Parton distribution functions with percent level precision

Third HiggsTools Annual Meeting, Turin

Zahari Kassabov

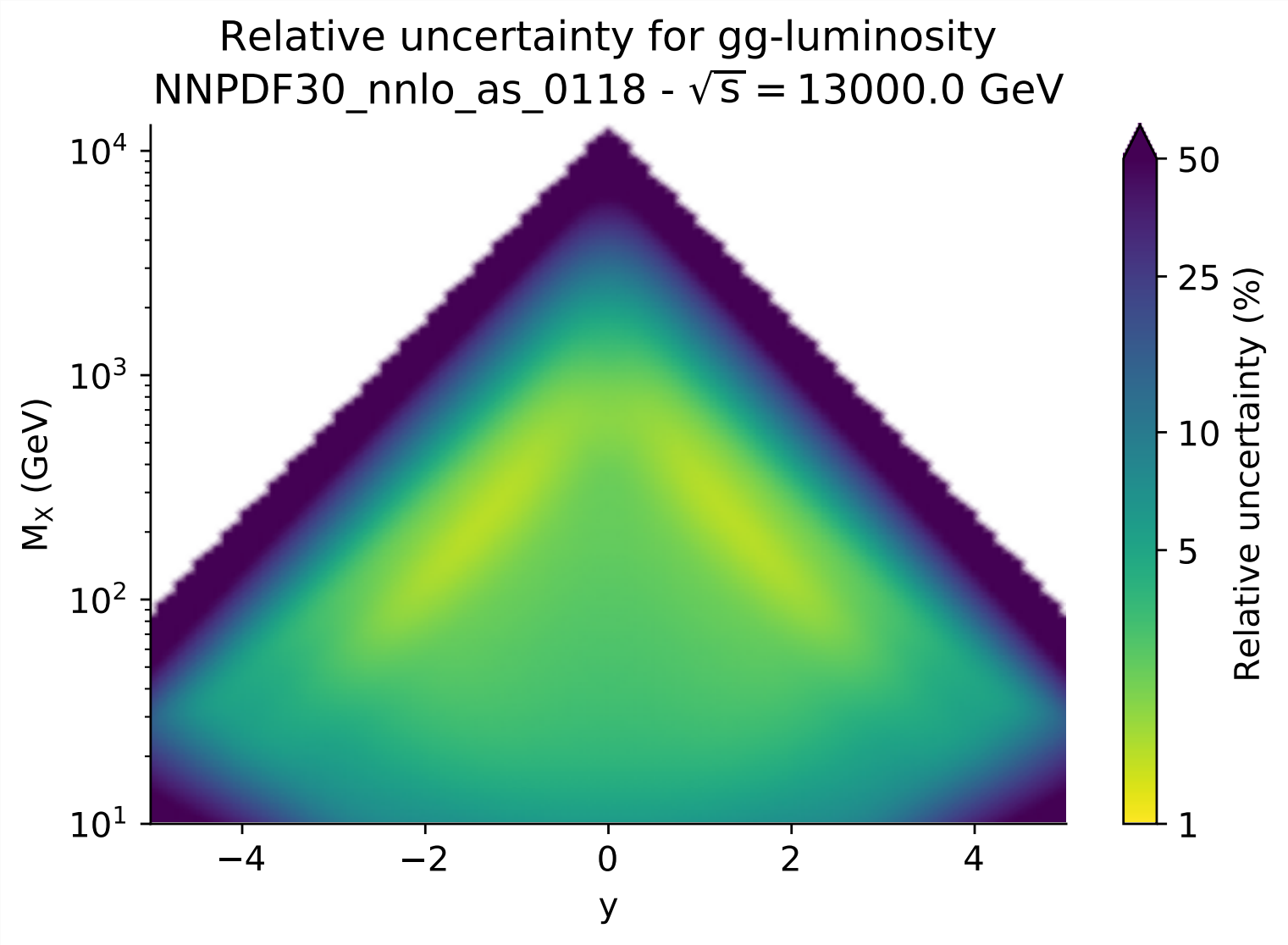
May 17th 2017

University of Turin, University of Milan

UNIVERSITA
DEGLI STUDI
DI TORINO

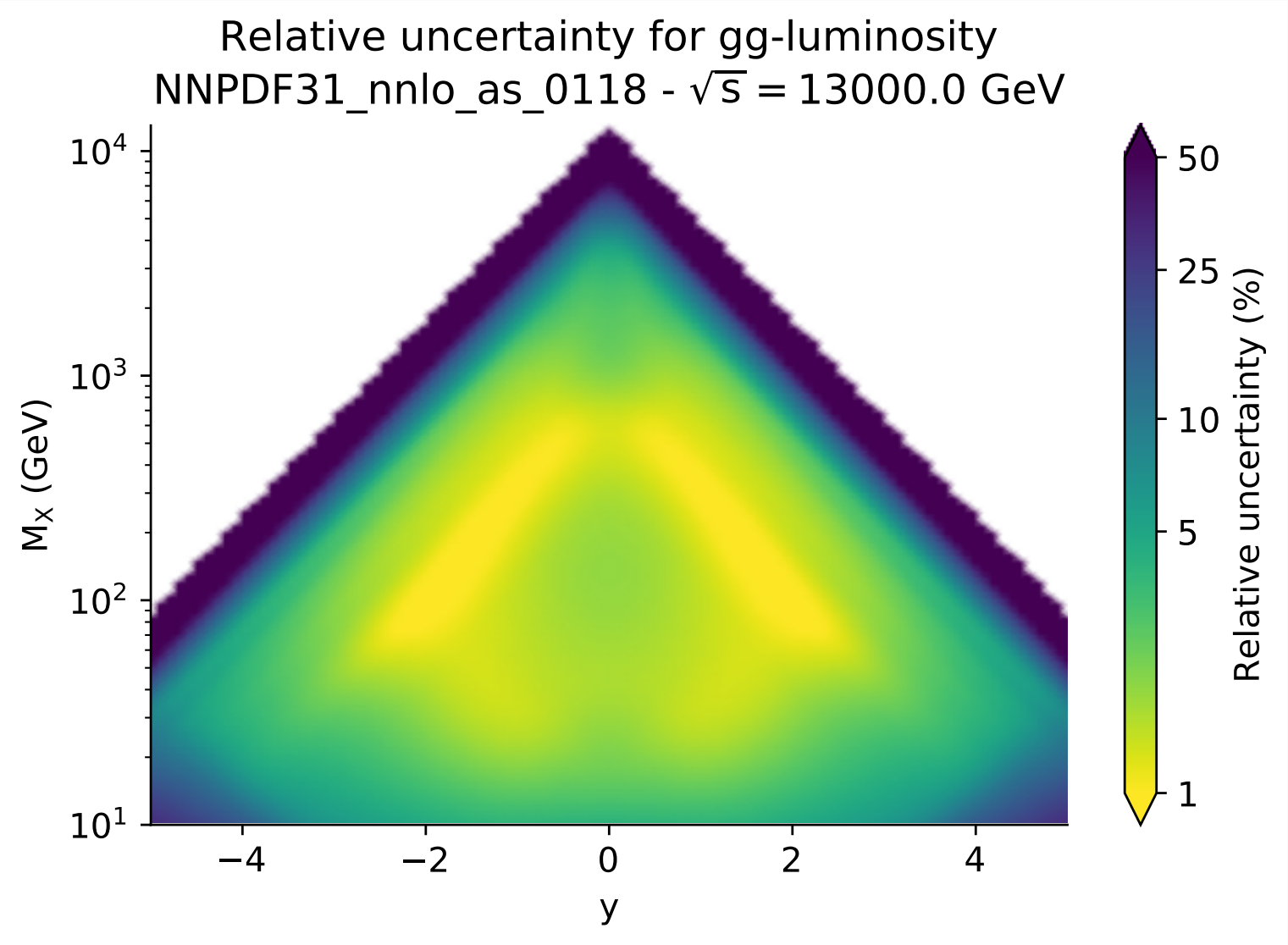
Turning the lights on...

NNPDF 3.0 (2014)



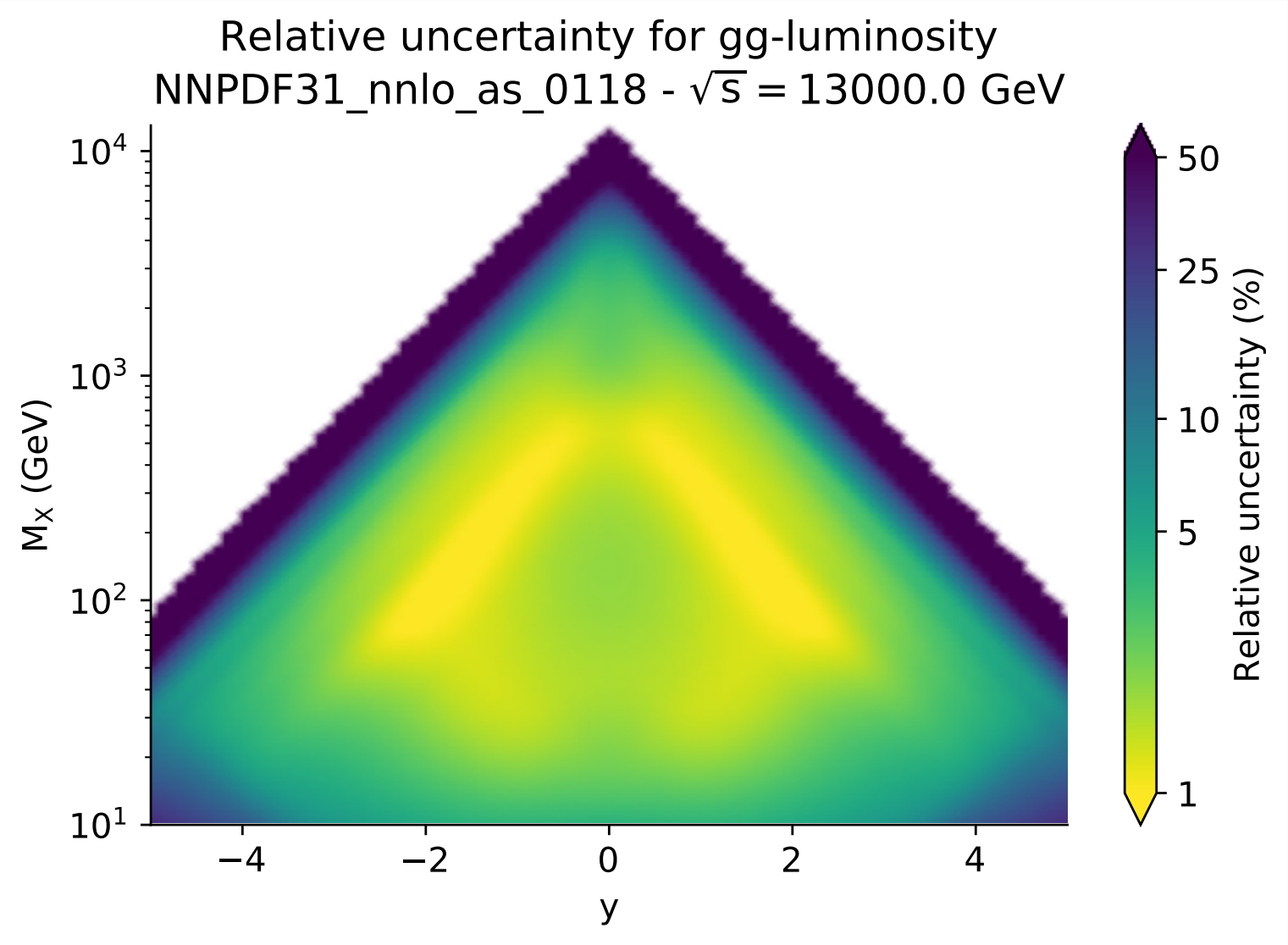
Turning the lights on...

NNPDF 3.1 (2017)



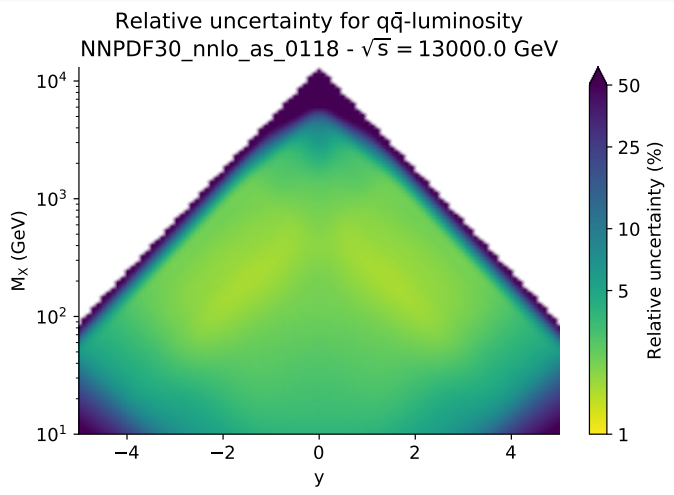
Turning the lights on...

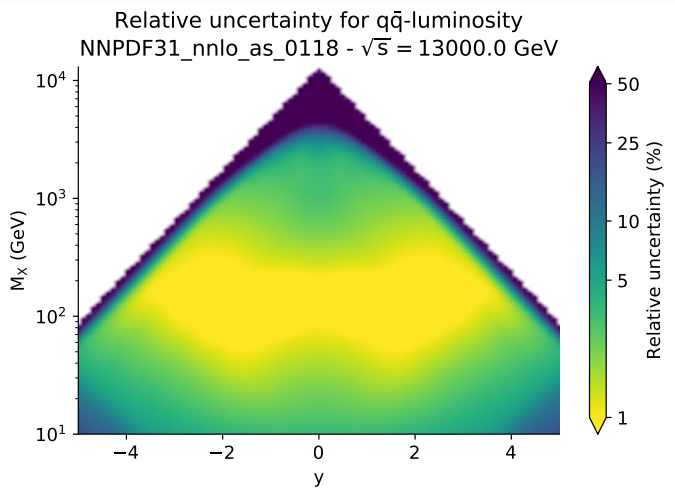
NNPDF 3.1 (2017) Available on LHAPDF!



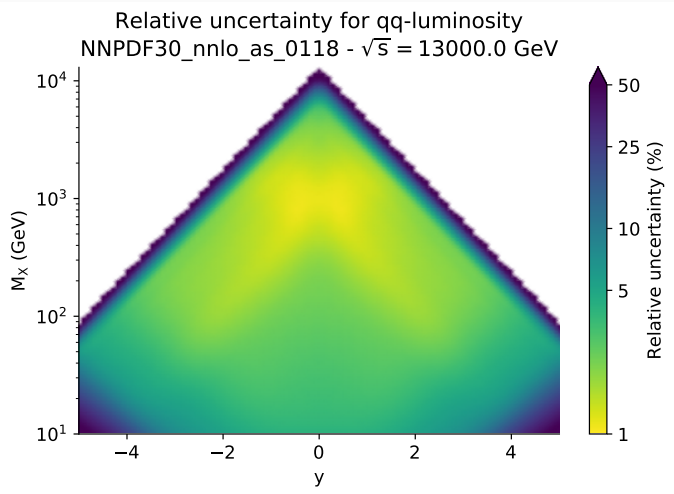
Similar situation for other channels

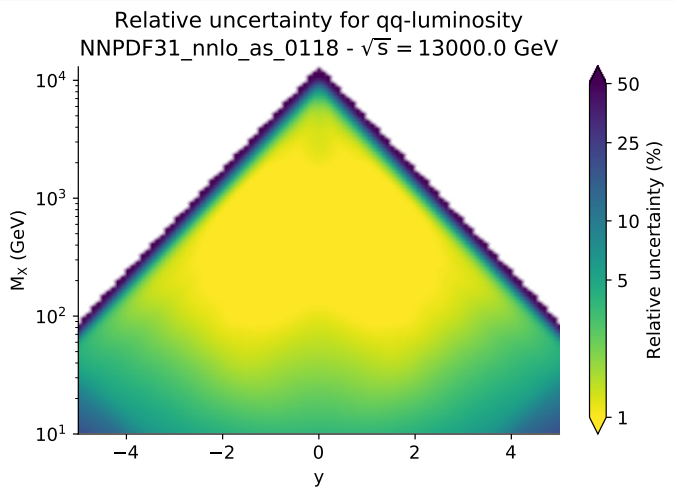
$q\bar{q}$





qq





Percent level PDF uncertainty

Confidently claiming improvements in PDF uncertainties requires:

- Wealth of *precise and accurate* experimental data.
- Corresponding theory calculations.
- Control over theoretical uncertainties (α_s,m_c).
- Control over uncertainties related to fitting methodology (*cross validation, closure testing*).

However:

- Improvements in one of these areas increases the relative importance of the others.
- Many non obvious problems appear that previously could be overlooked.

This talk

- What's new in NNPDF 3.1.
- Challenges associated to the increased precision.
- Future directions.

Fitting the charm PDF

Idea Do not assume that the charm PDF is perturbatively generated, but fit it.

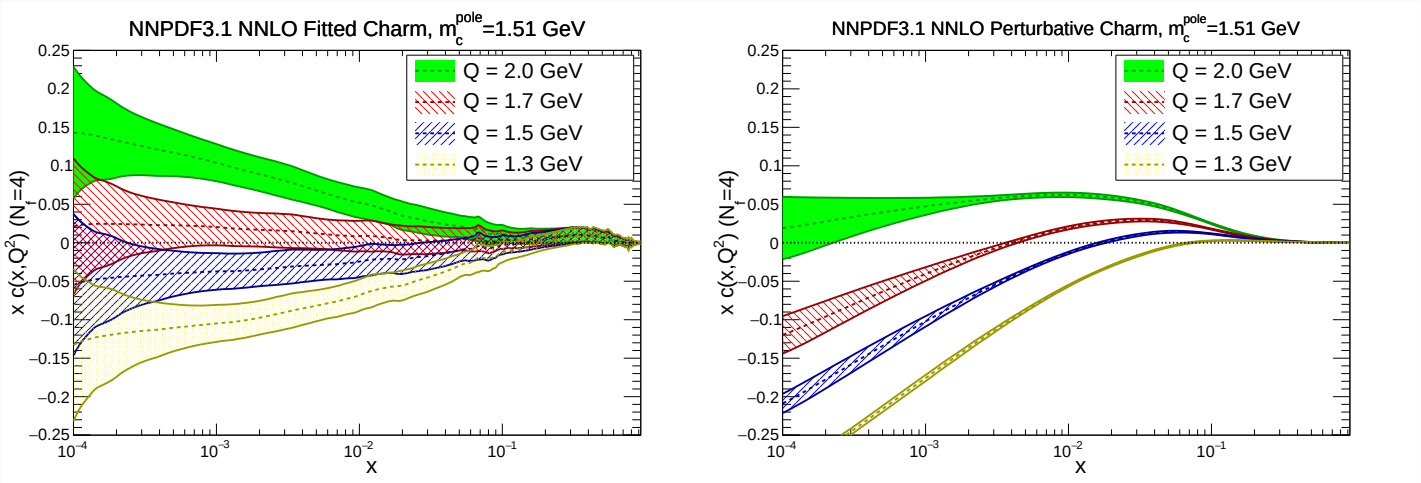
Main Effect Eliminate possible source of bias. Reduce the dependency on the charm mass.

Required Extend FONLL to deal with charm initiated contributions. [Ball, Bonvini, Rottoli, JHEP 15, Ball et all PLB]

Scale dependence of the charm PDF

Compared to a purely perturbative charm PDF,

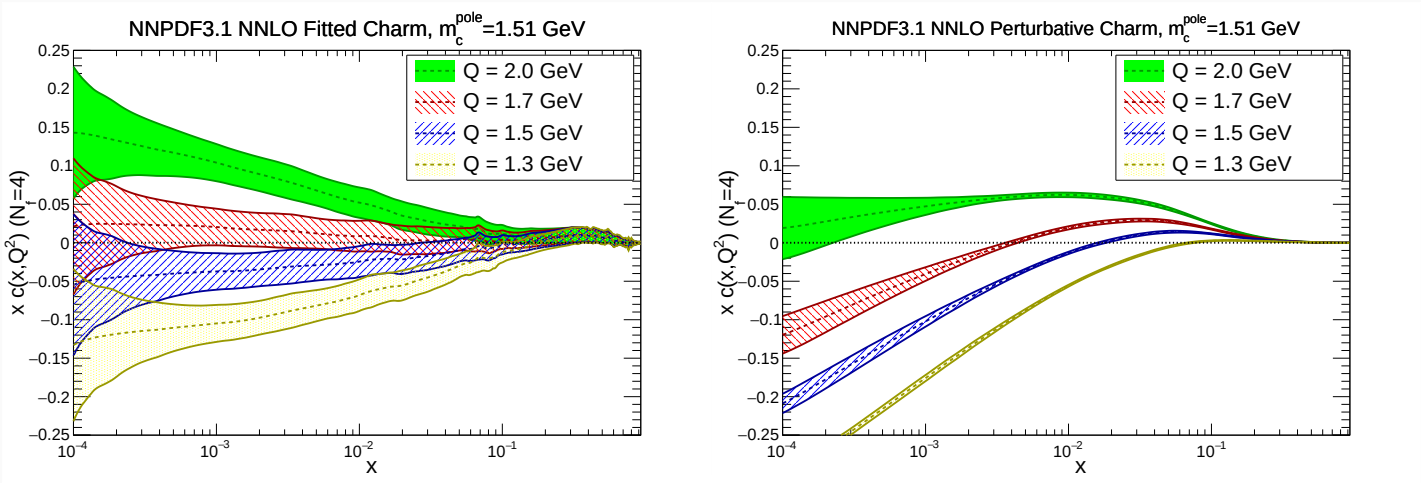
- Reduced scale dependence + charm vanishing at $Q^2 \gtrsim m_c$
  $\rightarrow$Increased strangeness at high $Q^2 \rightarrow$ Better fit to the LHC data.

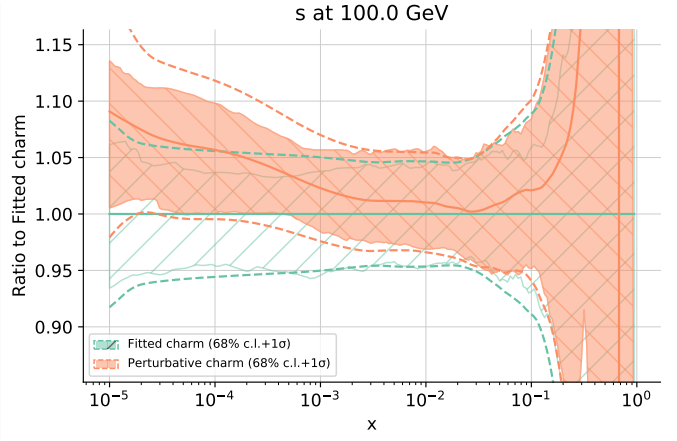


Scale dependence of the charm PDF

Compared to a purely perturbative charm PDF,

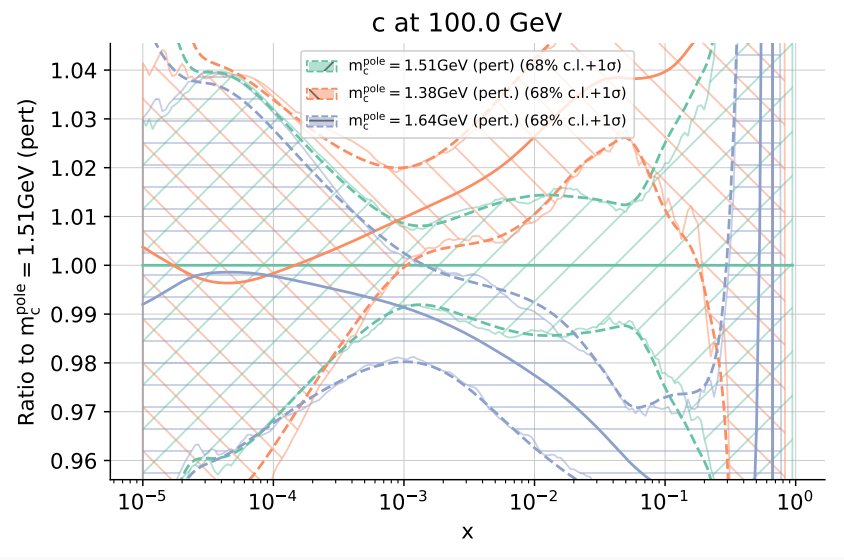
- Reduced scale dependence + charm vanishing at $Q^2 \gtrsim m_c$
  $\rightarrow$Increased strangeness at high $Q^2 \rightarrow$ Better fit to the LHC data.

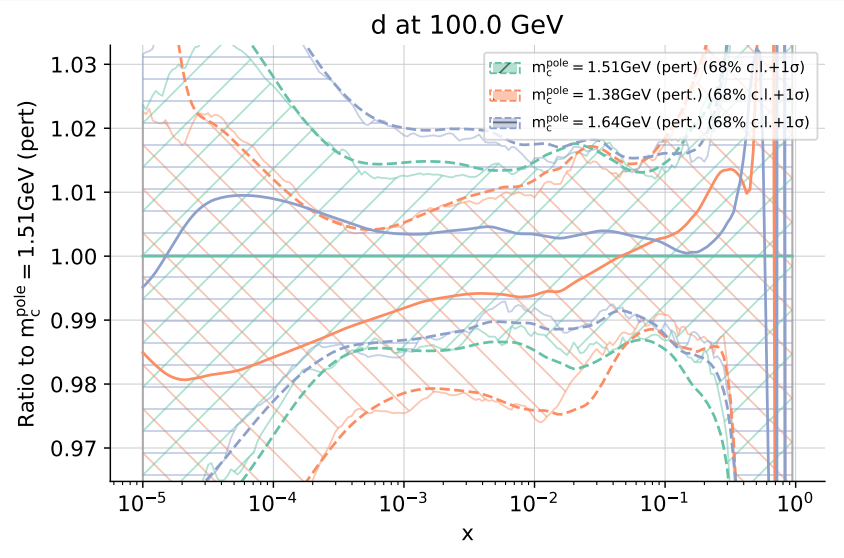




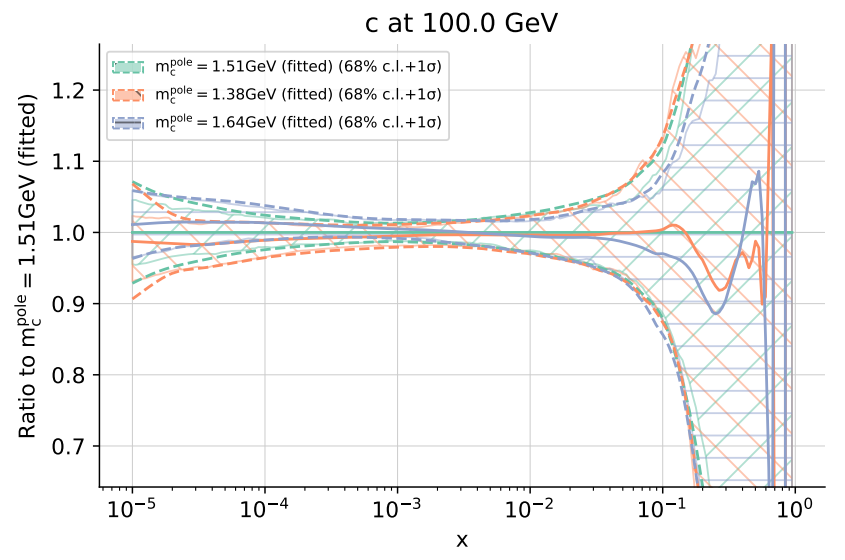
PDF dependence on charm mass

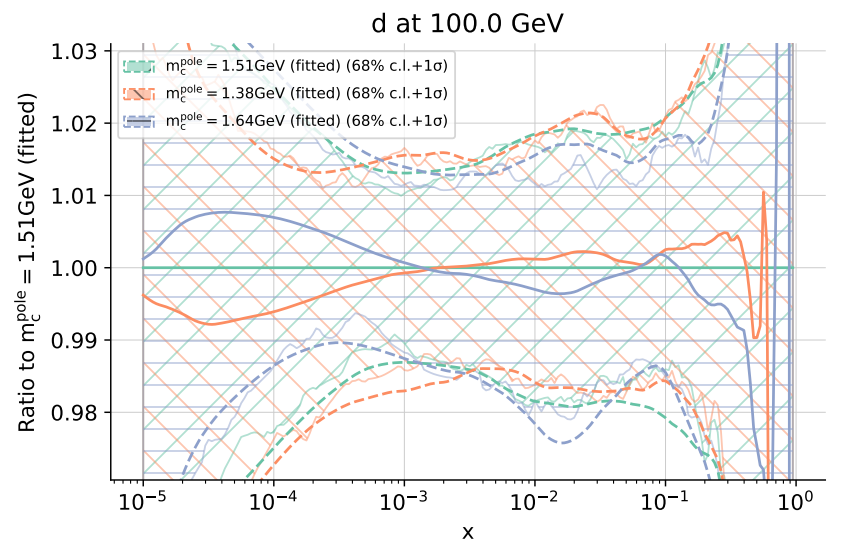
Perturbative charm





Fitted charm

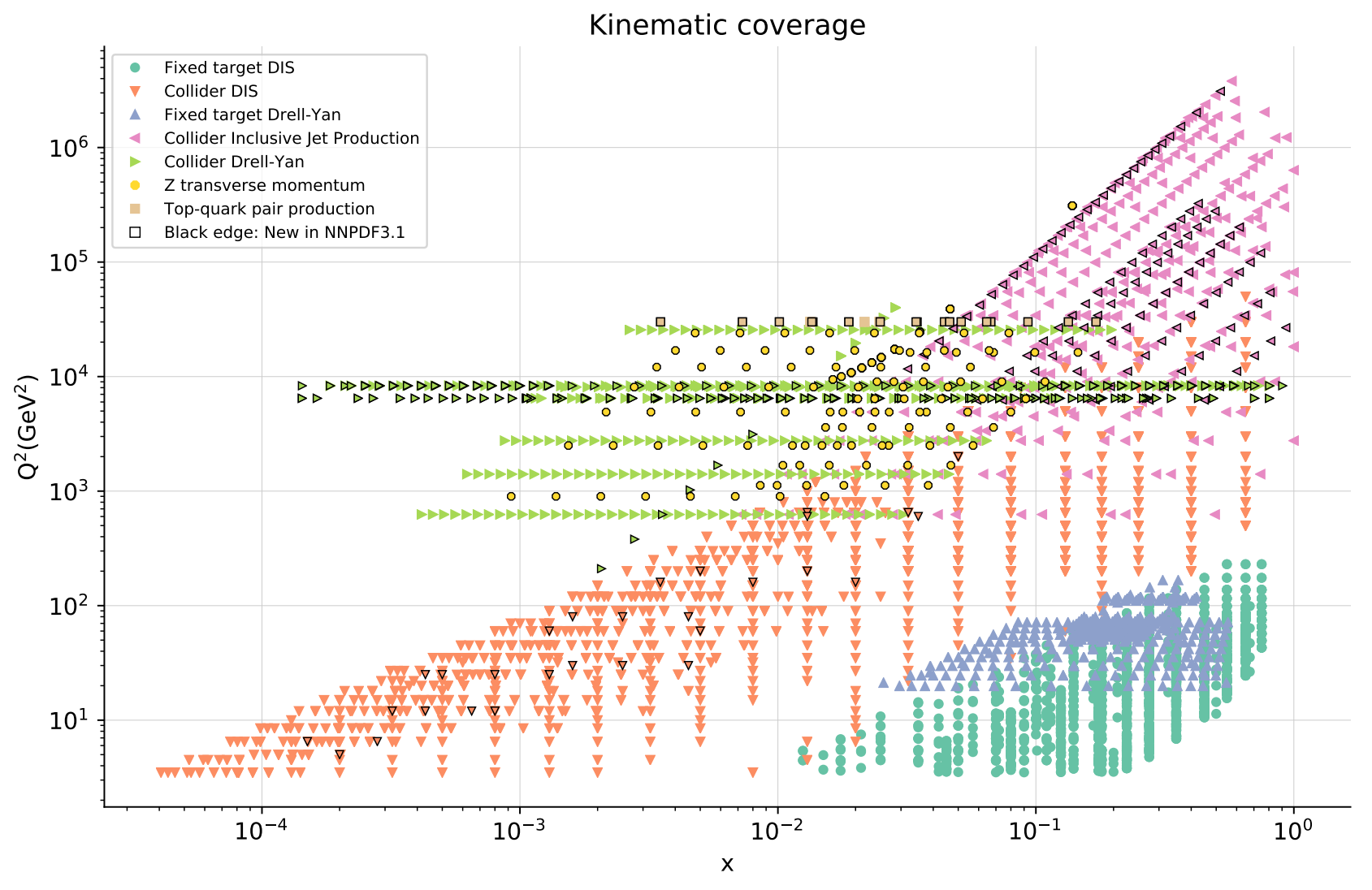




New data in NNPDF 3.1

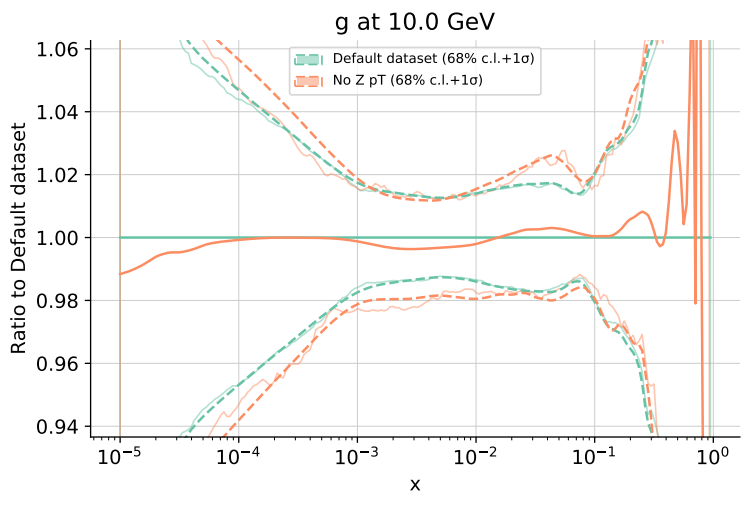
Combined HERA data	Run I+II
D0 assymetries	Run II
ATLAS W,Z rapidity 7 TeV	2011
ATLAS inclusive jets	2011
ATLAS low Mass DY 7 TeV	2010-2011
ATLAS, CMS Z p_T 8 TeV	2012
ATLAS and CMS tt differential 8 TeV	2012
CMS 2.76 TeV Jets	2012
LHCb W,Z rapidity 7 and 8 TeV	2011-2012

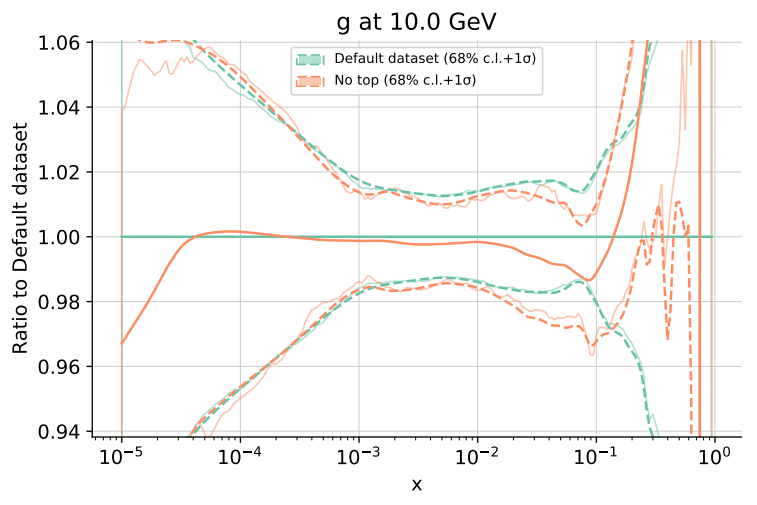
Kinematic coverage (LO)

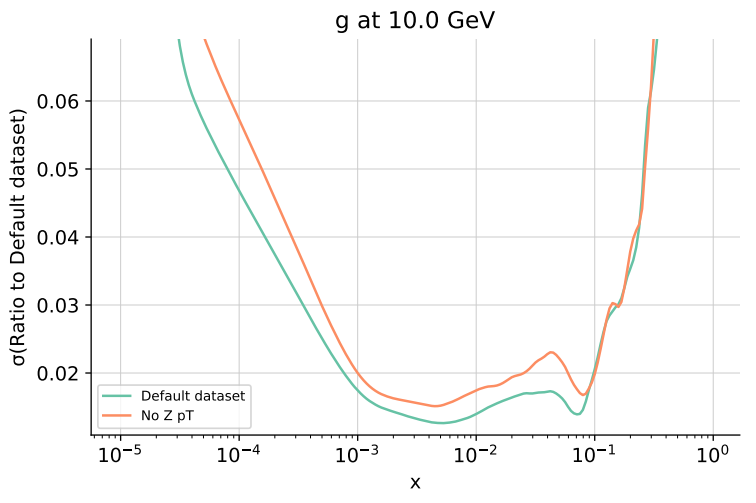


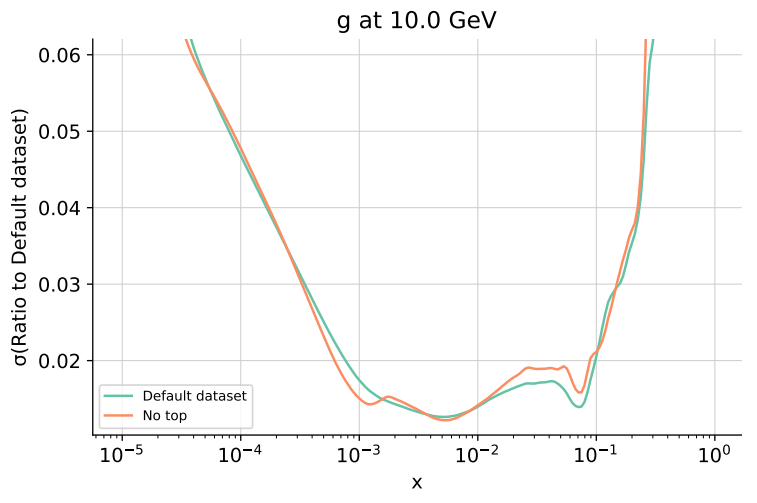
Individual effect of the new data

- New included datasets are broadly in agreement and have a moderate impact individually.



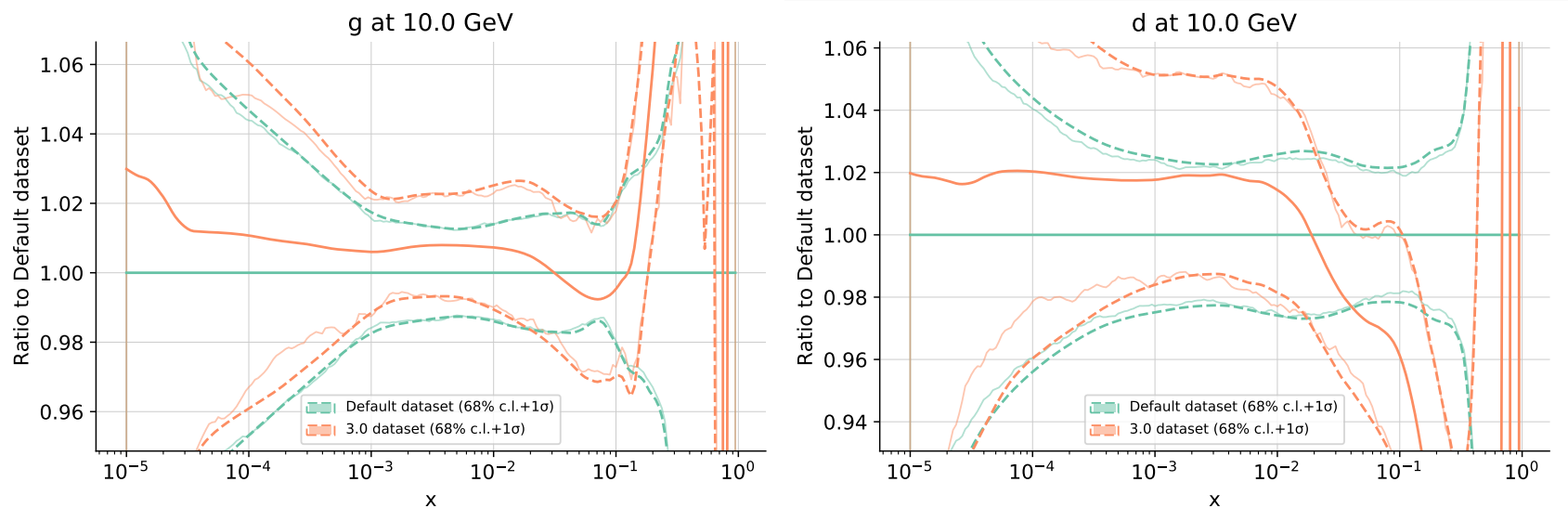


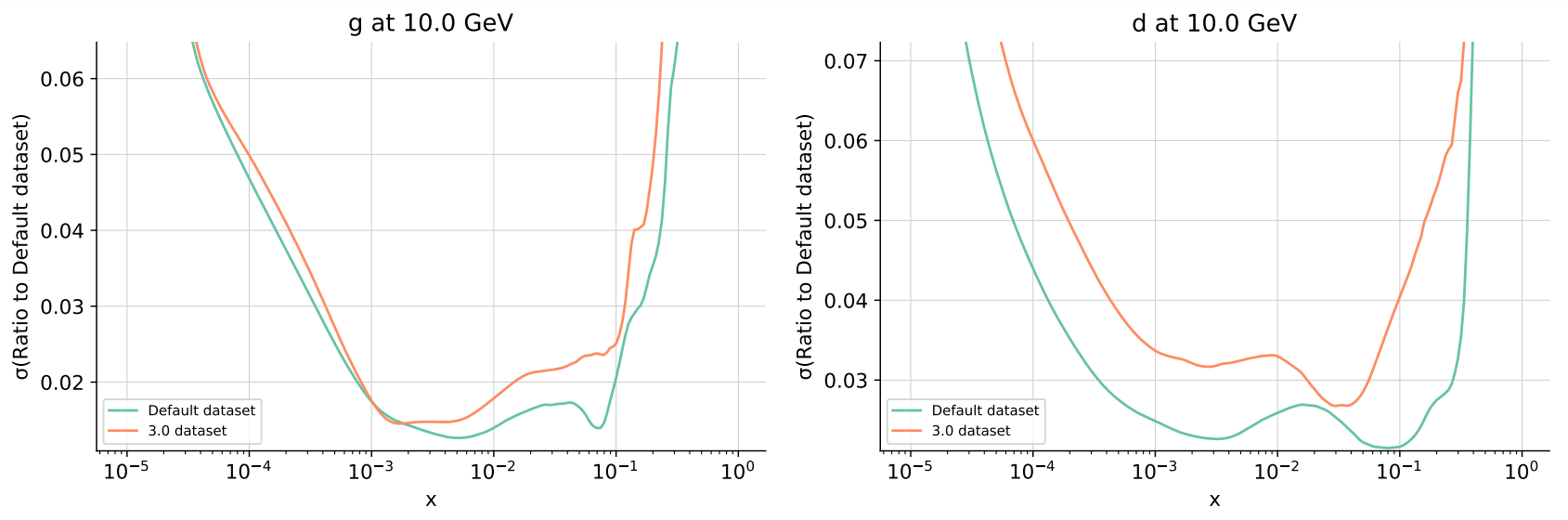




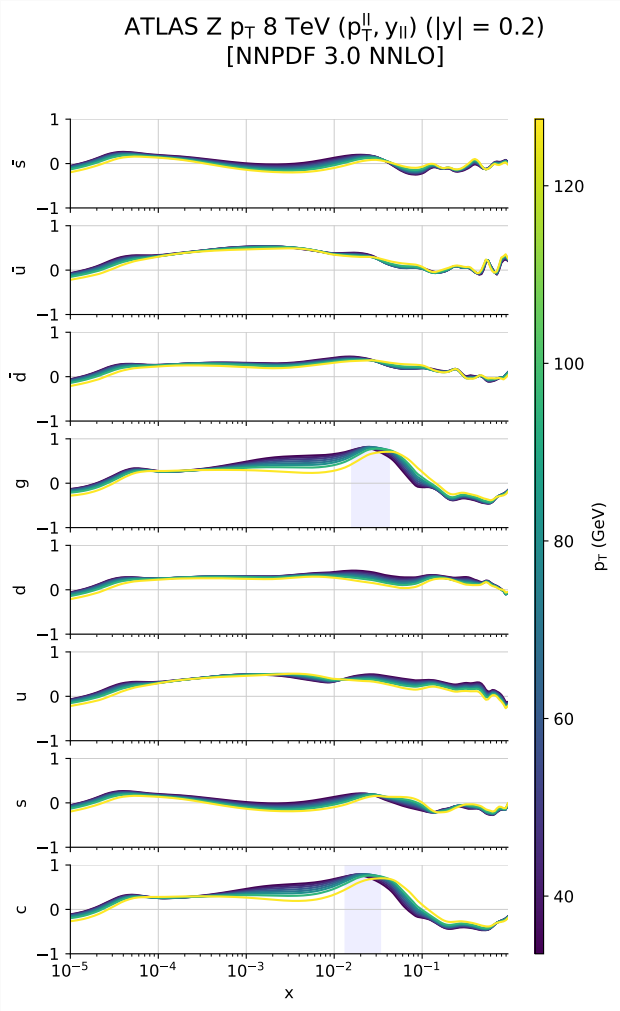
Global effect of the new data

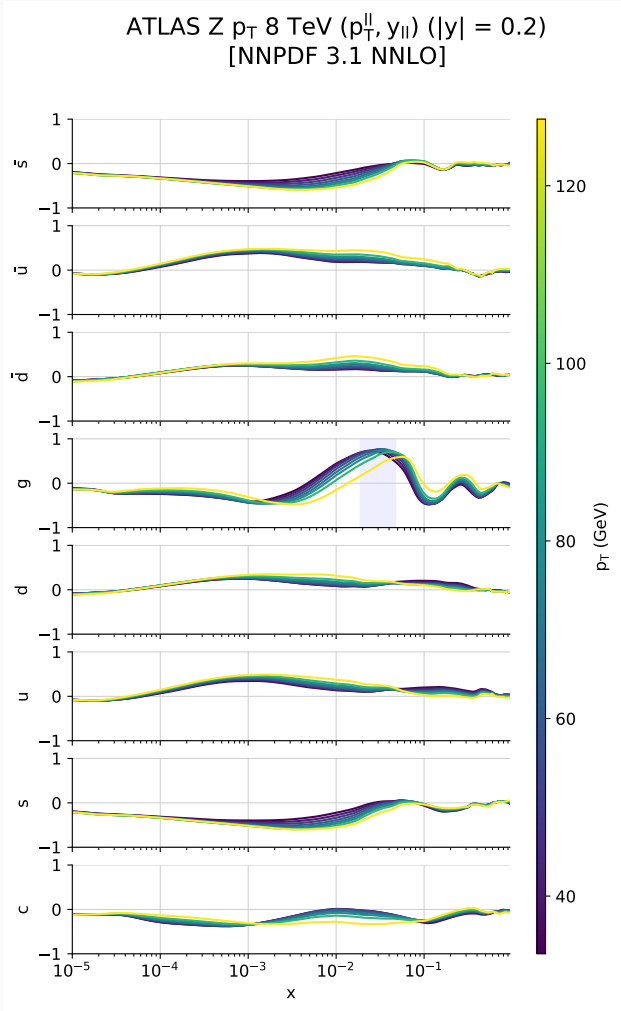
Collectively, up tp $\sim 1\sigma$ deviations, $\sim 30\%$ smaller uncertainties.





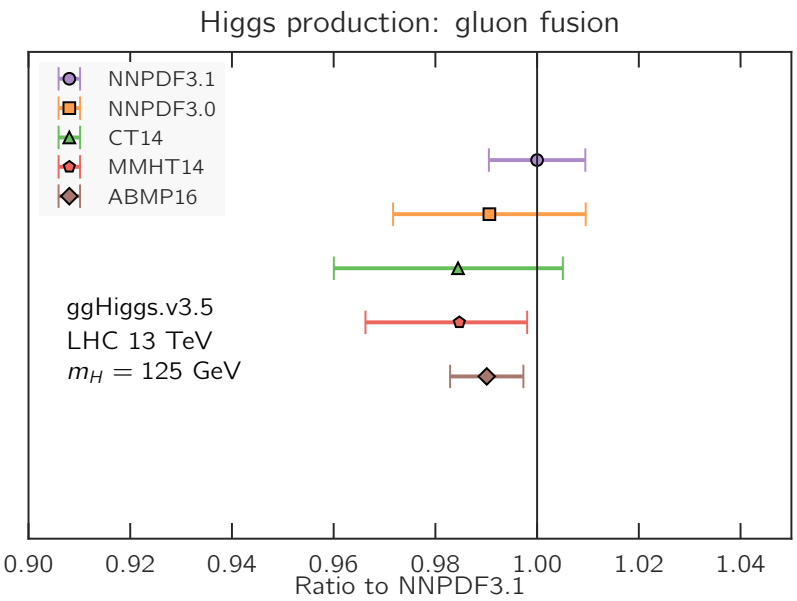
Correlations between data and PDF fluctuations

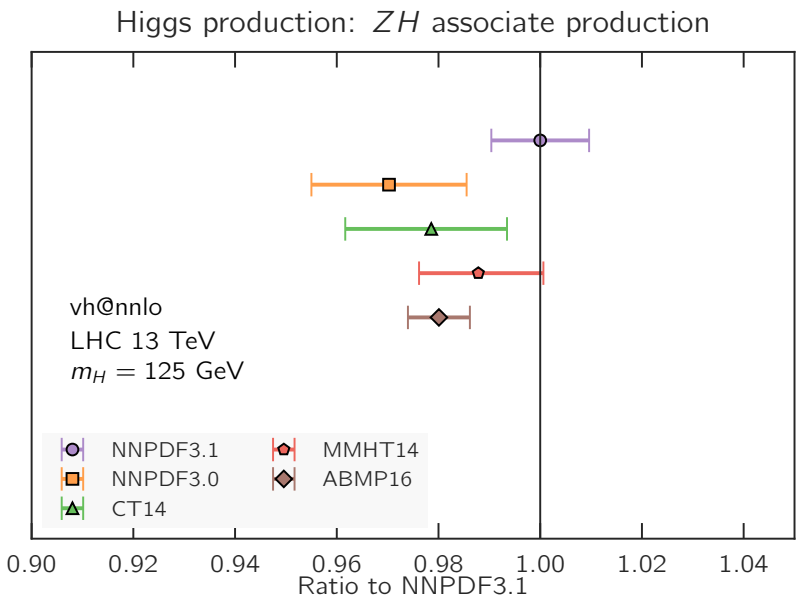


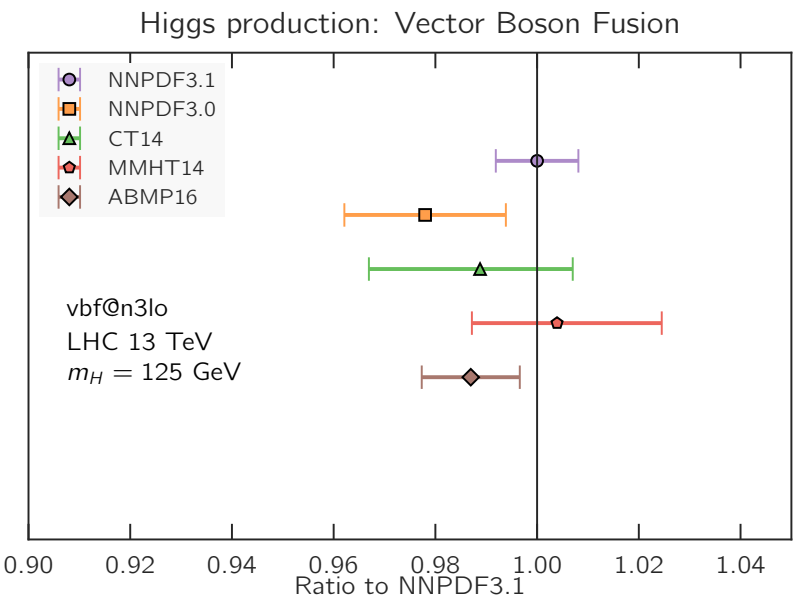


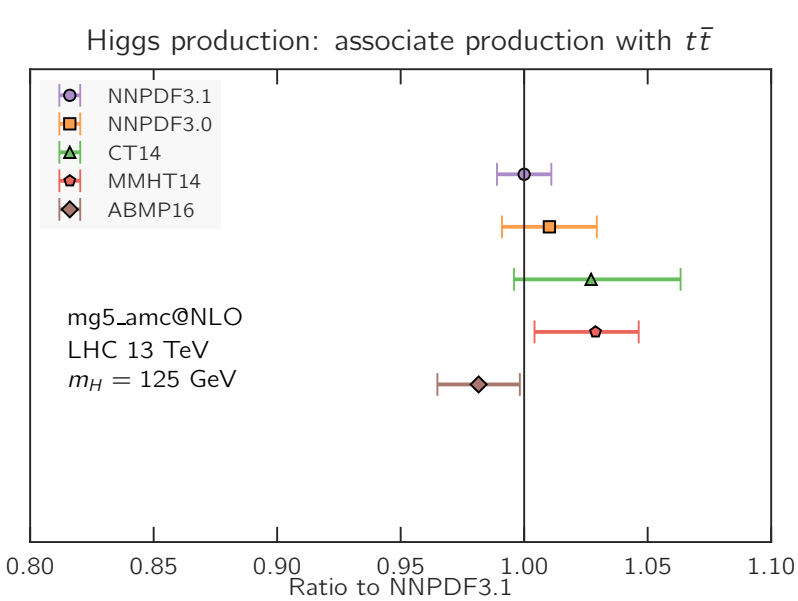
- Gluon dependence more localized.
- Charm less important even with increased fluctuations.
- Constraining strangeness becoming more important.

Impact on the Higgs Cross section

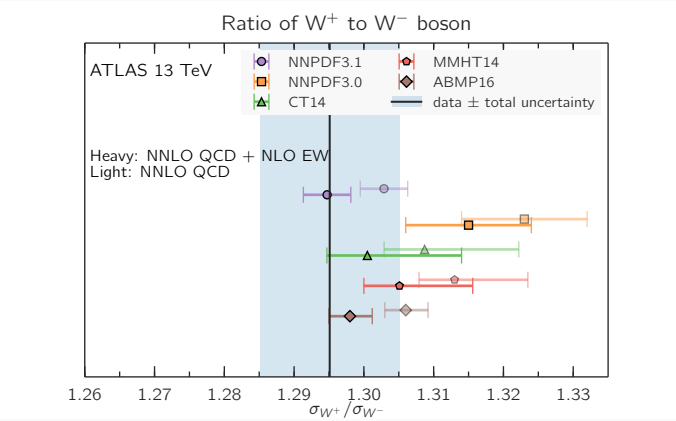


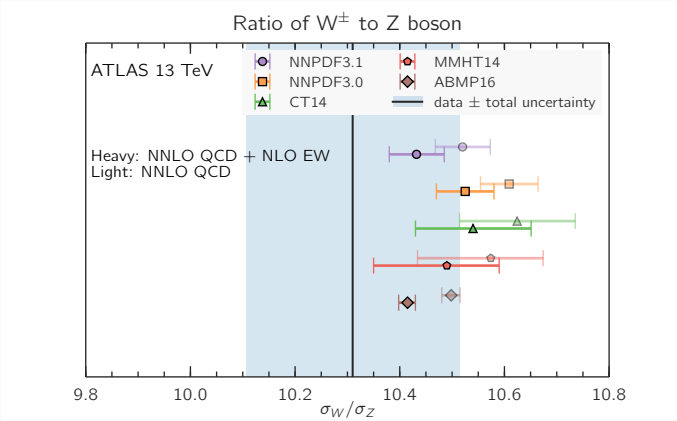


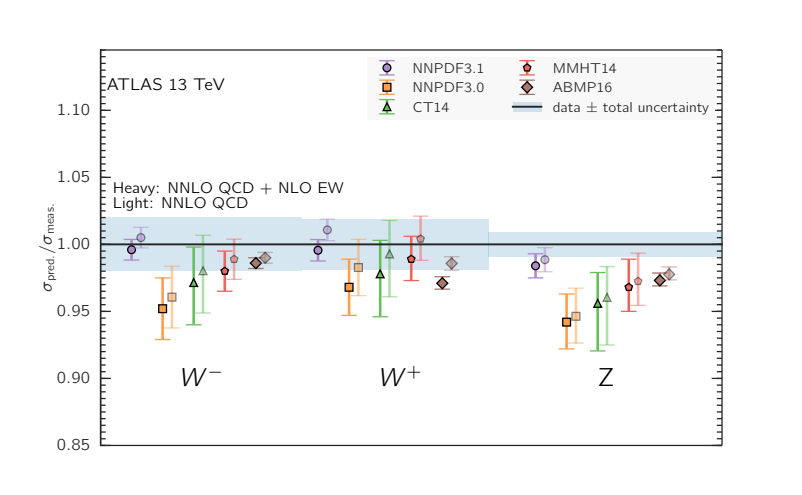




Impact on the W and Z cross sections



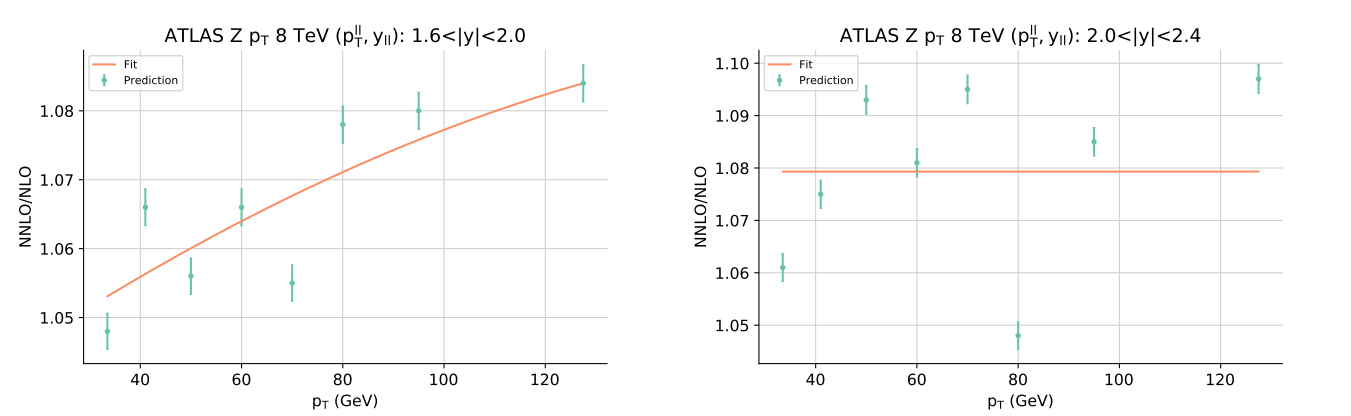




Note:

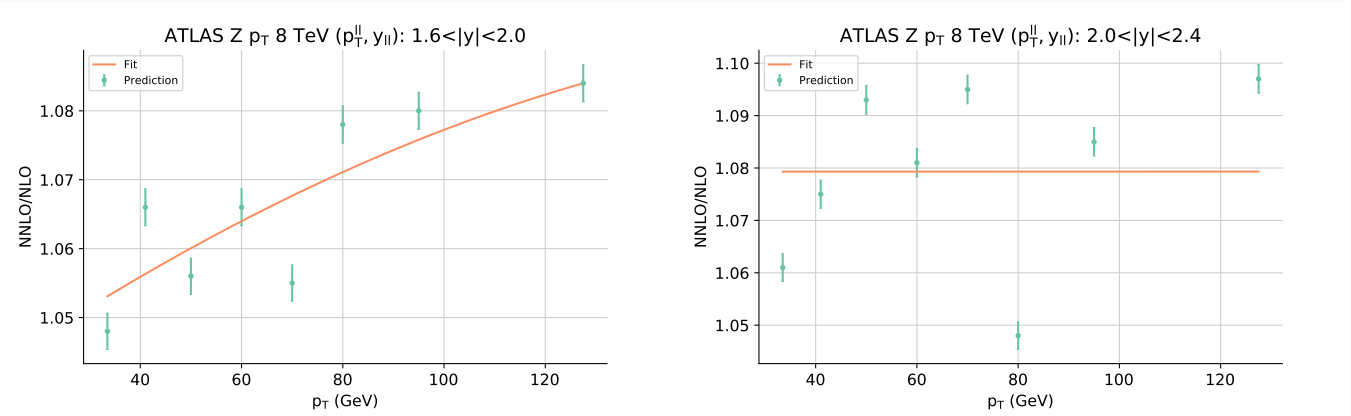
- Effect of EW corrections.
- Better agreement thanks to charm suppression.

Including the *Z pT* at NNLO



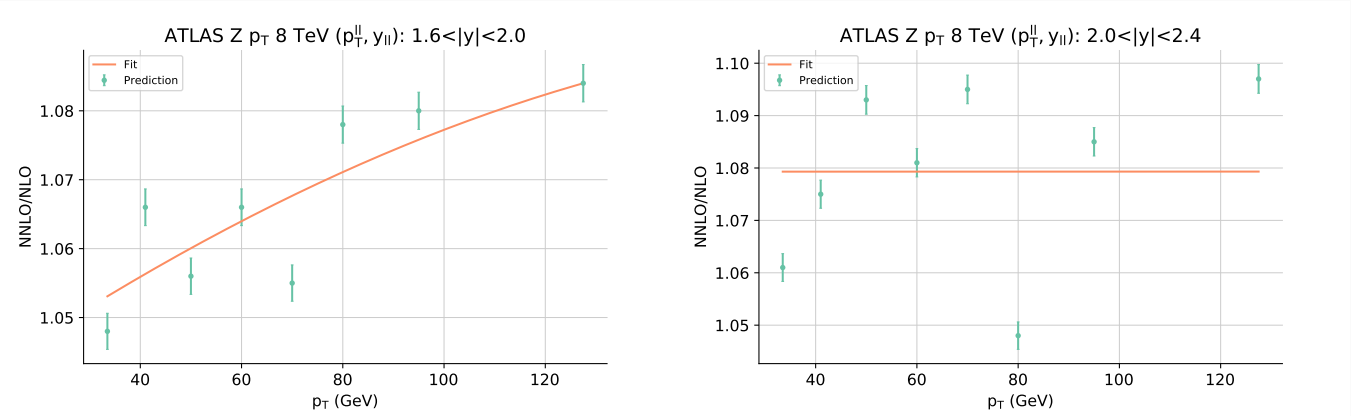
- We expect the NNLO/NLO [Boughezal et al, PRL 2016] k-factor to be a *smooth* function of *pT*.

Including the *Z pT* at NNLO



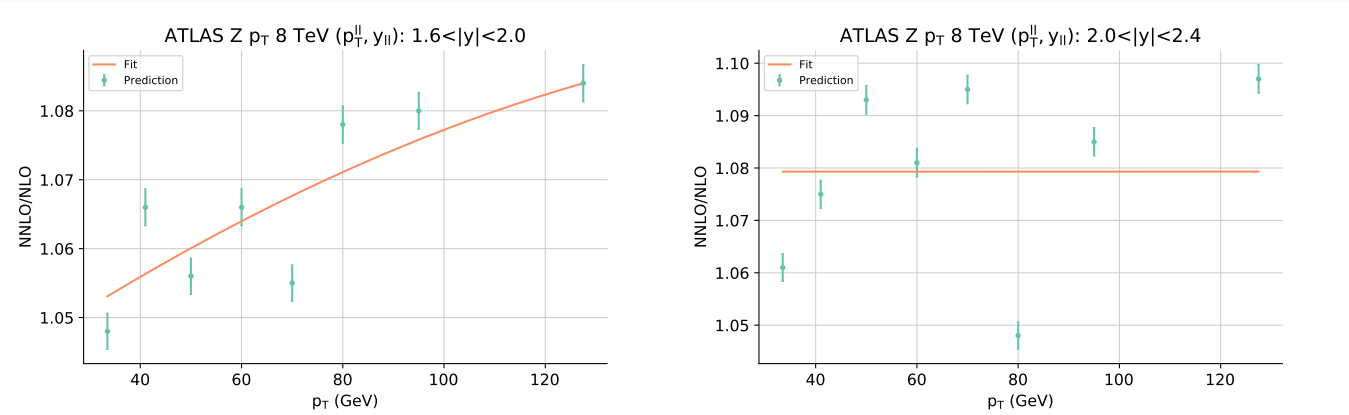
- We expect the NNLO/NLO [Boughezal et al, PRL 2016] k-factor to be a *smooth* function of *pT*.
- MC errors too small to account for the lack of smoothness.

Including the *Z pT* at NNLO

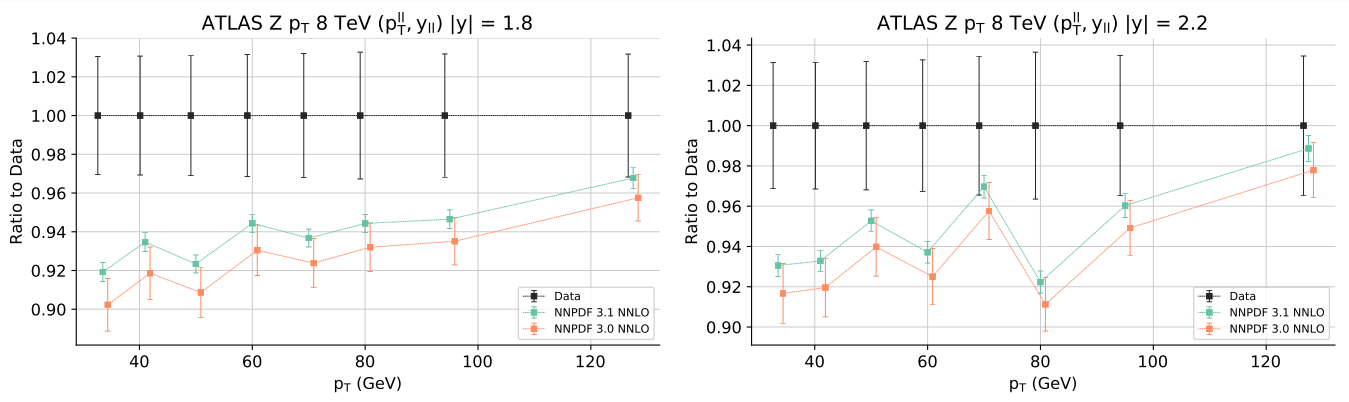


- We expect the NNLO/NLO [Boughezal et al, PRL 2016] k-factor to be a *smooth* function of *pT*.
- MC errors too small to account for the lack of smoothness.
- Data doesn't support the fluctuations either.

Including the *Z pT* at NNLO



- We expect the NNLO/NLO [Boughezal et al, PRL 2016] k-factor to be a *smooth* function of *pT*.
- MC errors too small to account for the lack of smoothness.
- Data doesn't support the fluctuations either.

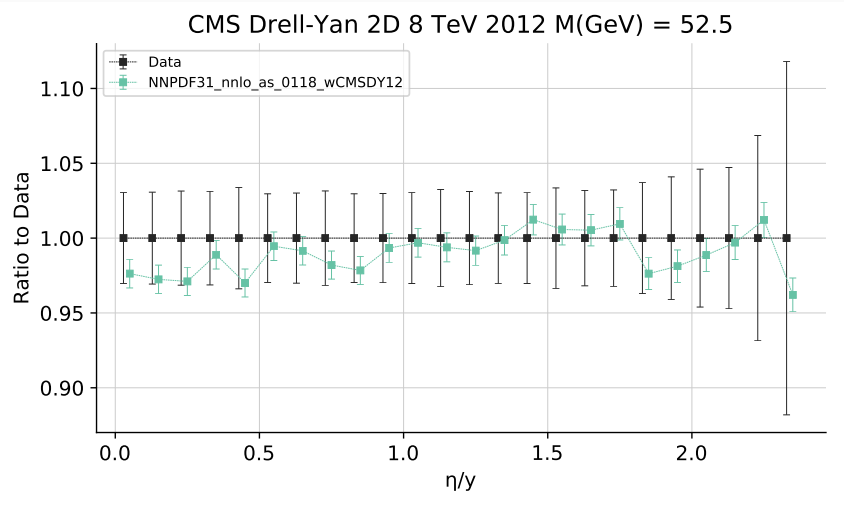


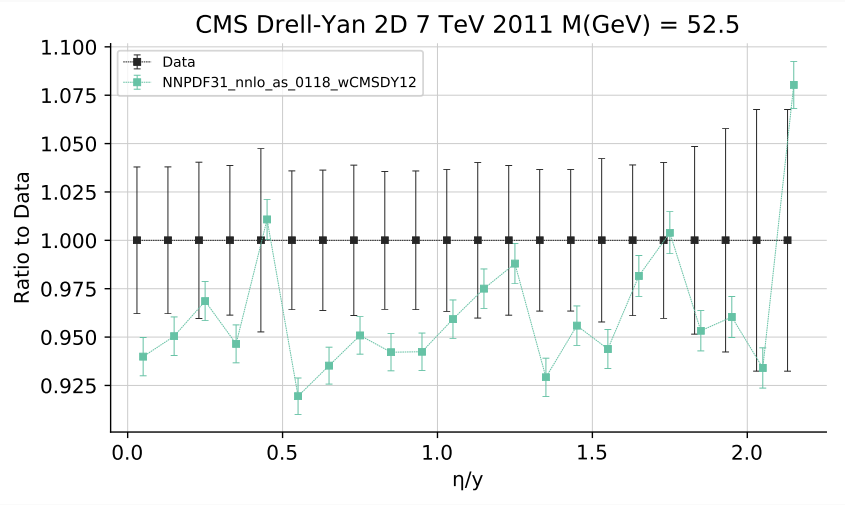
Dealing with MC errors

- The solution is to add a 1% uncorrelated error consistent with the distance to the smoothed curve.
- This improves the quality of the description while leaving the PDFs unchanged [Boughezal, Guffanti, Petriello, Ubiali 1705.00343]
- Can we do better?
- MC errors previously considered negligible compared to data. Now of the same order.

(Not) including CMS Drell-Yan 2D 2012

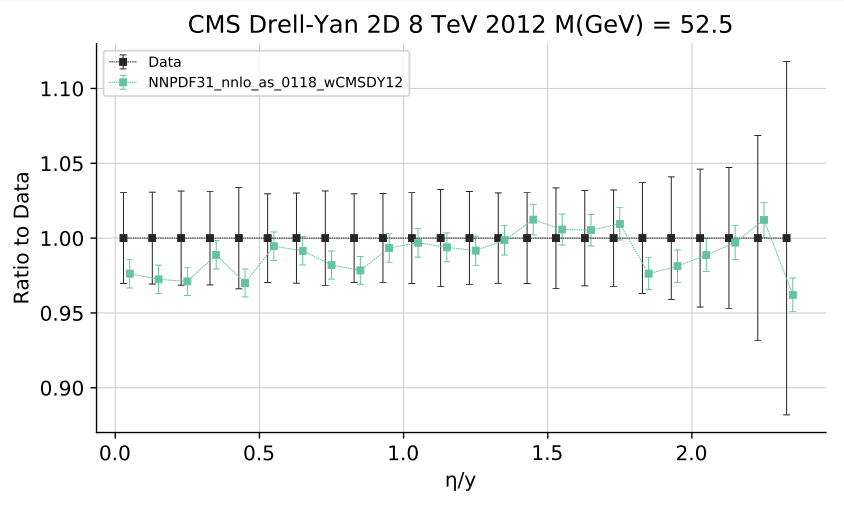
We tried to include the CMS 2012 8 TeV double differential measurement (left):

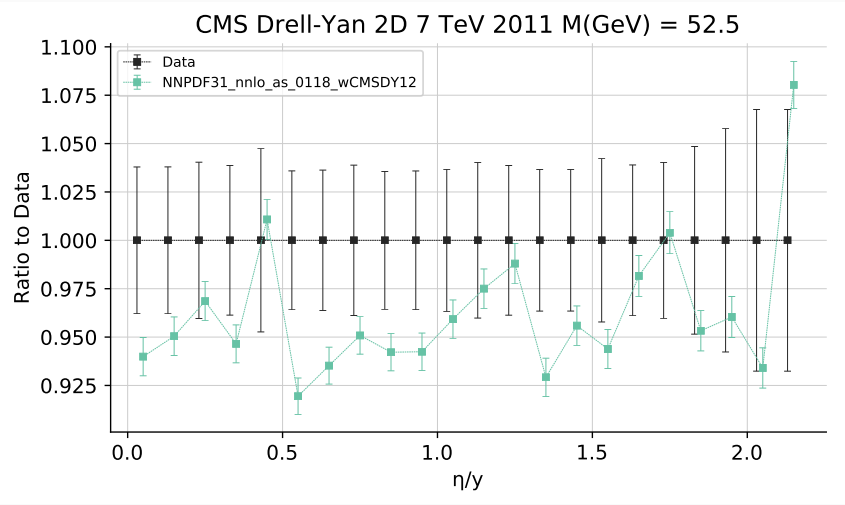




(Not) including CMS Drell-Yan 2D 2012

We tried to include the CMS 2012 8 TeV double differential measurement (left):



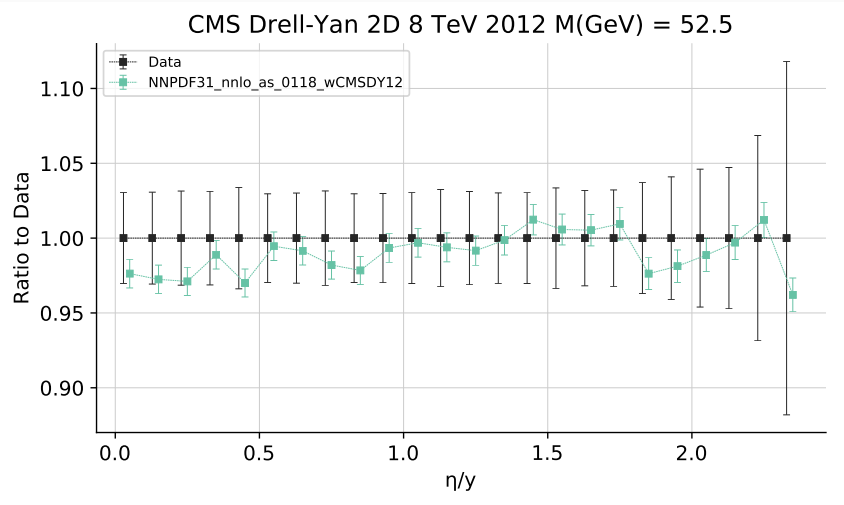


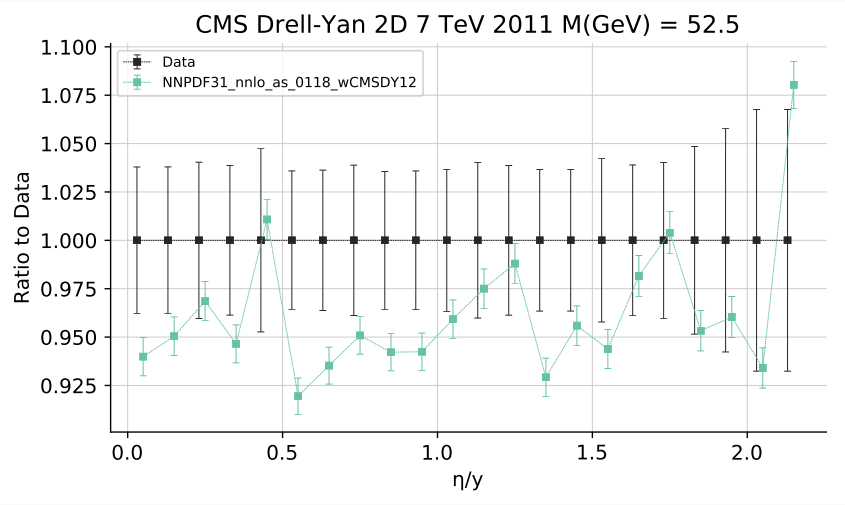
$\chi^2/N_{\text{dat}} = 2.88$

$\chi^2/N_{\text{dat}} = 1.24$

(Not) including CMS Drell-Yan 2D 2012

We tried to include the CMS 2012 8 TeV double differential measurement (left):

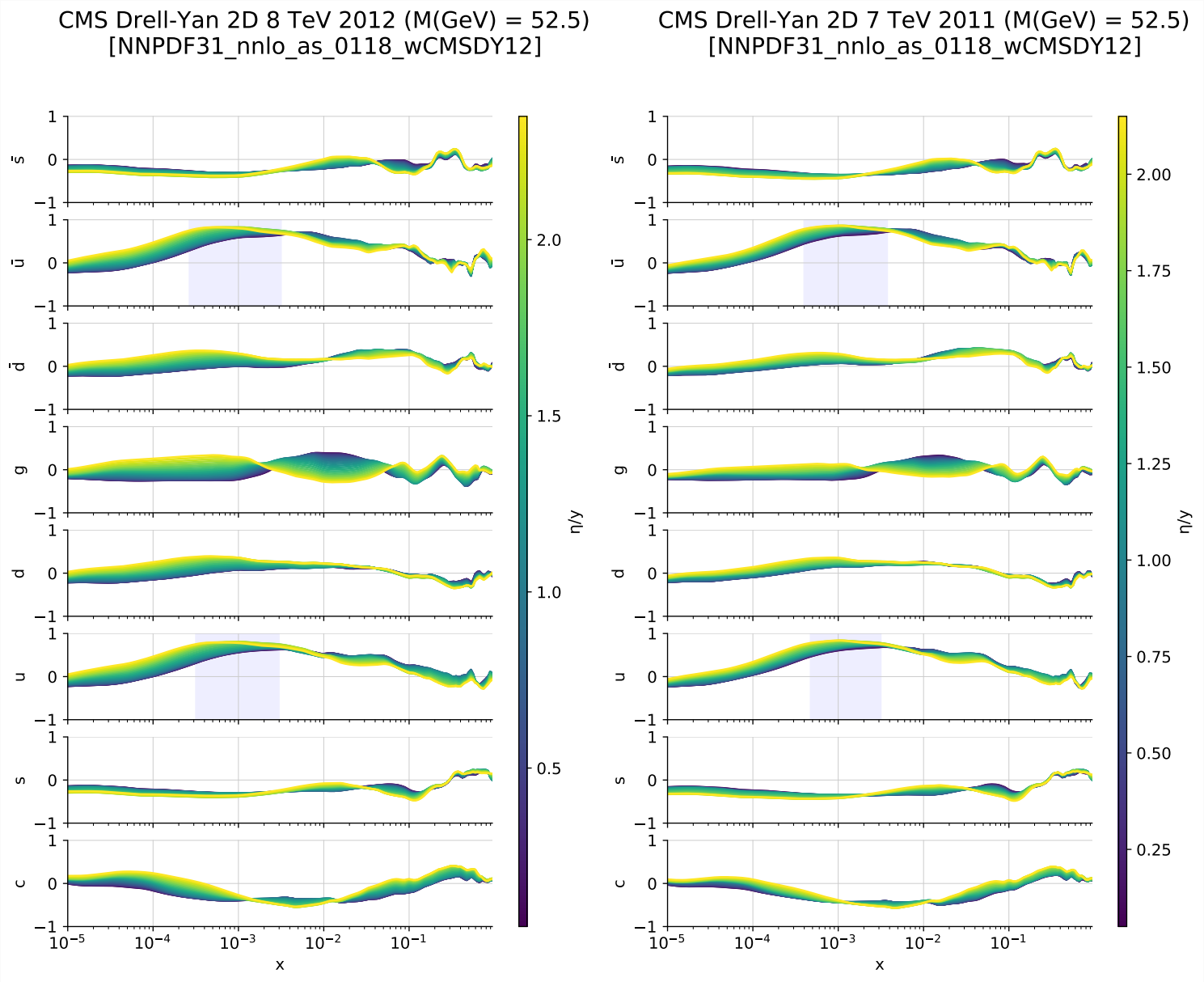




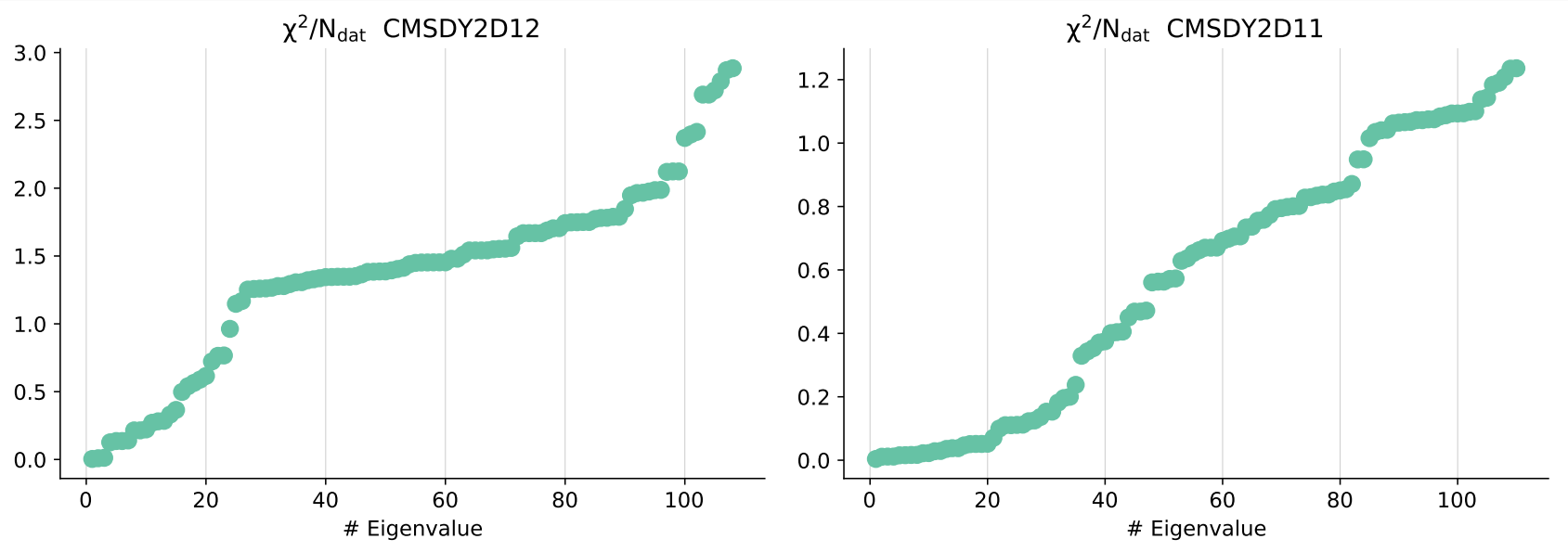
$\chi^2/N_{\text{dat}} = 2.88$ $\chi^2/N_{\text{dat}} = 1.24$

Similar kinematics and apparent agreement to the 2011 data, yet much worse actual χ^2.

Similar PDF dependence



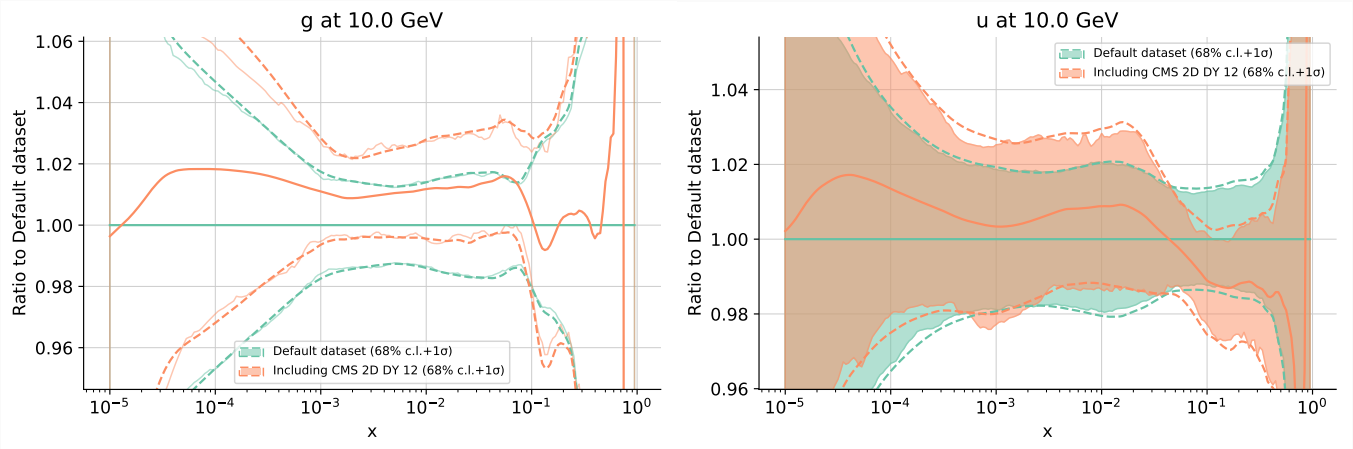
Checking the correlations



- Most of the disagreement on the ’12 data comes from a few eigenvectors of the covariance matrix.
- The 2012 data have extremely small statistical uncertainties.
  - Covariance matrix close to singular. Systematics need to be very accurate.
- Unfortunately couldn’t study much further:
  - No breakdown of systematics available.

Effects of including the ’12 data

- Bad fit quality for the ’12 data.
- Deterioration of the fit quality of other datasets (precise HERA data).
- Big change in the PDFs (particularly gluon).



- Including the data doesn’t seem advantageous.
- Also, similar conclusions by other fitting groups.

Dealing with uncertain covariance matrices

- Luminosity upgrades shift the weight of the uncertainty from statistical to systematic components.
- Covariance matrices increasingly more unstable and sensitive to small errors.
- How should we deal with this? Can we model the uncertainty of the uncertainty somehow? Wishart models?

Precise PDF calculations in practice

Conceptually:

$$\sigma_{\text{had}} = \hat{\sigma} \otimes f(x, Q^2)$$

In practice:

- $NN(x, Q_0^2)$ determined at the initial scale.

Precise PDF calculations in practice

Conceptually:

$$\sigma_{\text{had}} = \hat{\sigma} \otimes f(x, Q^2)$$

In practice:

- $NN(x, Q_0^2)$ determined at the initial scale.
- DGLAP solved over an *interpolated* grid in x by APFEL.

Precise PDF calculations in practice

Conceptually:

$$\sigma_{\text{had}} = \hat{\sigma} \otimes f(x, Q^2)$$

In practice:

- $NN(x, Q_0^2)$ determined at the initial scale.
- DGLAP solved over an *interpolated* grid in x by APFEL.
- An *interpolated* grid in (x, Q^2) written to LHAPDF.

Precise PDF calculations in practice

Conceptually:

$$\sigma_{\text{had}} = \hat{\sigma} \otimes f(x, Q^2)$$

In practice:

- $NN(x, Q_0^2)$ determined at the initial scale.
- DGLAP solved over an *interpolated* grid in x by APFEL.
- An *interpolated* grid in (x, Q^2) written to LHAPDF.
- LHAPDF *interpolates* the values in (x, Q^2) requested by the user.

Precise PDF calculations in practice

Conceptually:

$$\sigma_{\text{had}} = \hat{\sigma} \otimes f(x, Q^2)$$

In practice:

- $NN(x, Q_0^2)$ determined at the initial scale.
- DGLAP solved over an *interpolated* grid in x by APFEL.
- An *interpolated* grid in (x, Q^2) written to LHAPDF.
- LHAPDF *interpolates* the values in (x, Q^2) requested by the user.
- Convolution performed in a finite grid in (x, Q^2) where f and $\hat{\sigma}$ are *interpolated*.

Precise PDF calculations in practice

Conceptually:

$$\sigma_{\text{had}} = \hat{\sigma} \otimes f(x, Q^2)$$

In practice:

- $NN(x, Q_0^2)$ determined at the initial scale.
- DGLAP solved over an *interpolated* grid in x by APFEL.
- An *interpolated* grid in (x, Q^2) written to LHAPDF.
- LHAPDF *interpolates* the values in (x, Q^2) requested by the user.
- Convolution performed in a finite grid in (x, Q^2) where f and $\hat{\sigma}$ are *interpolated*.
- ...plus the ~reverse path to fit the PDFs.

Precise PDF calculations in practice

Conceptually:

$$\sigma_{\text{had}} = \hat{\sigma} \otimes f(x, Q^2)$$

In practice:

- $NN(x, Q_0^2)$ determined at the initial scale.
- DGLAP solved over an *interpolated* grid in x by APFEL.
- An *interpolated* grid in (x, Q^2) written to LHAPDF.
- LHAPDF *interpolates* the values in (x, Q^2) requested by the user.
- Convolution performed in a finite grid in (x, Q^2) where f and $\hat{\sigma}$ are *interpolated*.
- ...plus the ~reverse path to fit the PDFs.

All those interpolations need to work at percent level!

The future (personal view)

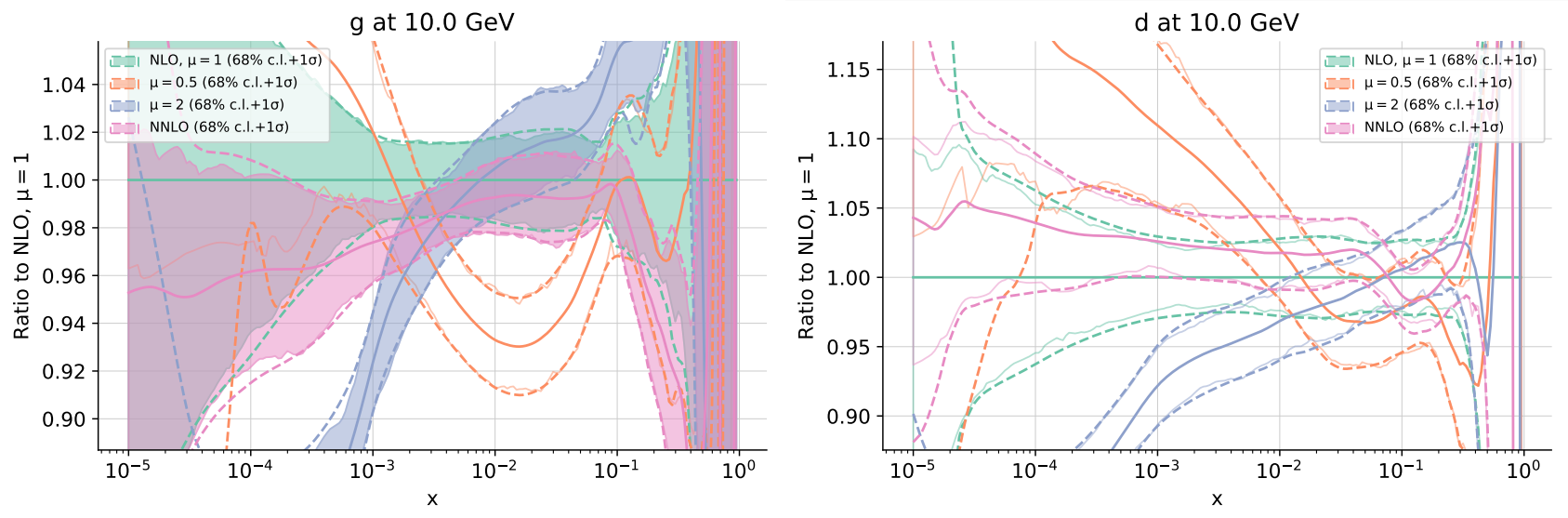
- Future progress will focus on methodological upgrades:
  - Can we use better minimization algorithms? Will the result be more precise? More accurate?
  - Can we remove all ad hoc settings with *deep learning* techniques?
  - Can we run fits (+precomputed grids) faster (GPUs)?
    - Or remove fits altogether (*reminimization* starting from a prior).
- Inclusion of more sources of uncertainty:
  - Scale variations: First results already available!

PDFs with scale variations

- Technology ready at NLO (can produce fits with arbitrary μ_R/μ_F).
- At NNLO we need to transition to *grids* as opposed to K factors first.
- Not clear how to correlate the scales for various processes or what's the final deliverable.

Example: All scales set multiplied and divided by two:

First results $\mu = 0.5$ and $\mu = 2$



Thank you!