t}) - p_T(t\bar{t})$	15	31	83	30	33	485 (429)	377 (310)	379 (264)	251 (212)	553 (426)	428 (415)	382 (378)
$M(t\bar{t}) - \Delta\phi(t\bar{t})$	11	21	21	10	17	354 (336)	293 (272)	296 (259)	148 (143)	386 (335)	329 (324)	297 (295)

Monte Carlo

- MP: MadGraph(CTEQ6) + Pythia6
- PP: Powheg(CT10) + Pythia6
- PH: Powheg(CT10) + Herwig6
- MH: MC@NLO(CTEQ6) + Herwig6

Overall description by MCs:

- best: Powheg + Herwig
- worst: MadGraph + Pythia

NLO

Some comments on PDFs:

- MMHT14, CT14 and NNPDF30: use LHC data
- other PDFs: no LHC data

Overall description by different PDFs:

- best: MMHT14, CT14, NNPDF30, CJ12
- worse: HERAPDF2.0, JR14, ABM11nf5

Particular distributions described by "best" PDFs:

- bad description of $\Delta\phi(t\bar{t})$ and $p_T(t\bar{t})$
- also bad description of $\Delta\eta(t\bar{t})$
- reasonable description of the rest

BACKUP. Framework of QCD analysis

In short: similar to HERAPDF2.0 fit with $\mathrm{t\bar{t}}$ data included

(any publicly reproducable PDF fit would serve the purpose)

- Platform: xFitter (former HERAFitter) [www.xfitter.org]
- Input data: HERA $e^{\pm}p$ inclusive data [1506.06042], $Q^2_{\mathrm{min}} > 3.5$ GeV^2 + CMS W asym. 8TeV [1603.01803] + $t\bar{t}$ normalised 2D data
- RT optimal variable-flavour-number scheme with $n_f = 5$, $\alpha_s^{n_f=5}(M_Z) = 0.118$, $M_c = 1.47$ GeV, $M_b = 4.50$ GeV
- Predictions for $t\bar{t}$:
  - MNR calculations [NPB373 (1992) 295] via MCFM $\otimes$ ApplGrid $\otimes$ xFitter
  - pole top mass $m_t = 172.5$ GeV
  - scales $\mu_r^2 = \mu_f^2 = m_t^2 + (p_T^2(t) + p_T^2(\bar{t}))/2$
- PDF parametrisation: 18 free parameters HERAPDF style
- Uncertainties:
  - Experimental: from $\Delta\chi^2 = 1$
  - Model: from theoretical and model parameter variations
  - Parametrisation: from μ^2_{f0} and parameterisation form variation

(pictures taken from arXiv:1511.00549)

Is it OK to use NLO calculations?

Yes, for normalised distributions $\mathbf{p_T(t), y(t), M(t\bar{t}), y(t\bar{t})}$ with current data precision, because:

- These are the only publicly available calculations
- Missing higher order corrections (estimated by scale variation) for NLO are small w.r.t to data:
  - $\simeq 10\%$ fot total x-section
  - $\lesssim 3\%$ for $p_T(t)$, $y(t)$, $M(t\bar{t})$, $y(t\bar{t})$ shapes
  - **Confirmed by exact NNLO calculations** [Czakon et al. arXiv:1511.00549] (slightly larger effect on $p_T(t)$).