Global fits of PDFs and SM parameters: A modern perspective with xFitter

Sasha Zenaiev
(Hamburg University)

https://gitlab.cern.ch/fitters/xfitter

Mainz Institute for Theoretical Physics
25 Oct 2024

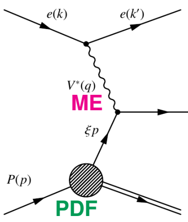
Overview

- I will discuss how we do global QCD fits: why, methodology, what issues may arise etc.
- Disclaimer: based on my limited experience with global QCD fits (mainly using xFitter):
  - fits to *ep* HERA data
    H1 and ZEUS Coll., EPJ C75 (2015) 580 EPJ C78 (2018) 473
  - fits to HERA+LHCb+ALICE data and astroparticle application
    PROSA Coll., EPJ C75 (2015) 396 JHEP 04 (2020) 118 EPJ C80 (2020) 658
  - global fits in the ABMP PDF framework
    Garzelli, Mazzitelli, Moch, Zenaiev JHEP 05 (2024) 321,
    Alekhin,Garzelli, Moch, Zenaiev arXiv:2407.00545
  - fits within xFitter developers' team
    Anataichuk et al. arXiv:2310.19638 etc.
- Most of these were done using open source `xFitter` (former `HERAfitter`) program
  - define theory model with some free parameters
  - select experimental data (uncertainties+correlations are crucial here)
  - compare theory to data and extract best theory parameters

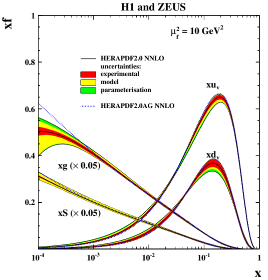
xFitter

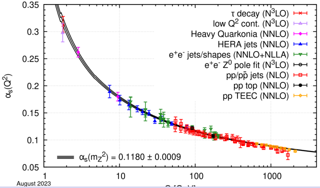
Gitlab page

Global QCD anlyses: PDFs,α_S,... $\xleftarrow{\text{pQCD}}$ $d\sigma/dO$ $\xleftarrow{\text{MC}}$ events



- Factorization theorem: $\boldsymbol{\sigma} = \textbf{PDF} \otimes \textbf{ME}$
- Parton distribution functions (PDFs) $\boldsymbol{f(x, \mu_f)}$ describe distribution of quarks and gluons in hadrons
- Matrix elements (ME) are calculated in perturbative QCD (pQCD) $\sigma = \Sigma_{i=0}^{n} \sigma_i \alpha_S^i$ requiring $\alpha_S(\mu_r) < 1$ ($\mu_r \gg \Lambda_{QCD}$)
- At low scales $\mu \sim 1$ GeV non-perturbative QCD effects are parametrised by PDFs **which are extracted using data**
  - typically shaped like $x^a(1-x)^b$ with a few tens of parameters (but there are different approaches e.g. NNPDF)
- At higher scales $\mu > 1$ GeV PDF evolution is predicted by pQCD
- Other unknown parameters: $\alpha_S(M_Z)$, masses of heavy quarks (fundamental free parameters of Standard Model)
  - can be fitted or fixed in global QCD analyses
- **Challenges:**
  - choose suitable PDF parametrization
  - select **PDF sensitive** and **consistent** data sets
  - use appropriate statistical method (typically minimizing χ^2)
  - **most challenging are PDF uncertainties:** very much depend on all above



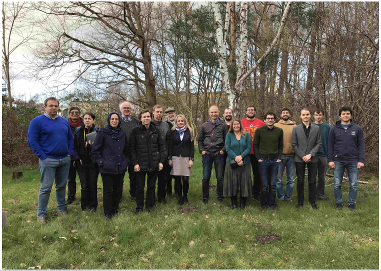


xFitter [https://gitlab.com/fitters/xfitter]

Welcome to xFitter (former HERAFitter)

Proton parton distribution functions (PDFs) are essential for precision physics at the LHC and other hadron colliders. The determination of the PDFs is a complex endeavor involving several physics processes. The main process is lepton-proton deep-inelastic scattering (DIS), with data collected by the HERA ep collider, covering a large kinematic phase space needed to extract PDFs. Further processes (fixed-target DIS, pp-bar collisions, etc.) provide additional constraining powers for flavor separation. In particular, the precise measurements obtained from or yet to come from the LHC will continue to improve the knowledge of the PDF.

The xFitter project is an open-source QCD fit framework designed to extract PDFs and assess the impact of new data. The framework includes modules allowing for various theoretical and methodological options, capable of fitting a large number of relevant datasets from HERA, Tevatron, and LHC. This framework is already used in many analyses at the LHC.

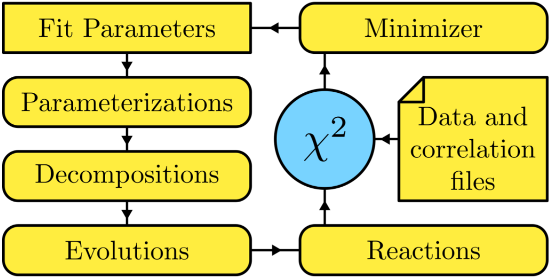


Downloads of the xFitter software package

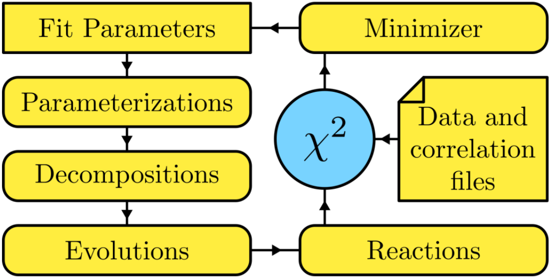
(!) **xFitter-2.0.1 release** is publicly available.
All the xFitter releases can be accessed HERE.
Description: arXiv:1410.4412

Overview of xFitter [gitlab.com/fitters/xfitter]

I. Novikov



- **xFitter (HERAfitter before 2015) is a unique open-source QCD fit framework:**
  - extract PDFs and theory parameters
  - assess impact of new experimental (pseudo-)data and check consistency
  - test different theoretical assumptions
  - . . . any exercise which involves data vs. theory
- It is widely used by LHC experiments and theorists ($>$ 100 publications)

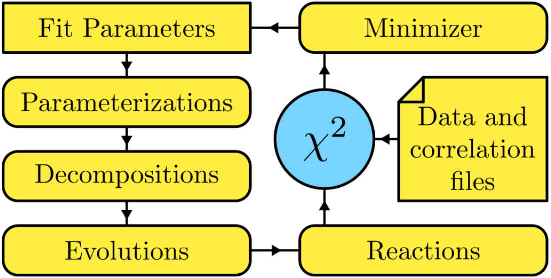


Why is xFitter UNIQUE and so VERSATILE/FLEXIBLE/ADAPTABLE?
Because it is fully modular.

E.g., hadron interactions are realized as:

- PDF parametrisation at starting scale: it is enough to type your favourite formulas
- PDF decomposition: construct valence, sea and gluon, apply sum rules (automatic numerical integration is available)
- PDF evolution: interfaced various codes (QCDNUM, OPENQCDRAD, APFEL, LHAPDF, **APFEL++, HOPPET (new!)** for PDF evolution up to **approximte N^{3}LO** (TMD PDF evolution is also available in (inofficial) branch)
- hard scattering (“reaction”): very many processes are available

Flexibility of xFitter (2)



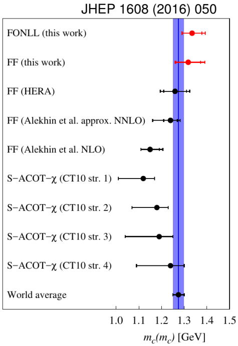
- hard scattering ("reaction"): very many processes are available
  - various heavy-quark schemes for *ep* DIS
  - some "simple" calculations, e.g. LO DY (used in this work)
  - interfaced external packages, e.g. HATHOR (NNLO total heavy-quark and single *t* hadroproduction) and HVQMNR (NLO heavy-quark differential hadroproduction)
  - but main emphasis is put on interfaces to **fast interpolation tables**, such as **fastNLO, ApplGrid, PineAppl**: allows us to get recent higher-order calculations (e.g. MCFM, MATRIX etc.) "for free"
- Flexible χ^2 implementation (treatment of experimental uncertainties)
- χ^2 minimisation: MINUIT, CERES
- **. . . and one can change/mix all these ingredients freely!**

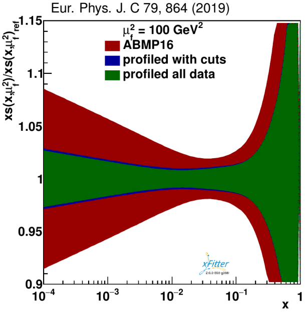
xFitter community

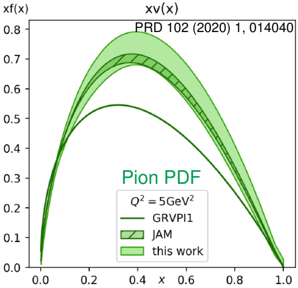
- $\sim$ 10 active developers (both experimentalists and theorists) + a few tens of users
- bi-weekly meetings via zoom + physical room at DESY (Wednesday 3pm)
- annual external meetings (spring):
  - 2020: DESY
  - 2022: Paris-Saclay University
  - 2023: CERN
  - 2025: Dubai (tentative $\sim$ May)
- CERN Gitlab `https://gitlab.cern.ch/fitters/xfitter`
  - read access for everyone (also for development branches)
  - needed CERN account to commit new code, or use mirror at `https://gitlab.com`
- DESY support: `naf-xfitter` machine + access to DESY NAF BIRD computing cluster (need DESY computing account)
- every winter/summer school at DESY $\sim$ 1 student (remote/on site) for xFitter
  - very successfull projects (e.g. A. Anataichuk et al., arXiv:2310.19638)
  - contact us if you have students willing to work on phenomenology topics matching xFitter scope
- mailing list: `xfitter-users@googlegroups.com`

Selected studies by the xFitter team

- "A determination of $m_c(m_c)$ from HERA data using a matched heavy flavor scheme" [JHEP 1608 (2016) 050]
- "Probing the strange content of the proton with charm production in charged current at LHeC" [Eur. Phys. J. C 79, 864 (2019)]
- "Parton Distribution Functions of the Charged Pion Within The xFitter Framework" [Phys.Rev.D 102 (2020) 1, 014040]







xFitter: practical aspects

- xFitter is interfaced to a number of libraries (QCDNUM, LHAPDF, APFEL, APFEL++, Hoppet, Hathor, APPLgrid, fastNLO, PineAppl . . .)
  - xFitter comes with installation script which installs everything from scratch
  - tested on several Linux distributions and macOS
- xFitter comes with built-in **tests** (examples)
  - serve for validation (CI on gitlab)
  - serve as examples how to use certain data and reproduce results from e.g. a certain paper: perfect starting point for new users
- Data are in separate repo: `https://gitlab.cern.ch/fitters/xfitter-datafiles`
  - contain necessary theory predictions (APPLgrid, fastNLO, PineAppl): a few GB
  - some recent data sets come with theory predictions on git LFS or Ploughshare (theory tables will be downloaded on request)
- Avaiable processes (either in master or development branches):
  - DIS: HERA [backbone of PDF fits] + fixed target (BCDMS, NMC, SLAC) [large x]
  - DY (Tevtron and LHC) [quarks PDFs]
  - $t\bar{t}$ abd single t (Tevtron and LHC) [gluon PDFs, m_t]
  - jets (LHC) [gluon PDFs, α_S]
    $\rightarrow$ **comprehensive set of data for a global PDF fit**
  - pseudo-data used for xFitter papers (LHeC, HL-LHC)
- Most of the data sets come with theory predictions at least at NLO
  - some (recent) come with NNLO predictions: jets, $t\bar{t}$
  - work-in-progress on N3LO theory predictions for DIS (APFEL++, Hoppet)

Basic ingedient for any data vs. theory fit: χ^2 expression

$$\chi^2_{\text{exp}}(\boldsymbol{m}, \boldsymbol{b}) = \sum_{ij} \left(m_i - \sum_\alpha \Gamma^i_\alpha b_\alpha - \mu_i \right) C^{-1}_{\text{stat},\ ij} \left(m_i - \sum_\alpha \Gamma^i_\alpha b_\alpha - \mu_i \right) + \sum_\alpha b^2_\alpha$$

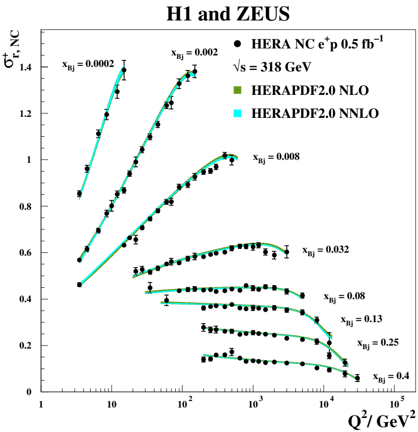
- m_i: data
- μ_i: theory
- $C_{\text{stat},\ ij}$: statistical covariance matrix
- b_α: nuisance parameters for correlated systematic uncertainties
- Γ_α: scaled correlated systematic uncertainties; might depend on m_i, μ_i:

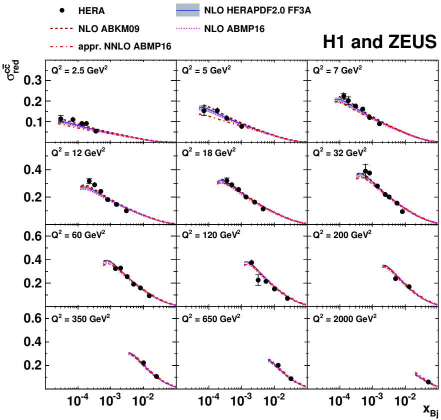
Treatment	Scaling rule (Γ^i_α)
Poisson	$\sqrt{m_i \mu_i}$
Multiplicative	m_i
Additive	μ_i

- Correlated uncertainties can be supplied as covariance matrix or source-by-source
- Also uncertainties can be included with offset method (external variations)

$\Rightarrow$ Need to know what are uncorrelated and correlated uncertainties, and how they scale (input from experiments which is not lways available)

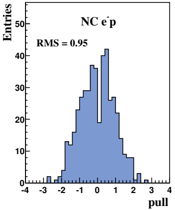
HERA DIS data

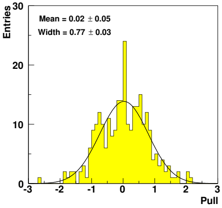




- HERA data on *ep* DIS scattering are a backbone of all global QCD analyses
  - direct constraints on valence and sea quark PDFs in a wide kinematic range
  - however only indirect sensitivity to gluon PDF and α_S
- HERA data on heavy quark (charm, bottom) and jet production in DIS:
  - direct constraints on gluon PDF, α_S, m_c, m_b

HERA DIS data: discussion which might be relevant for EIC

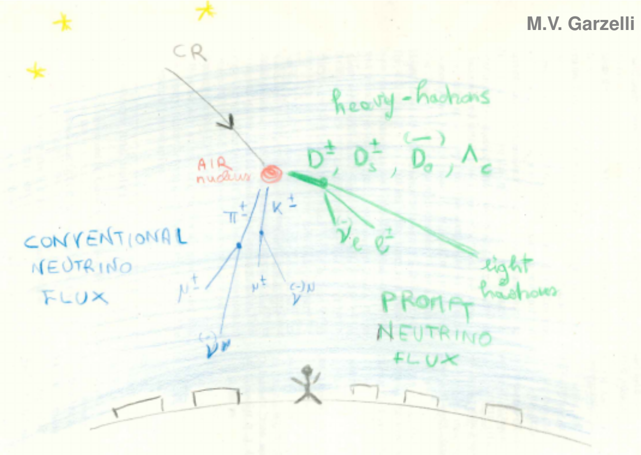




Q^2 GeV^2	x_{Bj}	$\sigma^+_{r,NC}$	δ_{stat} %	δ_{uncor} %	δ_{cor} %	δ_{rel} %	$\delta_{\gamma p}$ %	δ_{had} %	δ_1 %	δ_2 %	δ_3 %	δ_4 %	δ_{tot} %
70	0.922×10^{-3}	1.385	2.57	5.67	1.10	0.12	−0.16	−0.29	−0.04	−0.29	0.01	−0.80	6.39
70	0.100×10^{-2}	1.434	2.18	2.55	0.86	−0.08	−0.11	−0.20	−0.02	0.08	0.01	−0.93	3.60
70	0.110×10^{-2}	1.506	1.88	2.10	0.81	−0.08	−0.11	−0.17	−0.01	0.48	0.01	−0.97	3.14
70	0.124×10^{-2}	1.445	1.48	1.80	0.80	−0.15	−0.10	−0.16	−0.01	0.24	0.01	−0.95	2.66
70	0.130×10^{-2}	1.414	1.32	1.32	0.82	−0.16	−0.08	−0.12	0.00	0.45	0.00	−1.09	2.37
70	0.200×10^{-2}	1.246	1.15	1.36	0.78	−0.03	−0.09	−0.12	0.00	0.22	0.01	−0.96	2.19
70	0.250×10^{-2}	1.190	0.96	0.97	0.77	0.03	−0.09	−0.12	0.00	0.13	0.01	−0.91	1.82
70	0.320×10^{-2}	1.084	0.91	0.78	0.74	0.26	−0.10	0.00	0.01	−0.13	0.03	−0.52	1.53
70	0.500×10^{-2}	0.958	0.83	0.74	0.75	0.23	−0.09	−0.09	−0.01	−0.49	0.05	−0.43	1.51
70	0.800×10^{-2}	0.819	1.62	0.48	0.77	0.26	−0.05	−0.11	0.01	−0.06	0.07	−0.05	1.88
70	0.130×10^{-1}	0.716	1.89	0.51	0.86	0.46	−0.07	−0.62	0.01	−0.06	0.05	−0.04	2.27
70	0.200×10^{-1}	0.637	1.77	0.67	0.80	0.37	−0.11	−0.46	−0.01	−0.06	0.07	−0.05	2.14
70	0.320×10^{-1}	0.561	2.10	0.92	0.85	−0.02	−0.05	0.26	−0.04	−0.05	0.07	−0.08	2.46
70	0.500×10^{-1}	0.512	1.61	0.90	1.19	0.35	−0.16	−0.37	−0.12	−0.03	0.00	0.02	2.27
90	0.130×10^{-2}	1.479	0.66	1.49	0.87	0.06	−0.04	−0.13	0.00	−0.05	0.15	0.24	1.87
90	0.150×10^{-2}	1.418	1.12	1.03	1.33	1.02	−1.18	−0.21	0.01	−0.04	0.08	0.12	2.56
90	0.200×10^{-2}	1.328	0.79	0.95	0.76	0.53	−0.25	−0.01	−0.01	0.06	0.07	−0.10	1.57
90	0.320×10^{-2}	1.165	0.66	0.63	0.75	0.45	−0.10	−0.08	0.00	0.08	−0.01	−0.20	1.29
90	0.500×10^{-2}	1.023	0.71	0.58	0.74	0.46	−0.09	−0.06	0.00	0.04	−0.03	−0.13	1.27
90	0.800×10^{-2}	0.883	0.87	0.56	0.78	0.38	−0.07	−0.12	0.00	−0.07	0.06	−0.16	1.37
90	0.130×10^{-1}	0.754	0.91	0.69	0.80	0.33	−0.08	−0.22	−0.01	−0.05	−0.22	−0.09	1.47
90	0.200×10^{-1}	0.649	0.95	0.62	0.81	0.36	−0.09	−0.19	0.00	−0.04	0.19	−0.08	1.47

- HERA DIS data are **final combined** H1 and ZEUS data
  - essentially provided as a single data set (no overlap)
  - combinantion served as a data consistency test
- Very complete description of correlated uncertainties
- Bin-by-bin unfolding (very good resolution of kinematic variables Q^2,x_{Bj})
  - however, sometimes at phase space corners a coarse binning had to be used
- Data are reported at (Q^2,x_{Bj}) **values**
  - although experimental measurements were done in *intervals* of Q^2,x_{Bj}
  - these intervals were different in H1 and ZEUS measurements
  - interpolation procedure (**swimming**) was applied to provide data at (Q^2,x_{Bj}) values
  - → potential model dependence, however, corresponding uncertainties were estimated (also older fixed-target DIS data sets were provided at (Q^2,x_{Bj}) values)
  - recent ZEUS analysis "Study of proton parton distribution functions at high x using ZEUS data" PRD 101 (2020) 11, 112009 published event counts and response matrices, but it is not easy to use these data together with the combined H1+ZEUS data

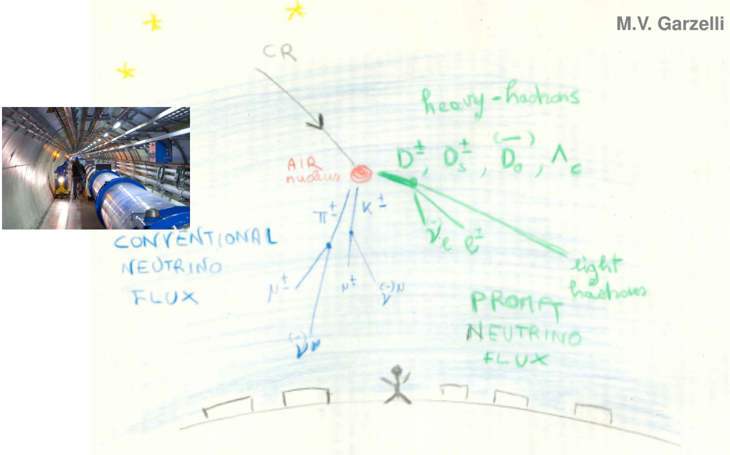
Atmospheric ν **are background for** *astrophysical* ν**:**



Production of atmospheric ν:

- cosmic rays (CR) + atmospheric nuclei
  $\rightarrow$ light and heavy hadrons $\rightarrow$ conventional and **prompt** ν fluxes
- spectra of conventional and **prompt** ν fluxes are different because of different hadroproduction cross sections and decay properties

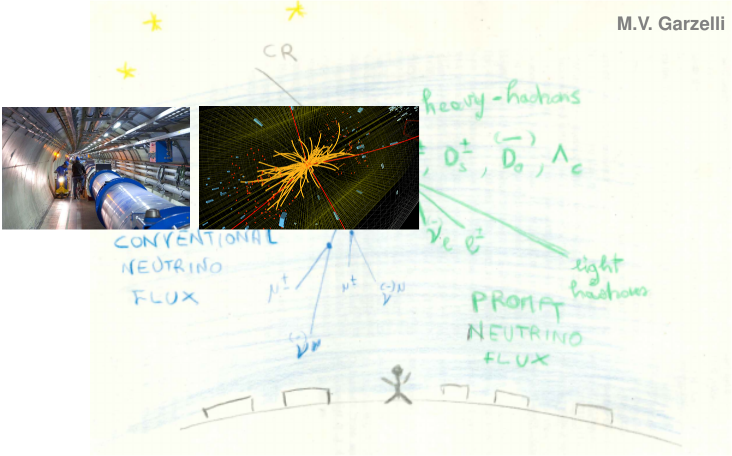
Atmospheric ν **are background for** *astrophysical* ν**:**



Production of atmospheric ν:

- cosmic rays (CR) $+$ atmospheric nuclei
  $\rightarrow$ light and heavy hadrons $\rightarrow$ conventional and **prompt** ν fluxes
- spectra of conventional and **prompt** ν fluxes are different because of different hadroproduction cross sections and decay properties

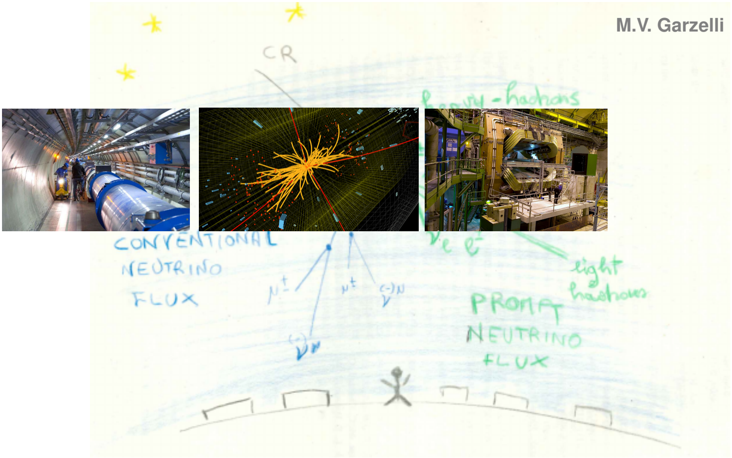
Atmospheric ν **are background for** *astrophysical* ν**:**



Production of atmospheric ν:

- cosmic rays (CR) + atmospheric nuclei
  $\rightarrow$ light and heavy hadrons $\rightarrow$ conventional and **prompt** ν fluxes
- spectra of conventional and **prompt** ν fluxes are different because of different hadroproduction cross sections and decay properties

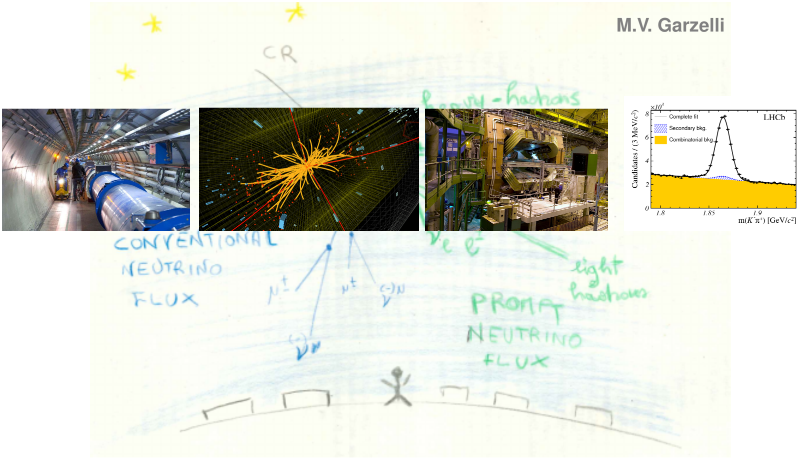
Atmospheric ν **are background for** *astrophysical* ν**:**



Production of atmospheric ν:

- cosmic rays (CR) $+$ atmospheric nuclei
  $\rightarrow$ light and heavy hadrons $\rightarrow$ conventional and **prompt** ν fluxes
- spectra of conventional and **prompt** ν fluxes are different because of different hadroproduction cross sections and decay properties

Atmospheric ν **are background for** *astrophysical* ν**:**

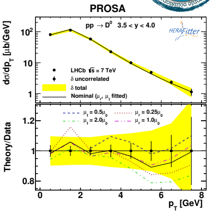


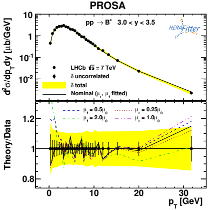
Production of atmospheric ν:

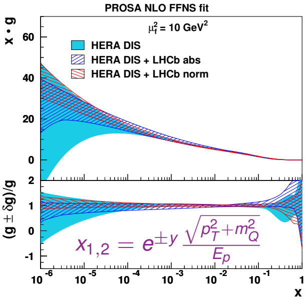
- cosmic rays (CR) $+$ atmospheric nuclei
  $\rightarrow$ light and heavy hadrons $\rightarrow$ conventional and **prompt** ν fluxes
- spectra of conventional and **prompt** ν fluxes are different because of different hadroproduction cross sections and decay properties

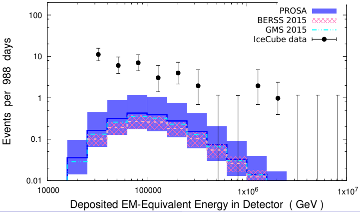
Charm production at LHC → gluon at low x → atmosphere ν fluxes

- LHCb measured:
  - charm $0 < p_T < 8$ GeV, $2 < y < 2.5$ [NPB871 (2013) 1]
  - beauty $0 < p_T < 40$ GeV, $2 < y < 2.5$ [JHEP08 (2013) 117]
- First QCD analysis of these data: [PROSA Coll., EPJ C75 (2015) 396]
- Improved gluon and sea-quark distributions up to $x \gtrsim 5 \times 10^{-6}$ (not covered by other experimental data)
  - used in next paper to predict IceCube background for very high energy cosmic ν [PROSA Coll., JHEP05 (2017) 004]
  - further update with ALICE and LHCb data [JHEP04 (2020) 118]



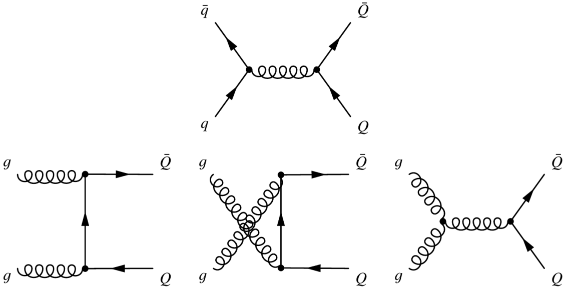






Ingredients:

- parton distribution functions (PDFs)
- hard-scattering perturbative partonic cross sections
- fragmentation functions / parton shower + hadronization



$$\sigma_{N_1 N_2 \to H+X} = \Sigma_{ab} PDF_a^{N_1}(x_a, \mu_F) PDF_b^{N_2}(x_b, \mu_F) \otimes$$
$$\otimes \hat{\sigma}_{ab \to cX}(x_a, x_b, z, \mu_F, \mu_R, \alpha_S(\mu_R), m_c) \quad (+\text{fragmentation})$$

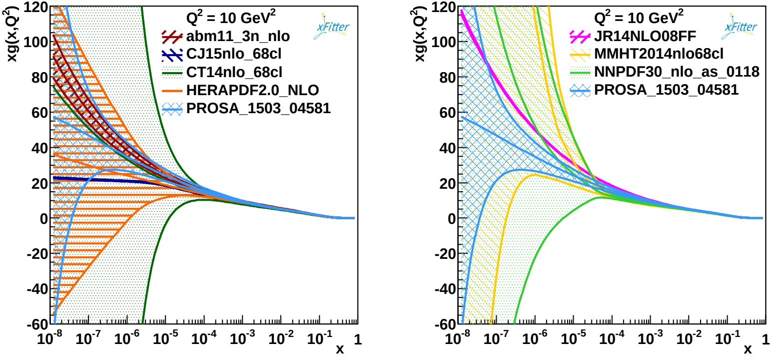
Theoretical description of charm hadroproduction

$$\sigma_{N_1 N_2 \to H+X} = \Sigma_{ab} PDF_a^{N_1}(x_a, \mu_F) PDF_b^{N_2}(x_b, \mu_F) \otimes$$
$$\otimes \hat{\sigma}_{ab \to cX}(x_a, x_b, z, \mu_F, \mu_R, \alpha_S(\mu_R), m_c) \quad (+\text{fragmentation})$$

Uncertainties in computation:

- proton/nuclear PDF uncertainties at small parton memonetum fraction x (can be very large for high $E \longleftrightarrow$ small parton x)
- perturbative hadroproduction cross section: NLO is state-of-the-art fixed-order prediction for differential charm production cross sections
  $\rightarrow$ various ways of treating heavy quark mass, fixed-flavour-number scheme (FFNS) vs. general-mass variable-flavour-number scheme (GM-VFNS)
- fragmentation $c \rightarrow D, \Lambda$

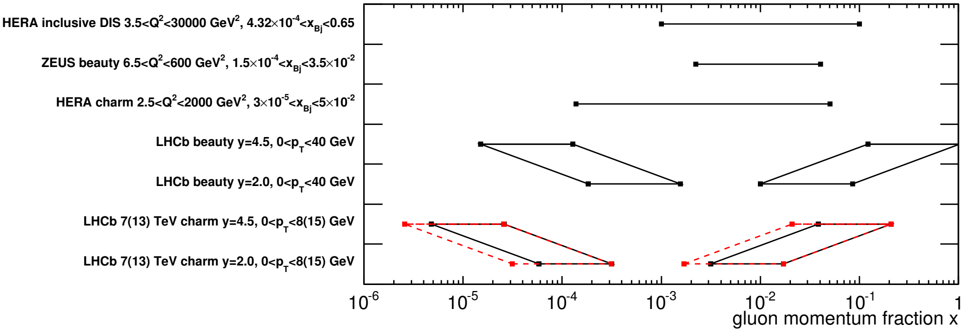
Gluon PDFs (in 2015)



[1611.03815]

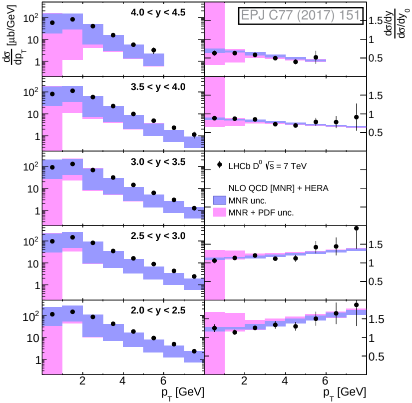
- Production of prompt neutrinos requires as input gluon PDF at low x, previously not constrained by any data
  - depends on additional assumption (e.g. PDF parametrisation)
  - *gluon PDF uncertainties can be arbitrary large at low x!*
- PROSA fit was the first QCD analysis of heavy-flavour (HF) data from LHCb, which extended coverage of gluon PDF to $x \gtrsim 10^{-6}$ [1503.04581]
- Similar studies were later presented by R. Gauld et al. [1506.08025, 1610.09373]

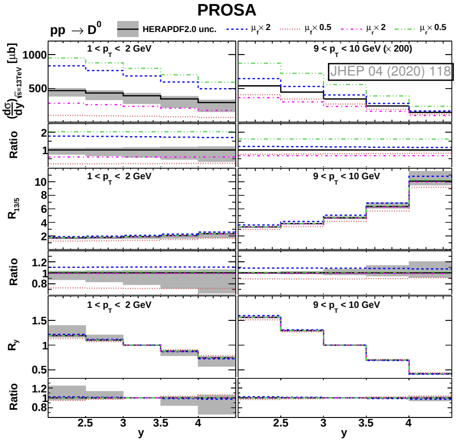
Kinematics of HF hadroproduction at LHCb



- LHCb measured double-diff. $\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_T\mathrm{d}y}$ for c and b at 7 TeV, $2.0 < y < 4.5$:
  - charm: $0 < p_T < 8$ GeV [NPB871 (2013) 1]
  - beauty: $0 < p_T < 40$ GeV [JHEP08 (2013) 117]
- At LO: $x_{1,2} = \frac{\sqrt{p_T^2 + m_Q^2}}{E_P} \mathrm{e}^{\pm y}$
- c @ 7 TeV data: $x \gtrsim \frac{1.4}{3500} \mathrm{e}^{-4.5} \approx 4 \times 10^{-6}$ (uncovered by HERA data)
- c @ 13 TeV: even lower x, however not in PROSA 2015 fit
- b @ 7 TeV: $x \lesssim 1$, however not yet studied carefully

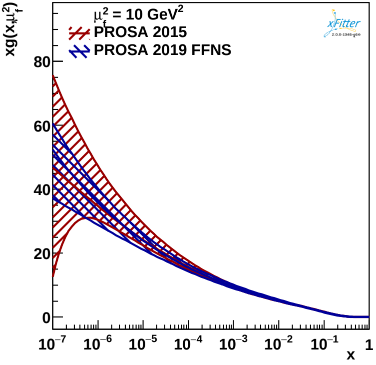
Charm production at LHC: dealing with large NLO scale variation uncertainties

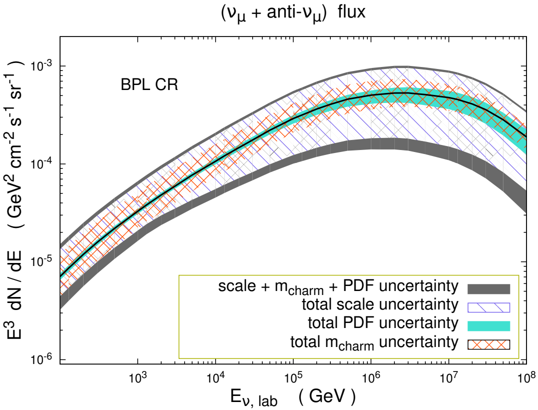




- Taking **ratios** of data → reducing scale uncertainties by one order of magnitude
- Ratio to another kinematic region might be better than (traditional) ratio to another c.m.e.
- This relies on assumption that scale unc. are correlated across different kinematic regions:
  - can be tested e.g. by using another (dynamical) scale choice

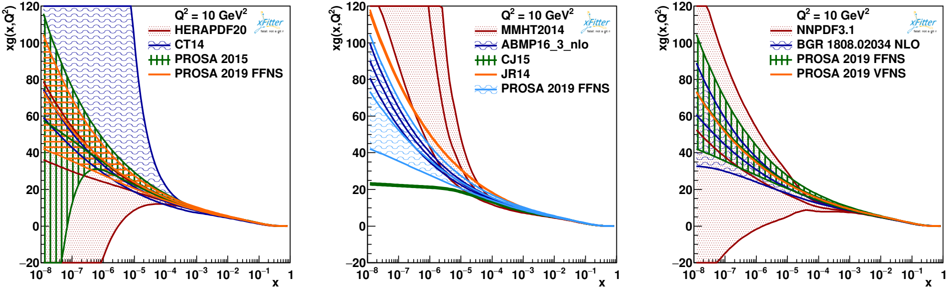
PROSA PDF fit [arXiv:1911.13164, JHEP 04 (2020) 118]





- Stronger constraints on gluon PDF at low *x* in PROSA2019
- Reduced PDF uncertianties for prompt ν flux at high energies

Comparison with other PDF sets



- In general, reduced gluon uncertainties at low *x* compared to most of other global PDF fits (MMHT, CT, NNPDF, HERAPDF)
  - other fits do not use LHCb heavy-flavour data
- Though some other PDF fits have smaller uncertainties purely because of rigid gluon parametrisation at low *x*
  - ABMP16 turns out to be compatible with PROSA2019
- Good agreement with results from 1808.02034 (also using LHCb charm data), both for central values and uncertainties

LHCb, ALICE data: discussion

Typical description of correlated systematic uncertainties

LHCb 5 TeV JHEP06 (2017) 147

Table 2: Fractional systematic uncertainties, in percent. Uncertainties that are computed bin-by-bin are expressed as ranges giving the minimum to maximum values. Ranges for the correlations between p_{T}-y bins and between modes are also given, expressed in percent.

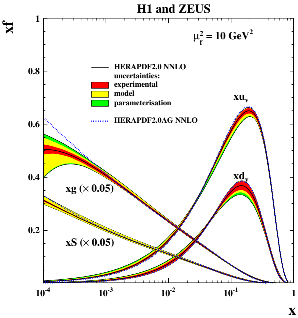
	Uncertainties (%)				Correlations (%)	
	D^0	D^+	D_s^+	D^{*+}	Bins	Decay modes
Luminosity		3.8			100	100
Tracking	3–5	5–7	4–7	5–7	90–100	90–100
Branching fractions	1.2	2.1	5.8	1.5	100	0–95
Simulation sample size	0–10	0–10	2–9	1–10	0	0
Simulation modelling	0.3	0.7	0.6	2	0	0
PID sample size	0–1	0–1	0–2	0–2	0–100	0–100
PID binning	0–30	0–10	0–20	0–20	0	0
Fit model shapes	0–3	0–3	0–3	0.0–1.0	0	0

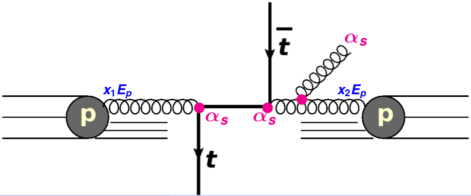
This information is not really sufficient:

- need to know contributions of different systematic uncertainties for each bin (not just ranges)
- need to know correlation betwen different *D* and energies
- *total* covariance matrices were provided for some LHCb data sets, but
  - some of them appeared to be not positive definite (issue of rounding?) $\rightarrow$ similar problem with ATLAS $t\bar{t}$ data
  - they still do not allow one to properly correlate different data sets

Global ABMP16tt fit [Alekhin, Garzelli, Moch, Zenaiev arXiv:2407.00545]

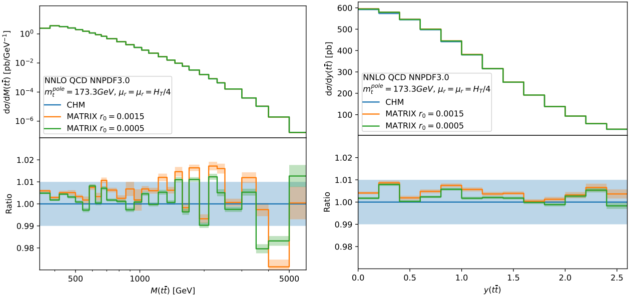
- Follows ABMP16 PDF fit PRD 96 (2017) 014011
  - used ABMP framework, xFitter was only partially used for cross checks, plotting etc.
- In this work we focus on gluon PDF: it is a challenge because it is constrained by DIS data (backbone of PDF fits) only indirectly
  - especially at large x (important for BSM searches at LHC)
  - in addition, g is strongly correlated with α_S
- Processes to constrain g:
  - heavy-quark production: NNLO, but depend on m_Q
  - jet production: larger dependence on NNLO scale choice
- We focused on double-differential $t\bar{t}$ LHC data $d^2\sigma/M(t\bar{t})y(t\bar{t})$:
  - $M(t\bar{t}) \approx \sqrt{sx_1 x_2}$ provides sensitivity to m_t and PDFs
  - $y(t\bar{t}) \approx \frac{1}{2} \ln \frac{x_1}{x_2}$ is sensitive especially to PDFs





Theoretical calculations and scope of our work

- NNLO calculations for total and fully differential $t\bar{t}$ (q_T subtraction) are publicly available with MATRIX framework [Catani et al. PRD99 (2019) 5, 051501] [Catani et al. JHEP 07 (2019) 100]
  - fully differential NNLO calculations were also published in [Czakon, Heymes, Mitvo JHEP 04 (2017) 071] [CHM], but no public code available (apart from the HighTEA datbase)



- We have interfaced MATRIX $t\bar{t}$ to PineAPPL interpolation library: **no NNLO/NLO K-factors**
- [Garzelli, Mazzitelli, Moch, Zenaiev JHEP 05 (2024) 321]: extraction of m_t^{pole} from the total and differential $t\bar{t}$ data using different PDF sets
- **[Alekhin, Garzelli, Moch, Zenaiev arXiv:2407.00545]: global PDF+α_S+$m_t(m_t)$ fit within the ABMP16 framework**
  - using collider and fixed-target DIS, DY (see BACKUP), updated single top and $t\bar{t}$ data

PineAPPL grids for 2D $t\bar{t}$ & xFitter

- PineAPPL grids store genuine NNLO predictions with their exact PDFs, α_S, $\mu_{r,f}$ dependence
  - 2M CPU hours (billions of events): 3 months of real time
  - numerical uncertainty in bins $\lesssim 0.5\%$
  - the resulting tables occupy 6 GB
- the grids were produced for $165 \leq m_t^{pole} \leq 177.5$ GeV with 2.5 GeV step: m_t^{pole} dependence via interpolation
- fine binning in $M(t\bar{t})$ abd $y(t\bar{t})$ which is suitable for all existing Run 1 and 2 ATLAS and CMS measurements at parton level via re-binning
  - can be used at another $\sqrt{s}$
  - can be used for another (slightly) different binning via intepolation
  - $\rightarrow$ hopefully, can be re-used at least for Run-3 measurements
- **work-in-progress to make our grids public via Ploughshare and provide xFitter example how to use them**

χ^2 for different variants of the ABMPtt fit

Experiment	Data set	$\sqrt{s}$ (TeV)	Reference	NDP	χ^2		
					I	II	III
ATLAS	$ATLAS\,13_{ljet}$	13	[25]	19	34.2	27.2	–
	$ATLAS\,13_{had}$	13	[26]	10	11.8	12.1	–
CMS	$CMS\,13_{ll}$	13	[24]	15	21.1	–	19.9
	$CMS\,13_{ljet}$	13	[22]	34	42.2	–	40.8

I: both ATLAS and CMS

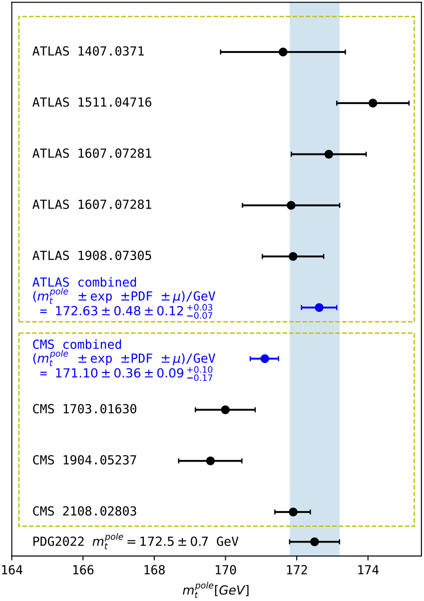
II: only ATLAS

III: only CMS

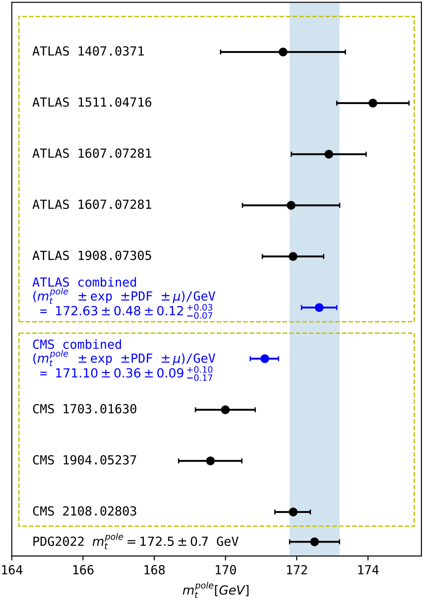
$\rightarrow$ **Overall good description of data by NNLO theoretical predictions, but some tension between ATLAS and CMS differential $t\bar{t}$ data is noticeable**

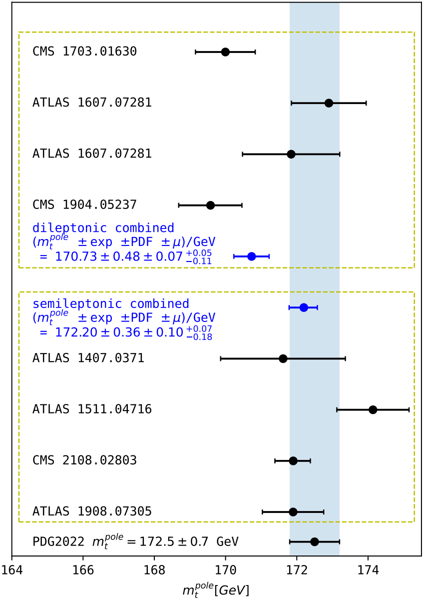
$\rightarrow$ **ATLAS + CMS effort on combining differential $t\bar{t}$ cross sections will be useful**

Extraction of m_t^{pole} using ABMP16 JHEP 05 (2024) 321 [arXiv:2311.05509]

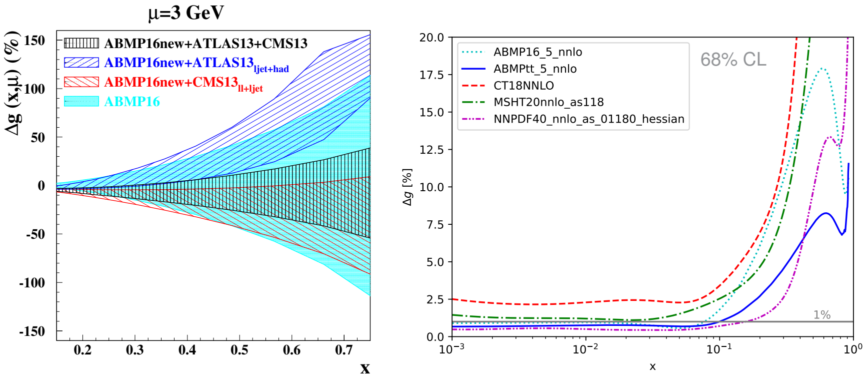


Extraction of m_t^{pole} using ABMP16 JHEP 05 (2024) 321 [arXiv:2311.05509]





Fitted gluon PDF



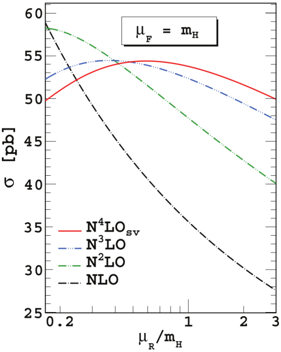
- Significant reduction of the gluon PDF uncertainty once differential $t\bar{t}$ data are included
- The fitted gluon PDF is consistent with ABMP16
- Fitted $\alpha_S(M_Z) = 0.1150 \pm 0.0009$ in ABMPtt (vs fixed $\alpha_S(M_Z) = 0.118$ in other fits)

→ **we have confirmed ABMP16 gluon and α_S with new data**

PDF Name	N2LO	N3LO	N4LOsv
ABMP16	$(45.4\pm4.6)^{+0.7}_{-0.7}$	$(49.6\pm2.6)^{+0.8}_{-0.8}$	$(50.8\pm1.9)^{+0.9}_{-0.9}$
ABMPtt	$(45.0\pm4.6)^{+0.6}_{-0.6}$	$(49.2\pm2.6)^{+0.7}_{-0.7}$	$(50.4\pm1.9)^{+0.8}_{-0.8}$
CT18NNLO	$(47.4\pm5.1)^{+1.3}_{-1.7}$	$(52.0\pm2.9)^{+1.4}_{-1.9}$	$(53.4\pm2.1)^{+1.5}_{-1.9}$
MMHT2014nnlo68cl	$(47.7\pm5.1)^{+0.6}_{-0.8}$	$(52.3\pm2.9)^{+0.7}_{-1.0}$	$(53.8\pm2.2)^{+0.7}_{-1.0}$
MSHT20nnlo_as118	$(47.4\pm5.1)^{+0.5}_{-0.6}$	$(52.0\pm2.9)^{+0.6}_{-0.6}$	$(53.4\pm2.1)^{+0.6}_{-0.6}$
NNPDF40_nnlo_as_01180	$(47.8\pm5.1)^{+0.3}_{-0.3}$	$(52.4\pm2.9)^{+0.3}_{-0.3}$	$(53.8\pm2.2)^{+0.3}_{-0.3}$
PDF4LHC21_40	$(47.6\pm5.1)^{+0.8}_{-0.8}$	$(52.3\pm2.9)^{+0.9}_{-0.9}$	$(53.7\pm2.2)^{+0.9}_{-0.9}$
MSHT20an3lo_as118	$(45.0\pm4.8)^{+0.8}_{-0.7}$	$(49.4\pm2.8)^{+0.9}_{-0.8}$	$(50.7\pm2.0)^{+0.9}_{-0.8}$

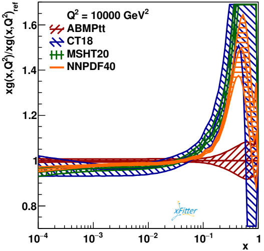
Table 1: Higgs cross-section along with the absolute error obtained from seven-point scale variation around $(\mu_R{}^c, \mu_F{}^c) = (1,1)m_H$ as well as intrinsic PDF uncertainty using LHAPDF. $\sqrt{S} = 14$ TeV, α_S from LHAPDF (NNLO value).

PLB 807 (2020) 135546



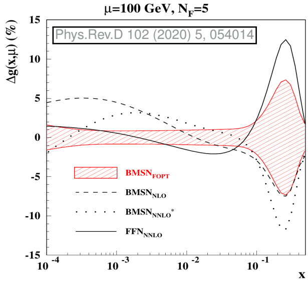
- Das, Moch, Vogt, Phys.Lett.B 807 (2020) 135546:
  - N4LOsv: soft virtual ggF corrections at 4 loops
  - N3LO: effective theory for $m_t \gg m_H$
  - N2LO: full theory for $m_H \lesssim m_t$
  - → apparent convergence of perturbative series (further details and predictions with $\mu/2$ in BACKUP)
- N4LOsv estimates missing higher-order corrections: 2%
- Larger differences originate from PDF + α_S sets:

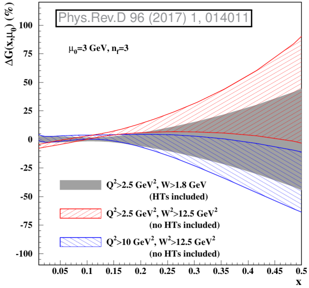
  7% (1995) → 12% (2020) → **7% (2024)** (more in BACKUP)
- Expect smaller effect of NNLO→N3LO PDFs



Some reasons for difference vs. other global fits

- Heavy-flavor PDF evolution and heavy flavour scheme for DIS
  Phys.Rev.D 102 (2020) 5, 054014
  - fixed-flavour number scheme in ABMP fits is accurate to NNLO for light and heavy quark production (exact or to the best available approximation) and works very well for HERA kinematics
  - CT, MSHT and NNPDF fits use different variable-flavour number schemes which miss some NNLO corrections for heavy quark production and have further theoretical uncertainties $\Rightarrow$ very relevant for LHC phenomenology
- Higher-twist (HT) corrections for DIS
  Phys.Rev.D 96 (2017) 1, 014011
  - ABMPtt/ABMP16 fits HT terms
  - other fits apply cuts to reduce HT: not sufficient
- Correlation between gluon PDF, α_S and m_t:
  - fully accounted for in ABMPtt/ABMP16
  - α_S and m_t are fixed in other fits
- **These issues need to be solved independently of switching to N^{3}LO evolution**





Target mass corrections (TMC)

- Description from Phys.Rev.D 86 (2012) 054009 (ABM11):

  In DIS the power corrections arise from kinematic considerations once the hadron mass effects are taken into account, i.e., the so-called target mass correction (TMC). The TMC can be calculated in a straightforward way from the leading twist PDFs within the OPE [57]. In our analysis the TMC are taken into account in the form of the Georgi-Politzer prescription [57]. For relevant observables, i.e., the structure function F_2 and the transverse one F_T it reads

$$F_T^{\mathrm{TMC}}(x,Q^2) = \frac{x^2}{\xi^2\gamma} F_T(\xi,Q^2) + 2\frac{x^3 M_N^2}{Q^2\gamma^2} \int_\xi^1 \frac{d\xi'}{\xi'^2} F_2(\xi',Q^2)\,, \quad (2.10)$$

  and

$$F_2^{\mathrm{TMC}}(x,Q^2) = \frac{x^2}{\xi^2\gamma^3} F_2(\xi,Q^2) + 6\frac{x^3 M_N^2}{Q^2\gamma^4} \int_\xi^1 \frac{d\xi'}{\xi'^2} F_2(\xi',Q^2)\,, \quad (2.11)$$

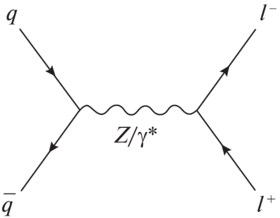
  respectively, which holds up to $\mathcal{O}(M_N^2/Q^2)$. Here $\xi = 2x/(1+\gamma)$ and $\gamma = (1+4x^2M_N^2/Q^2)^{1/2}$ is the Nachtmann variable [58]. The quantities on the right hand side of eqs. (2.10) and (2.11) are the leading twist structure functions introduced in eq. (2.7) above.

  [57] H. Georgi and H. Politzer, Phys.Rev. D14, 1829 (1976)

  [58] O. Nachtmann, Nucl.Phys. B63, 237 (1973).

- **“TMC is most important for SLAC, less important for BCDMS, almost unimportant for NMC, and negligible for HERA” [Phys.Rev. D63 (2001) 094022]**
  - **70% for some SLAC data points**
- TMC are avaibale in development xFitter branch

DY production at LHC

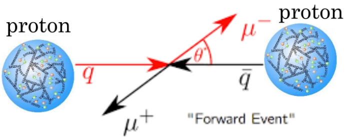


- Drell Yan (DY) lepton pair production at the LHC is a useful process to test SM and probe proton PDFs (quark and antiquark distributions). At LO QCD:

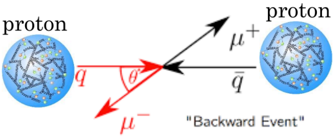
$$\frac{d\sigma}{dM(ll)dy(ll)dcos\theta} \sim \frac{d\sigma}{dM(ll)dy(ll)} \left[(1 + cos^2\theta) + A_4 cos\theta\right]$$

- The θ angle is defined w.r.t the direction of the boost of the di-lepton system *ll* (correlated with the direction of the incoming quark)

σ_F cross-section from forward events



σ_B cross-section from backward events

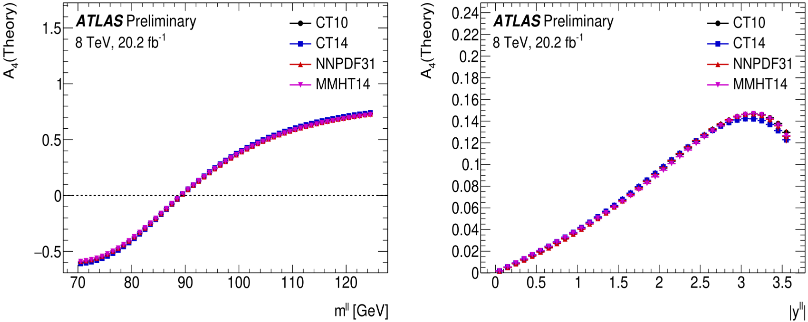


DY Forwrd-Backward Asymmetry (AFB)

- Forward-Backward Asymmetry (AFB) is a clean observable for which many experimental and theoretical uncertainties cancel:

$$A_{FB} = 3/8A_4 = \frac{\sigma_F - \sigma_B}{\sigma_F + \sigma_B}$$
$$= \frac{d^2\sigma/dM(ll)dy(ll)[cos\theta^* > 0] - d^2\sigma/dM(ll)dy(ll)[cos\theta^* < 0]}{d^2\sigma/dM(ll)dy(ll)[cos\theta^* > 0] + d^2\sigma/dM(ll)dy(ll)[cos\theta^* < 0]}$$

- ...but AFB is still sensitive to the PDFs [ATLAS-CONF-2018-037]:



- AFB was used to constrain PDFs (e.g. JHEP 10 (2019) 176)

- Traditionally AFB is used to measure the weak mixing angle (e.g. ATLAS-CONF-2018-037)
- **We explore AFB potential to constrain Z boson couplings at future HL-LHC**
- DY cross sections (and hence AFB) depend on the Z boson coupling to fermions:

$$\mathrm{d}\sigma/\mathrm{d}M\mathrm{d}y\mathrm{d}\cos\theta^* = F(g_V^{Zu}, g_A^{Zu}, g_V^{Zd}, g_A^{Zd}, g_V^{Ze}, g_A^{Ze})$$

- In the SM:

$$g_V^{Zu} = \frac{1}{2} - \frac{4}{3}\sin^2\theta_W, \quad g_A^{Zu} = \frac{1}{2}$$
$$g_V^{Zd} = -\frac{1}{2} + \frac{2}{3}\sin^2\theta_W, \quad g_A^{Zu} = -\frac{1}{2}$$

- In the SMEFT up to dimension $D = 6$:

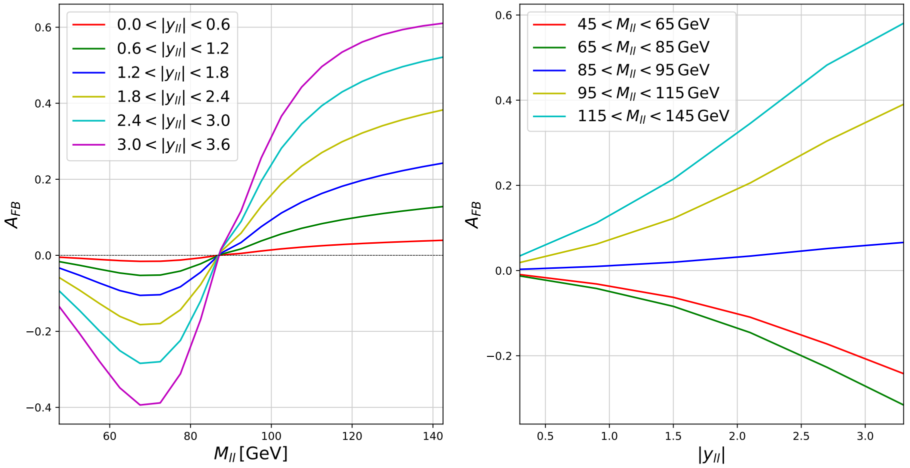
$$\mathcal{L} = \mathcal{L}^{(\mathrm{SM})} + \frac{1}{\Lambda^2}\sum_{j=1}^{N_6} C_j^{(6)}\,\mathcal{O}_j^{(6)}$$

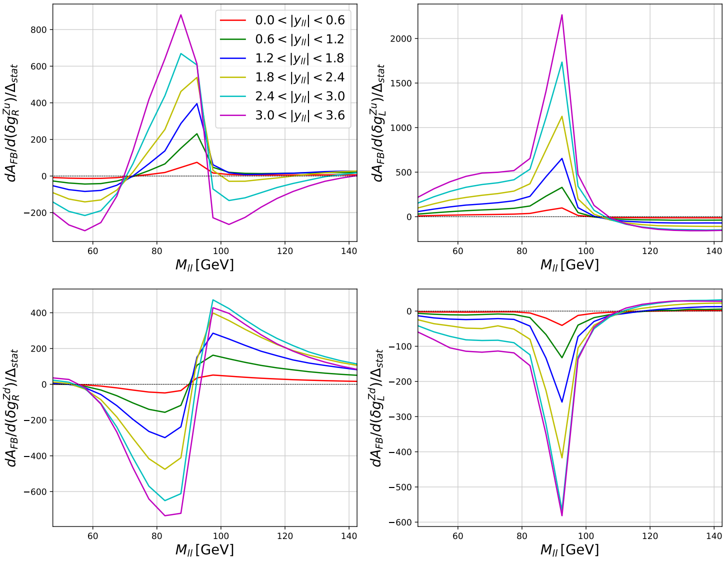
- In the dilepton mass region not too far from the Z-boson peak the whole effect of the $D = 6$ SMEFT Lagrangian is a modification of the vector boson couplings to fermions
- Couplings to leptons g_V^{Ze}, g_A^{Ze} are well constrained by LEP data
- 4-fermion operators are not included: $< 1\%$ at $M(ll) < 150$ GeV
- We fit four parameters δ (assuming $g_{A,V}^{Zu} = g_{A,V}^{Zc}$, $g_{A,V}^{Zd} = g_{A,V}^{Zs} = g_{A,V}^{Zb}$) which are $= 0$ in the SM (R, L couplings are linear combination of V, A couplings):

$$g_L^{Zu} \equiv g_{L(\mathrm{SMEFT})}^{Zu} = g_{L(\mathrm{SM})}^{Zu} + \delta g_L^{Zu}, \quad g_R^{Zu} \equiv g_{R(\mathrm{SMEFT})}^{Zu} = g_{R(\mathrm{SM})}^{Zu} + \delta g_R^{Zu}$$
$$g_L^{Zd} \equiv g_{L(\mathrm{SMEFT})}^{Zd} = g_{L(\mathrm{SM})}^{Zd} + \delta g_L^{Zd}, \quad g_R^{Zd} \equiv g_{R(\mathrm{SMEFT})}^{Zd} = g_{R(\mathrm{SM})}^{Zd} + \delta g_R^{Zd}$$

DY AFB as function of $M(ll)$ and $y(ll)$ [LO]

In order to maximize the sensistivity, we use double-differential AFB as function of $M(ll)$ and $y(ll)$

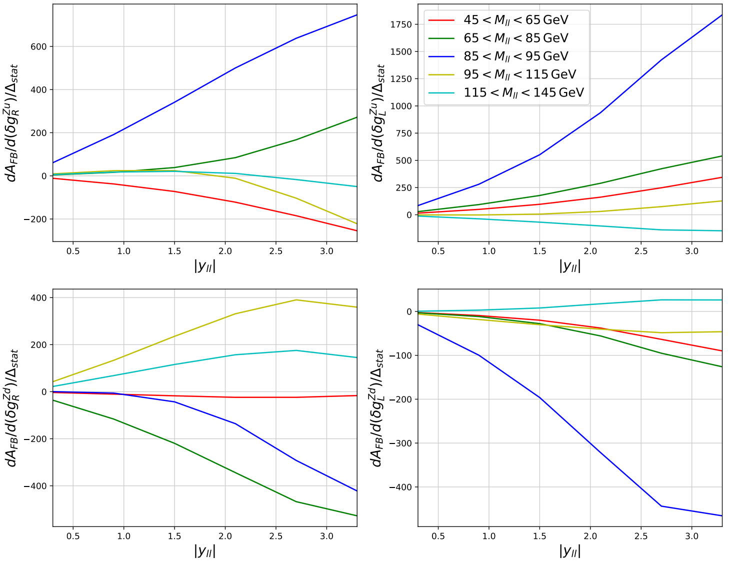




- Statistical uncertainties increase outside of the Z peak

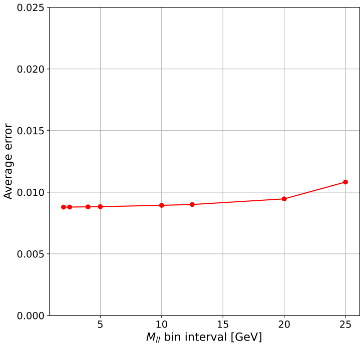
$\rightarrow$ we do not go to very low or high $M(ll)$ [also minimise impact of 4-fermion operators]

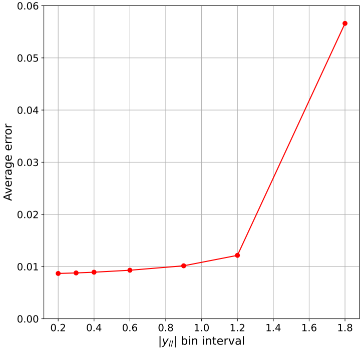
DY AFB $y(ll)$ derivatives w.r.t the couplings divided by stat. unc.



- AFB $\to 0$ as $y(ll) \to 0$ due to its definition at the LHC (w.r.t the longitudinal boost of ll)
- Best sensitivity comes from largest reachable $y(ll)$ values: limited by detector acceptance

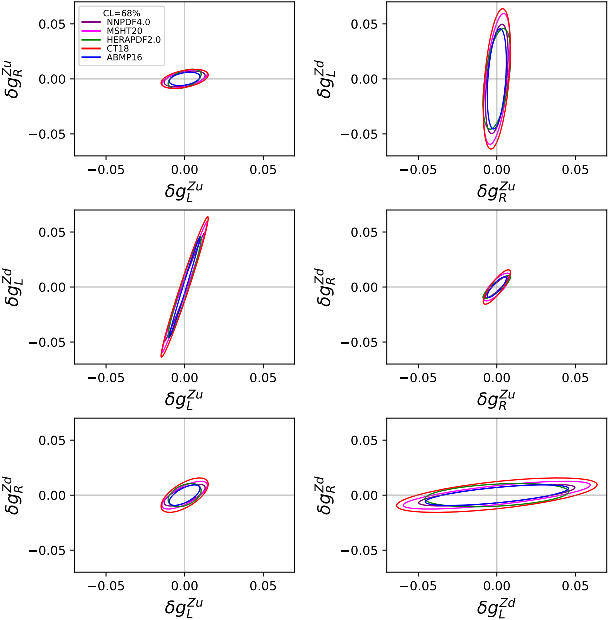
Binning scheme and analysis setup



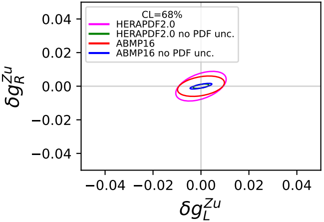


- We chose 5 GeV of $M(ll)$ and 0.6 bins of $y(ll)$ (experimentally feasible)
- Kinematic region: $45 < M(ll) < 145$ GeV, $0 < |y(ll)| < 3.6$
- Assume HL-LHC luminosity of 3000 fb^{-1} and 20% detector correction factor
- PDF uncertainties are included using the profiling technique (constrained by (pseudo-)data)
- The fits are done at LO (sensitivity study only) using **xFitter framework**

Results: fitted couplings using different PDF sets

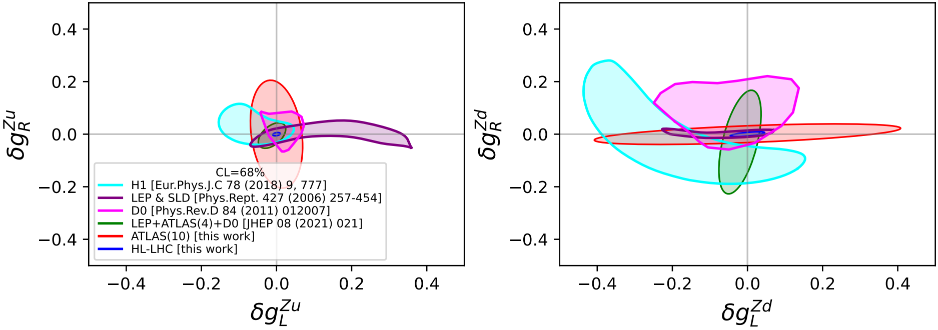


- Similar sensitivity is achieved when using different modern PDF sets
- A large fraction of the uncertainties are from PDFs



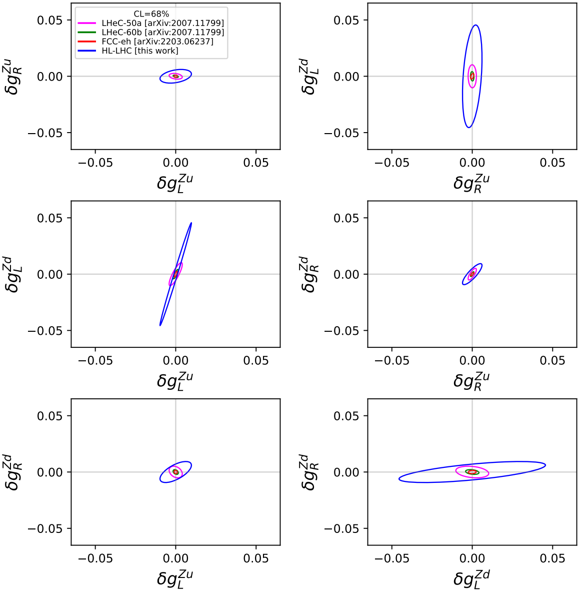
→ Ideally, a future extraction should be done in a simultaneous PDF fit

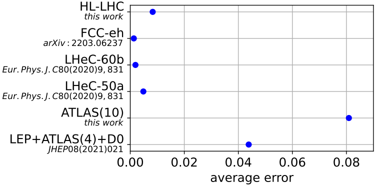
Results: comparison with existing extractions



- HL-LHC has significant potential for improving current constraints compared to the current LHC data [ATLAS(10)]

Results: comparison with other future experiments





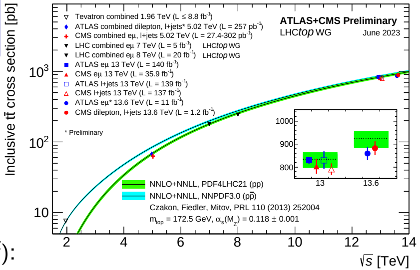
- HL-LHC significantly improves LHC limits, and approaches the precision of LHeC and FCC-eh

Outlook

- xFitter is a modern tool for data vs theory comparison:
  - https://gitlab.cern.ch/fitters/xfitter
- Data and theory predictions for a variety of processes (DIS, DY, $t\bar{t}$, jets) at NLO, NNLO and (partially) N3LO in SM and beyond
  - sufficient for a global PDF + SM (+ BSM) parameters fit
- Flexibility: one can do a fit, sensitivity study, just compare new data and theory to make plots etc.
- Modern C++ code with well developed interfaces: easy to contribute and write new code
- Nuclear PDFs, TMD PDFs are also available in separate (sometimes inofficial) branches: not covered in this talk, but feel free to ask
- Contact us: xfitter-users@googlegroups.com

BACKUP

Data used in this analysis



Selection of data:

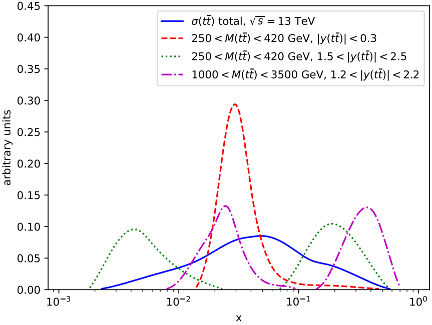
- LHC and Tevatron measurements of single t:
  - 6 data points from LHCtopWG
- LHC and Tevatron measurements of total $\sigma(t\bar{t})$:
  - 10 data points from LHCtopWG, including ATLAS+CMS combination at 7 & 8 TeV
- differential measurements $\frac{1}{\sigma(t\bar{t})}\frac{d\sigma(t\bar{t})}{dO}$ which satisfy following criteria:
  - as function of $M(t\bar{t})$ (if available, 2D $M(t\bar{t})$ and $y(t\bar{t})$)
  - unfolded to parton level (no cuts on p_T, y of leptons or jets): no LHCb data
  - bin-by-bin correlations are available (no Tevatron data)
  - normalized cross sections (to avoid unknown correlation with total $\sigma(t\bar{t})$ and to reduce unknown correlations between different data sets)
  - for the moment only Run-2 2D data included in the PDF fit (besides the total $t\bar{t}$ x-section data)

$$\frac{1}{\sigma(t\bar{t})}\frac{d\sigma(t\bar{t})}{dO}$$

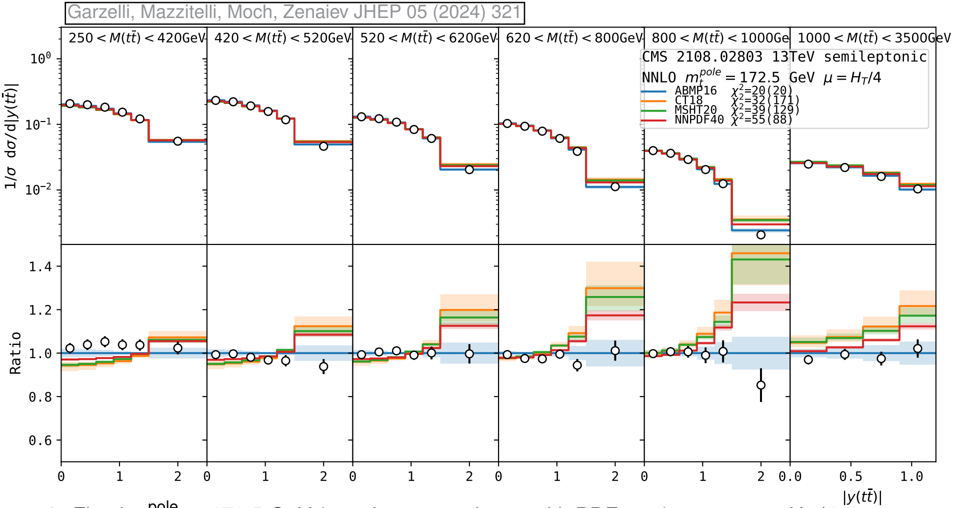
Experiment	decay channel	dataset	luminosity	$\sqrt{s}$	observable(s)	n
CMS	semileptonic	2016–2018	137 fb^{-1}	13 TeV	$M(t\bar{t})$, $\|y(t\bar{t})\|$	34
CMS	dileptonic	2016	35.9 fb^{-1}	13 TeV	$M(t\bar{t})$, $\|y(t\bar{t})\|$	15
ATLAS	semileptonic	2015–2016	36 fb^{-1}	13 TeV	$M(t\bar{t})$, $\|y(t\bar{t})\|$	19
ATLAS	all-hadronic	2015–2016	36.1 fb^{-1}	13 TeV	$M(t\bar{t})$, $\|y(t\bar{t})\|$	10
CMS	dileptonic	2012	19.7 fb^{-1}	8 TeV	$M(t\bar{t})$, $\|y(t\bar{t})\|$	15
ATLAS	semileptonic	2012	20.3 fb^{-1}	8 TeV	$M(t\bar{t})$	6
ATLAS	dileptonic	2012	20.2 fb^{-1}	8 TeV	$M(t\bar{t})$	5
ATLAS	dileptonic	2011	4.6 fb^{-1}	7 TeV	$M(t\bar{t})$	4
ATLAS	semileptonic	2011	4.6 fb^{-1}	7 TeV	$M(t\bar{t})$	4

LO: $x_{1,2} = (M(t\bar{t})/\sqrt{s}) \exp[\pm y(t\bar{t})]$

- Double-differential $t\bar{t}$ @ 13 TeV probes $0.002 \lesssim x \lesssim 0.7$
  - gg contributes $\approx 90\%$
- E.g. Higgs production at the LHC probes $x \sim m_H/\sqrt{s} \sim 0.01$ which is well covered by differential $t\bar{t}$ data
- Energy scales m_H, M_W, M_Z and m_t are similar
- DY production at the LHC probes a similar region $x \sim m_{W,Z}/\sqrt{s}$
  - sensitive to quark PDFs mostly



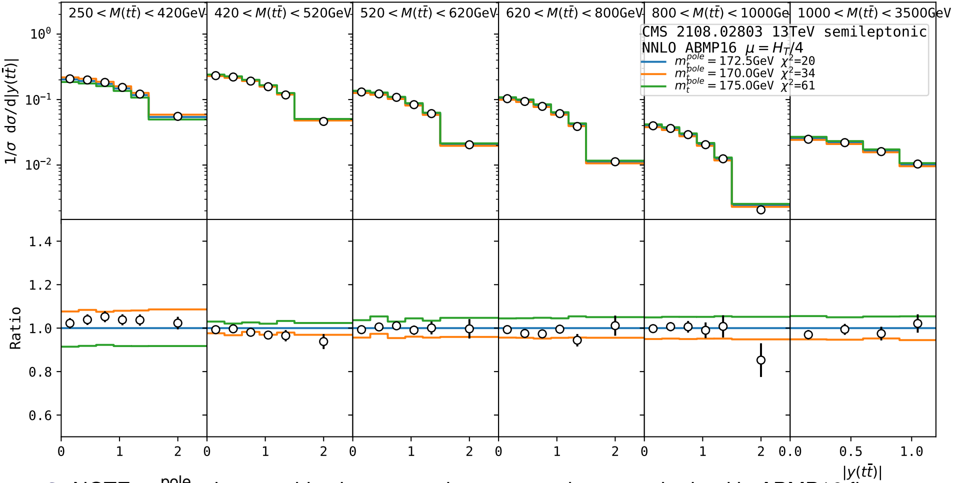
Example: CMS 2108.02803 vs NNLO predictions using different PDFs



- Fixed $m_t^{\text{pole}} = 172.5$ GeV (not always consistent with PDF sets), $\mu_r = \mu_f = H_T/4$
- Reported χ^2 values with (and without) PDF uncertainties
- All PDF sets describe data reasonably well, with best description by ABMP16
  - CT18, MSHT20 and NNPDF40 show clear trend w.r.t data at high $y(t\bar{t})$ (large x)

Example: CMS 2108.02803 vs NNLO predictions using different m_t^{pole}

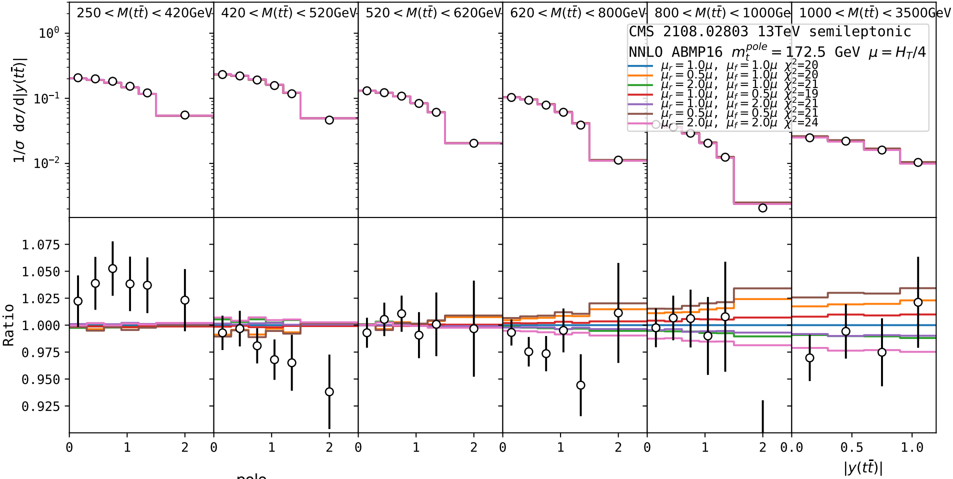
Garzelli, Mazzitelli, Moch, Zenaiev JHEP 05 (2024) 321



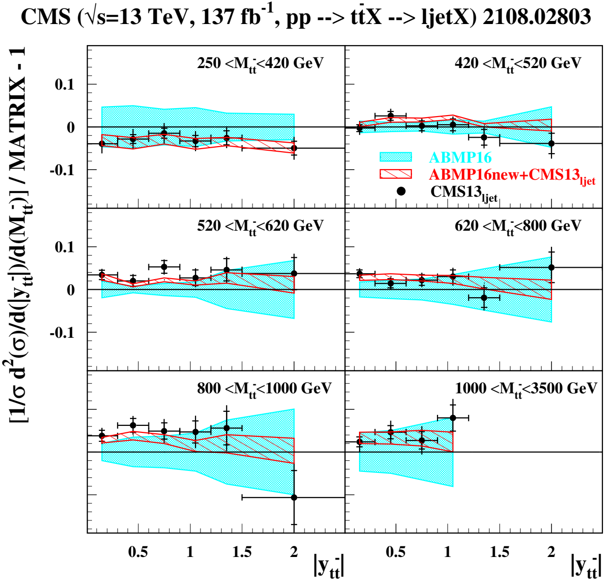
- NOTE: m_t^{pole} values on this plot are not the same as the ones obtained in ABMP16 fit (ABMP16 $m_t(m_t) = 160.9 \pm 1.1$ GeV corresponds at 3 loops to $m_t^{\text{pole}} = 170.4 \pm 1.2$ GeV)
- Low $M(t\bar{t})$: strong dependence on m_t^{pole} via threshold effects
- High $M(t\bar{t})$: opposite dependence due to cross-section normalization

Example: CMS 2108.02803 vs NNLO predictions: scale variations

Garzelli, Mazzitelli, Moch, Zenaiev JHEP 05 (2024) 321

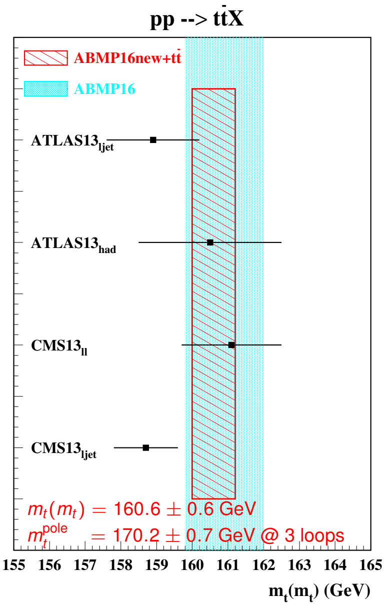


- ABMP16, fixed $m_t^{\text{pole}} = 172.5$ GeV
- Scale variations $< 1\%$ at low $M(t\bar{t})$ (largest cancellation), reach $\approx 4\%$ at high $M(t\bar{t})$
  - $\rightarrow$ **these data are useful to provide constraints on m_t and PDFs**

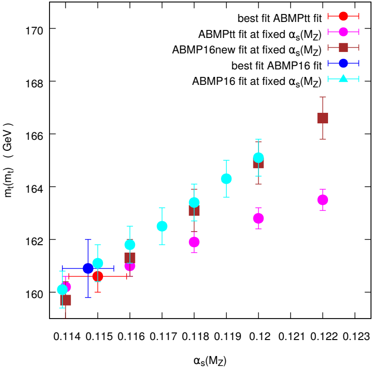


→ these data are in good agreement with NNLO theoretical predictions and put significant constraints on the PDFs

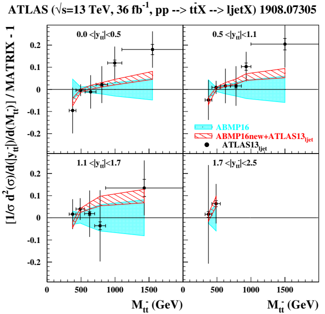
$m_t(m_t)$ and α_s in ABMP16 PDF fit

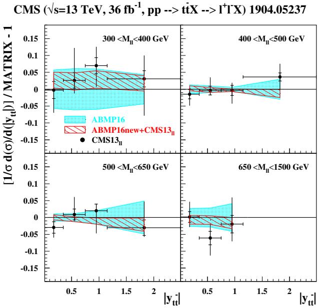


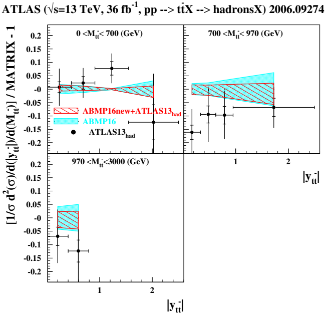
- Overall, good agreement among $m_t(m_t)$ extracted from different data sets
- Good agreement with ABMP16 fit and $\sim 50\%$ reduced uncertainty on $m_t(m_t)$
- ABMP16new: consistent with ABMP16, but with extra iteration for DY data
- Positive correlation between α_s and $m_t(m_t)$ reduced with $t\bar{t}$ differential data



Other $t\bar{t}$ differential data in ABMP16 PDF fit

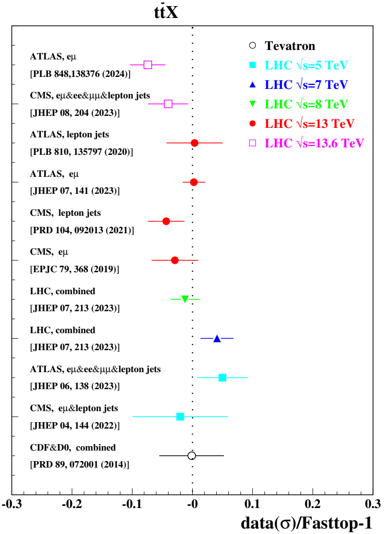


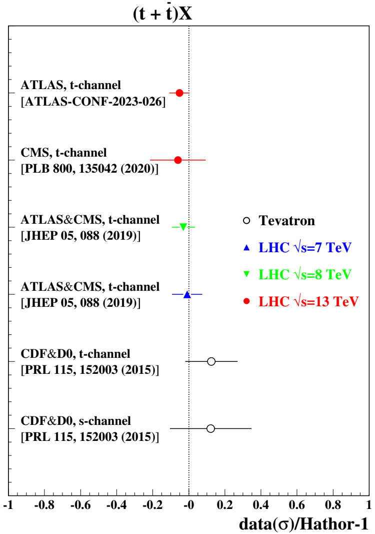




→ all data are in good agreement with NNLO theoretical predictions and put significant constraints on the PDFs

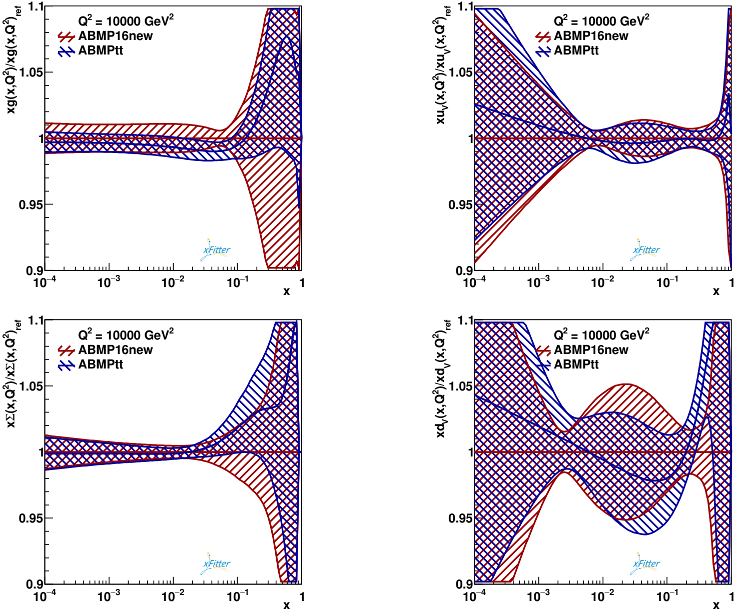
$\sigma(t\bar{t})$ and single t in ABMP16 PDF fit



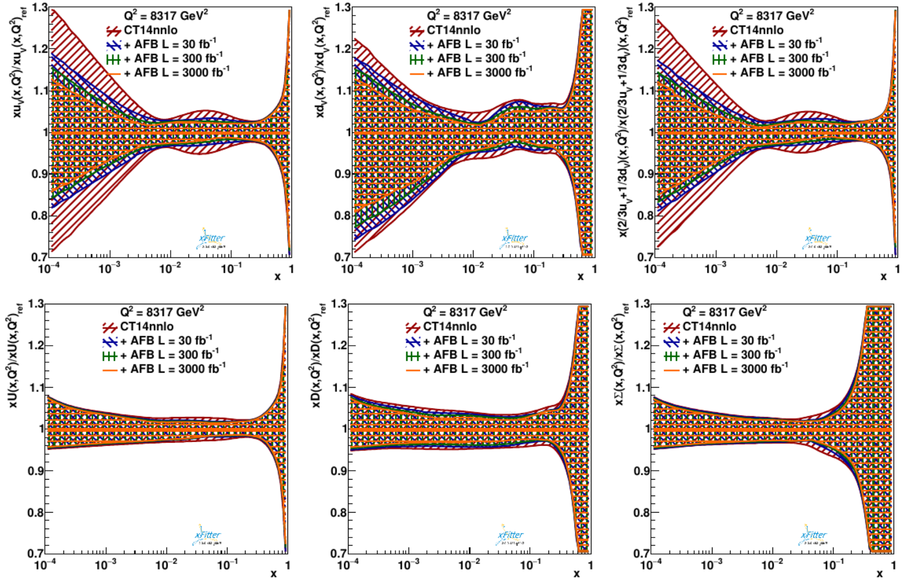


- good description of total $t\bar{t}$ and single t data
- single t data provide additional constraint on m_t (less correlated with g and α_S)

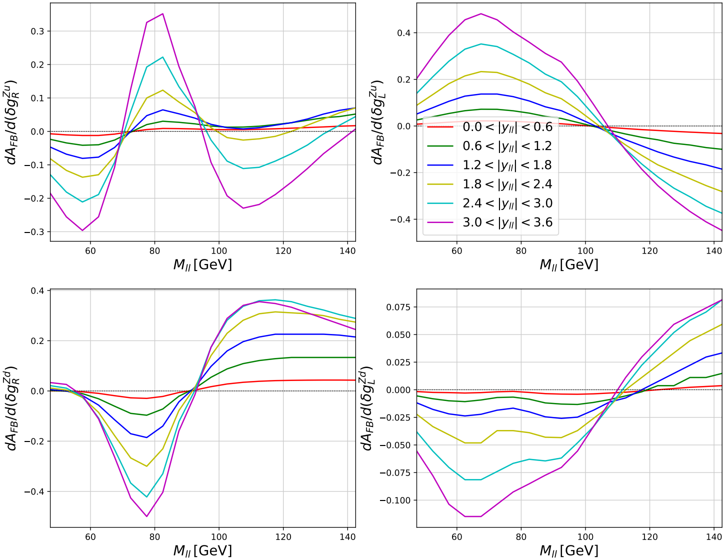
ABMPtt PDFs at $\mu = m_t$



DY AFB: PDF constraints at LHC and HL-LHC [JHEP 10 (2019) 176]

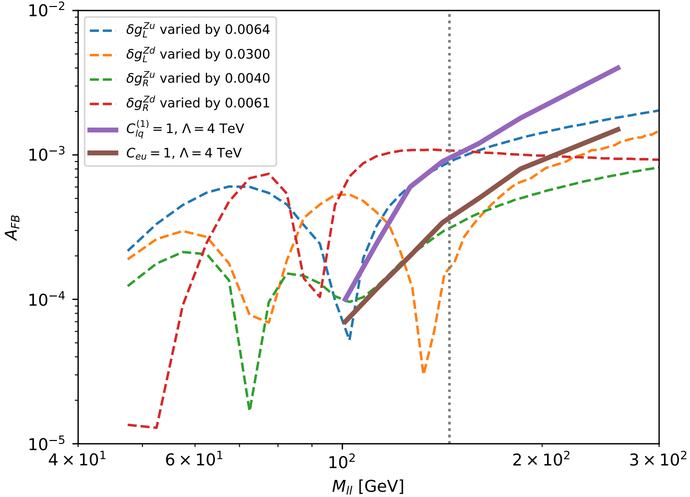


DY AFB derivatives w.r.t the couplings as a function of $M(ll)$ [LO]



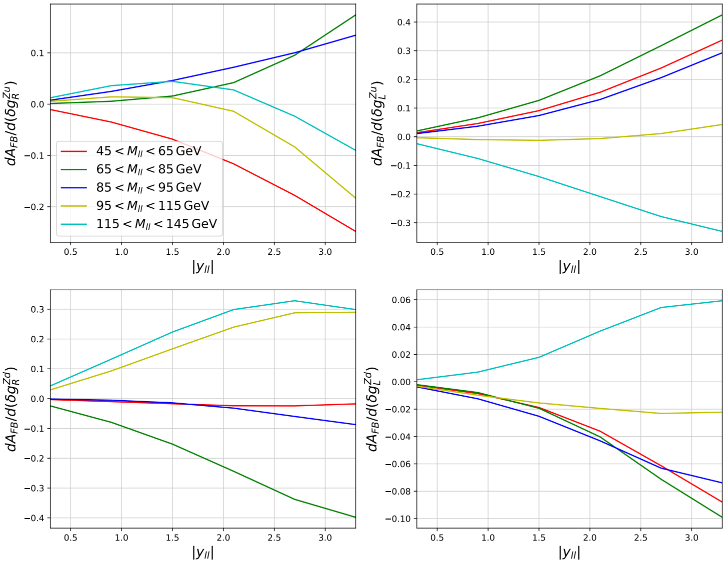
- Sensitivity to the couplings comes from AFB as a function of $M(ll)$

Impact of four-fermion operators: removed by $M(ll) < 150$ GeV



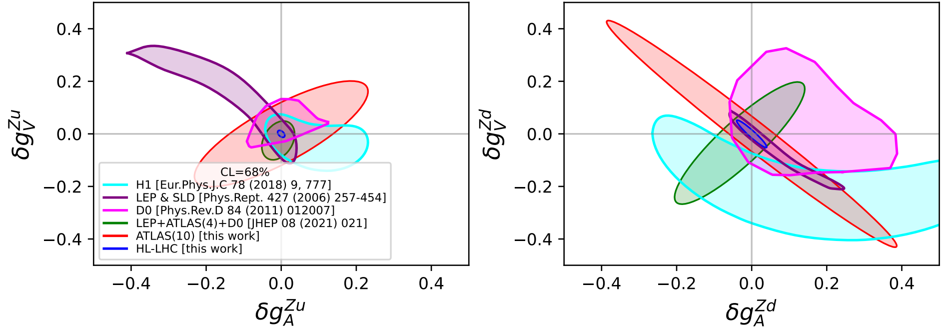
R. Boughezal, Y. Huang and F. Petriello, “Impact of high invariant-mass Drell-Yan forward-backward asymmetry measurements on SMEFT fits”, Phys. Rev. D 108 (2023) 076008 [arXiv:2303.08257].

DY AFB derivatives w.r.t the couplings as a function of $y(ll)$ [LO]



- AFB $\rightarrow$ 0 as $y(ll) \rightarrow 0$ due to its definition at the LHC (w.r.t the longitudinal boost of ll)
- Best sensitivity comes from largest reachable $y(ll)$ values: limited by detector acceptance

Results for axial and vector couplings



- All results were obtained also for axial and vector couplings:

$$g_V^{Zu} = g_R^{Zu} + g_L^{Zu} \,, \quad g_A^{Zu} = g_R^{Zu} - g_L^{Zu} \,,$$
$$g_V^{Zd} = g_R^{Zd} + g_L^{Zd} \,, \quad g_A^{Zd} = g_R^{Zd} - g_L^{Zd} \,.$$