Extracting the Strong Coupling at the EIC and other Future Colliders

May 5 - 7, 2025

Center for Frontiers in Nuclear Science, Stony Brook University

CT results on $\alpha_s(m_Z)$

J. Huston

for Alim Ablat, Sayip Dulat, Yao Fu, Kirtimaan Mohan, Pavel Nadolsky, Dan Stump, C.-P. Yuan

CTEQ-TEA group

Advertisement

CT18up enhanced precision LHAPDF grids (2023)

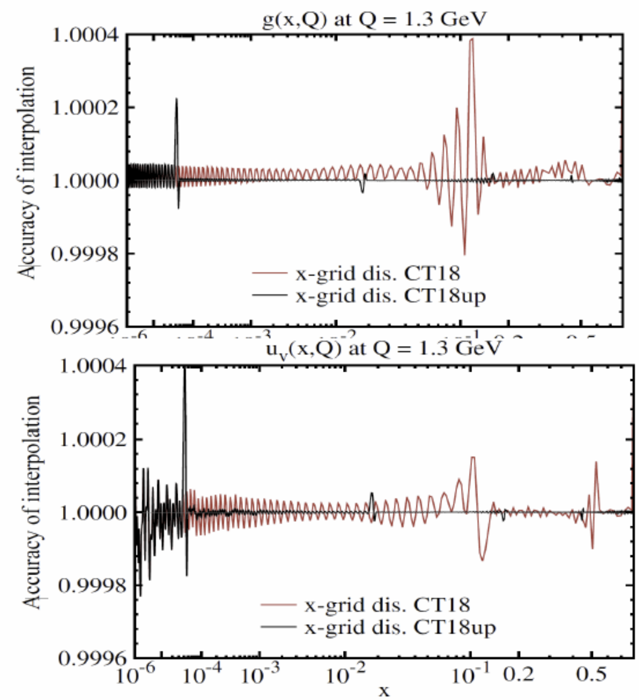
On https://cteq-tea.gitlab.io/project/00pdfs/

- CT18, A, X, Z NNLO PDFs (2019 edition) presented as LHAPDF grids with a 1.9x higher number of x and Q nodes; recommended for high-mass, precision calculations, estimates of the interpolation error
- Same PDFs as in the LHAPDF library, with even more precise interpolation at $10^{-4} \leq x \leq 1$
- 2019 grids ok in other cases

Numbers of x, Q nodes in LHAPDF grids

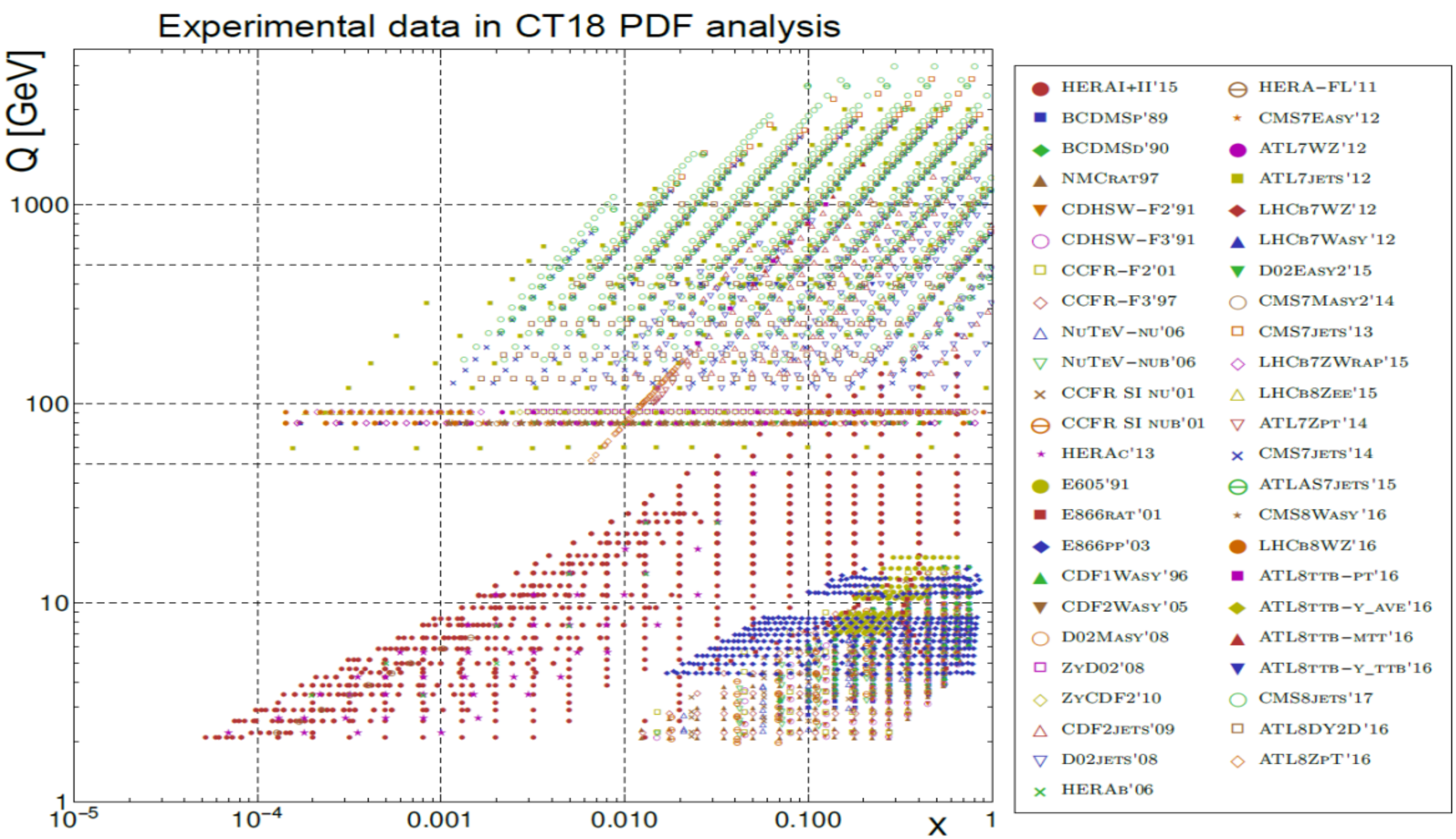
intervals in Q	CT18	CT18up
$[Q_0, m_c]$	2	4
$[m_c, m_b]$	8	11
$[m_b, m_t]$	14	18
$[m_t, Q_{\max}]$	13	16
Total	37	49

intervals in x	CT18	CT18up	intervals in x	CT18	CT18up
$[10^{-10}, 10^{-9}]$	1	1	$[0.1, 0.2]$	7	18
$[10^{-9}, 10^{-8}]$	11	11	$[0.2, 0.3]$	6	16
$[10^{-8}, 10^{-7}]$	12	12	$[0.3, 0.4]$	5	12
$[10^{-7}, 10^{-6}]$	11	11	$[0.4, 0.5]$	3	13
$[10^{-6}, 10^{-5}]$	12	12	$[0.5, 0.6]$	6	15
$[10^{-5}, 10^{-4}]$	11	15	$[0.6, 0.7]$	6	12
$[10^{-4}, 10^{-3}]$	12	23	$[0.7, 0.8]$	8	11
$[10^{-3}, 10^{-2}]$	11	23	$[0.8, 0.9]$	14	17
$[10^{-2}, 0.1]$	12	40	$[0.9, 1]$	15	38
			Total	161	300



Prelude: CT18

- There is a wide variety of data in modern global PDF analyses, over 3500 data points for CT18
- The data includes DIS, DY (including precision W/Z), jet production, top production
  - NB: ATLAS 7 TeV W/Z data not included in CT18, but rather in CT18A, because of tensions with other data sets
- All predictions at NNLO, all depending on α_s



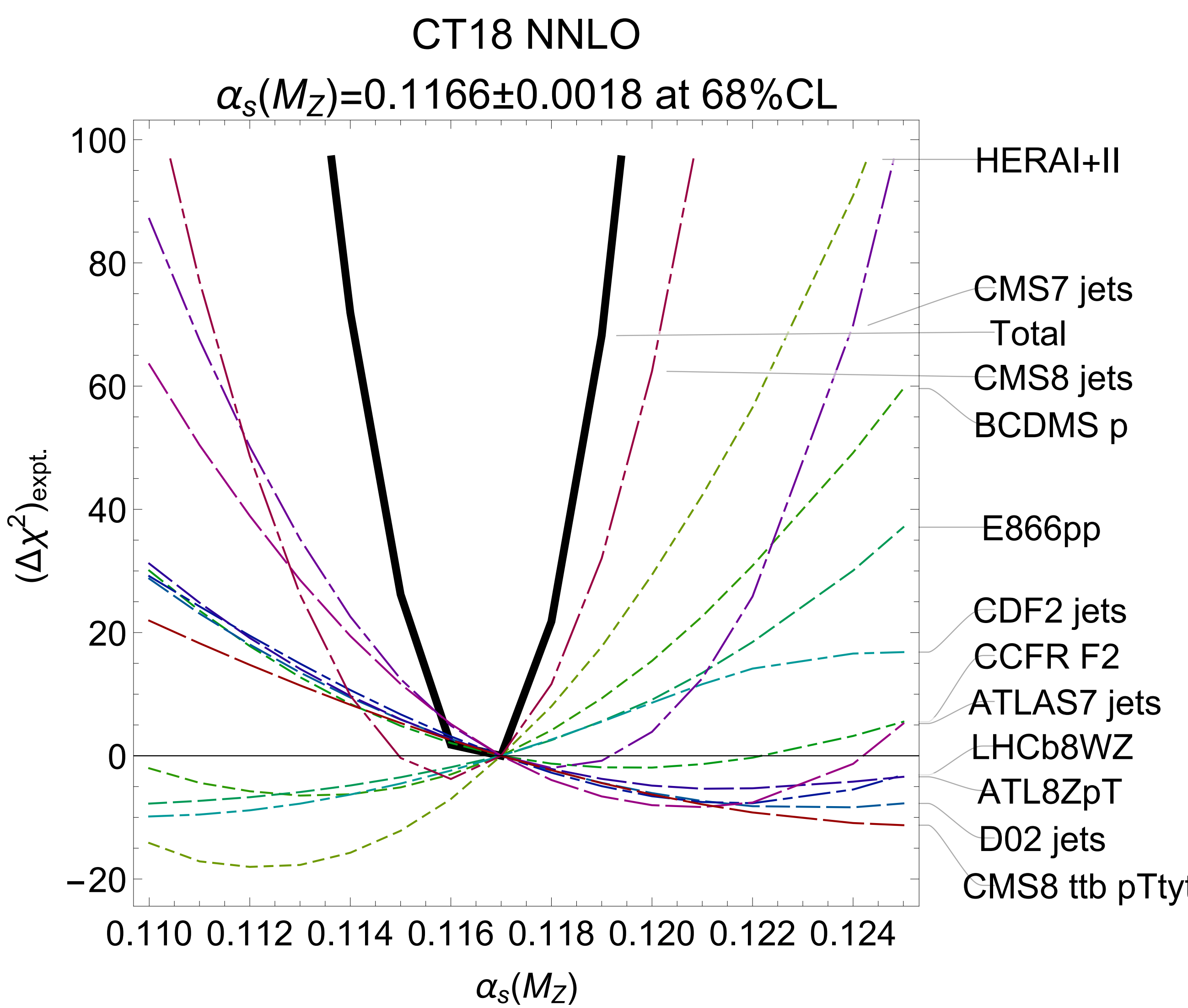
...but the power of α_s depends on the process

Born for DY is $\alpha_s{}^0$; Born for dijet/top production is $\alpha_s{}^2$

α_s from CT18 (*Phys.Rev.D* 103 (2021) 1, 014013)

- All of the experiments in the global fit do not speak with a unified voice, weakening the discrimination power
- We end up with a fairly parabolic χ^2 dependence of $\alpha_s(m_Z)$, but it's clear that different experiments have different preferences
- At 68% CL, $\alpha_s(m_Z)$=0.1166+/-0.0018 (for CT18 at NNLO)
- This assumes a 68%CL with a $\Delta\chi^2$ of 37
- Uncertainty of $\alpha_s(m_Z)$ depends on tolerance and on degree of conflict within data samples

Lagrange Multiplier scans



→ theme of this talk

Towards CT25

MSUHEP-24-020

The impact of LHC precision measurements of inclusive jet and dijet production on the CTEQ-TEA global PDF fit

Alim Ablat,[1] Sayipjamal Dulat,[1,2,*] Tie-Jiun Hou,[3] Joey Huston,[2]
Pavel Nadolsky,[2,4] Ibrahim Sitiwaldi,[1] Keping Xie,[2] and C.-P. Yuan[2]
(CTEQ-TEA Collaboration)

[1] *School of Physics Science and Technology, Xinjiang University, Urumqi, Xinjiang 830046 China*
[2] *Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA*
[3] *School of Nuclear Science and Technology, University of South China, Hengyang, Hunan 421001, China*
[4] *Department of Physics, Southern Methodist University, Dallas, TX 75275-0181, USA*

In this study, we investigate the impact of new LHC inclusive jet and dijet measurements on parton distribution functions (PDFs) that describe the proton structure, with a particular focus on the gluon distribution at large momentum fraction, x, and the corresponding partonic luminosities. We assess constraints from these datasets using next-to-next-to-leading-order (NNLO) theoretical predictions, accounting for a range of uncertainties from scale dependence and numerical integration. From the scale choices available for the calculations, our analysis shows that the central predictions for inclusive jet production show a smaller scale dependence than dijet production. We examine the relative constraints on the gluon distribution provided by the inclusive jet and dijet distributions, and also explore the phenomenological implications for inclusive H, $t\bar{t}$, and $t\bar{t}H$ production at the LHC at 14 TeV.

CONTENTS

Towards CT25

NNLO fits with new data (nD) from LHC at 8 and 13 TeV

χ^2/N_{pt} for CT18A + new data (*vs.* CT18A at right) NNLO fits; 68% CL

ID	Experimental data set	N_{pt}	CT25prel	CT18A
	Drell-Yan pair production			
211	ATLAS 8 TeV W	22	$2.37^{+1.30}_{-0.68}$	$3.56^{+2.26}_{-1.65}$
212	CMS 13 TeV Z	12	$2.10^{+2.20}_{-0.38}$	$2.13^{+3.46}_{-0.30}$
214	ATLAS 8 TeV Z 3D	188	$1.14^{+0.10}_{-0.04}$	$1.21^{+0.32}_{-0.15}$
215	ATLAS 5.02 TeV W,Z	27	$0.70^{+0.27}_{-0.06}$	$0.74^{+0.31}_{-0.08}$
217	LHCb 8 TeV W	14	$1.36^{+0.37}_{-0.34}$	$1.47^{+0.43}_{-0.38}$
218	LHCb 13 TeV Z	16	$1.06^{+0.76}_{-0.38}$	$1.29^{+0.95}_{-0.44}$
	13 TeV $t\bar{t}$ production			
521	ATLAS all-hadronic $y_{t\bar{t}}$	12	$1.07^{+0.08}_{-0.05}$	$1.07^{+0.12}_{-0.07}$
528	CMS dilepton $y_{t\bar{t}}$	10	$1.10^{+0.56}_{-0.40}$	$1.13^{+0.85}_{-0.53}$
581	CMS lepton+jet $m_{t\bar{t}}$	15	$1.38^{+0.65}_{-0.40}$	$1.44^{+0.89}_{-0.56}$
587	ATLAS lepton+jet $m_{t\bar{t}} + y_{t\bar{t}} + y^B_{t\bar{t}} + H^{t\bar{t}}_T$	34	$0.94^{+0.13}_{-0.11}$	$0.94^{+0.28}_{-0.09}$
	Inclusive jet production			
553	ATLAS 8 IncJet	171	$1.54^{+0.09}_{-0.06}$	$1.57^{+0.12}_{-0.07}$
554	ATLAS 13 IncJet	177	$1.25^{+0.07}_{-0.03}$	$1.26^{+0.08}_{-0.04}$
555	CMS 13 IncJet	78	$1.11^{+0.13}_{-0.09}$	$1.10^{+0.21}_{-0.10}$

(fits with 1 new process, 'nProcces')

nDY

nTT

nIncJet

nD

(combined fit)

also dijet data, but don't (officially) know how to combine with inclusive jets

776 new data points from DY, tT and jets for CT25

Towards CT25

FITS

NNLO fits with new data (nD) from LHC at 8 and 13 TeV

χ^2/N_{pt} for CT18A + new data (*vs.* CT18A at right) NNLO fits; 68% CL

(fits with 1 new process, 'nProcces')

ID	Experimental data set	N_{pt}	CT25prel	CT18A
	Drell-Yan pair production			
211	ATLAS 8 TeV W	22	$2.37^{+1.30}$	$2.56^{+2.26}$
581	CMS lepton+jet $m_{t\bar{t}}$	15	$1.38^{+0.65}_{-0.40}$	$1.44^{+0.89}_{-0.56}$
587	ATLAS lepton+jet $m_{t\bar{t}} + y_{t\bar{t}} + y^B_{t\bar{t}} + H^{t\bar{t}}_T$	34	$0.94^{+0.13}_{-0.11}$	$0.94^{+0.28}_{-0.09}$
	Inclusive jet production			
553	ATLAS 8 IncJet	171	$1.54^{+0.09}_{-0.06}$	$1.57^{+0.12}_{-0.07}$
554	ATLAS 13 IncJet	177	$1.25^{+0.07}_{-0.03}$	$1.26^{+0.08}_{-0.04}$
555	CMS 13 IncJet	78	$1.11^{+0.13}_{-0.09}$	$1.10^{+0.21}_{-0.10}$

nIncJet

Les Houches 2023 - Physics at TeV Colliders: Report on the Standard Model Precision Wishlist

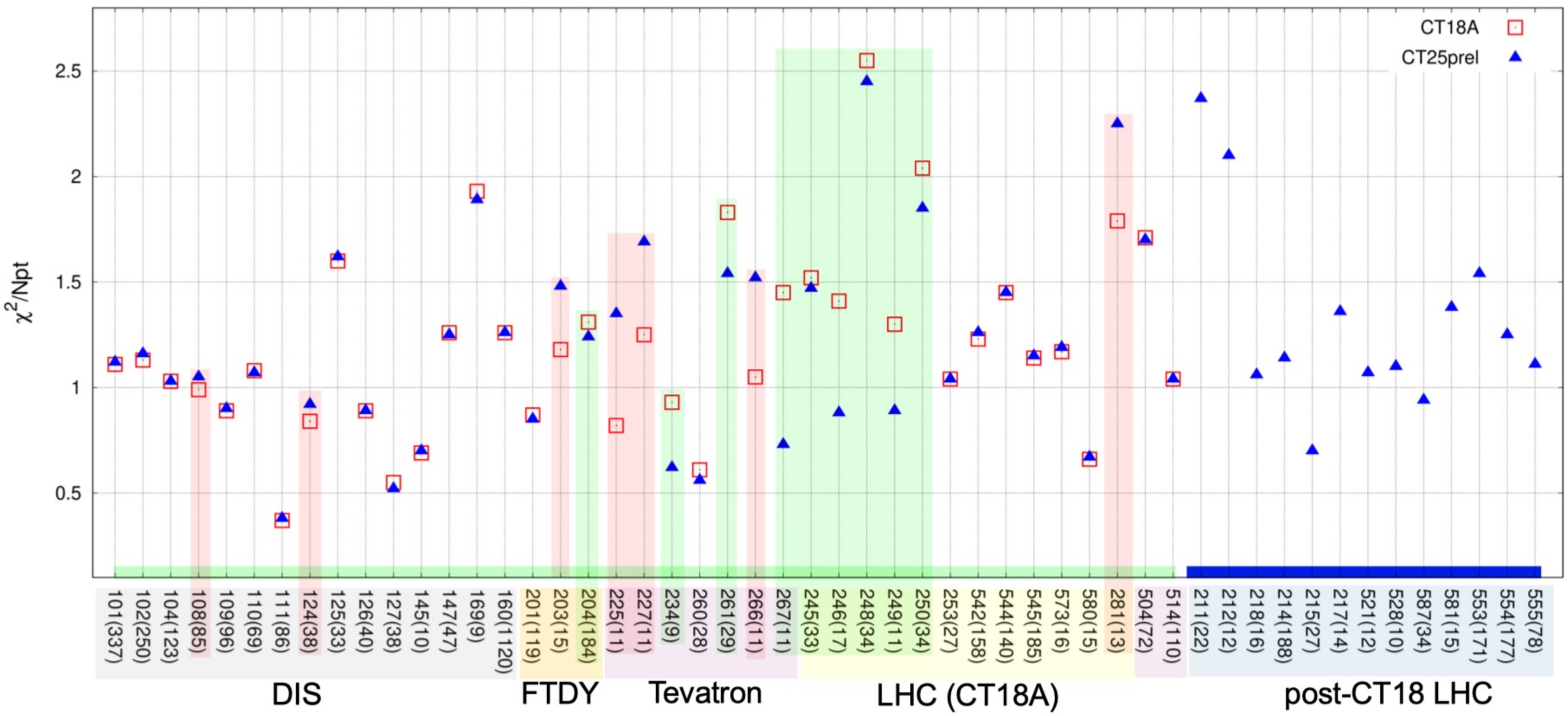
arXiv:2504.06689

Alexander Huss[1], Joey Huston[2], Stephen Jones[3], Mathieu Pellen[4], Raoul Röntsch[5]

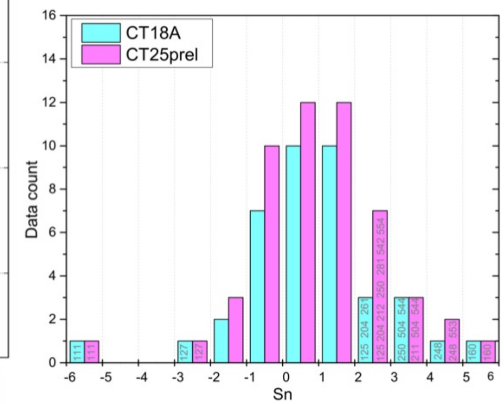
776 new data points from DY, tT and jets for CT25

also dijet data, but don't know how to combine with inclusive jets

CT18A baseline vs. CT25prel NNLO (= CT18A + nDY + nTT +nIncJet)



-ness



⟶ new experiments generally fit well; high-precision sets can have (slightly) elevated χ^2/N_{pt}, which, specifically,

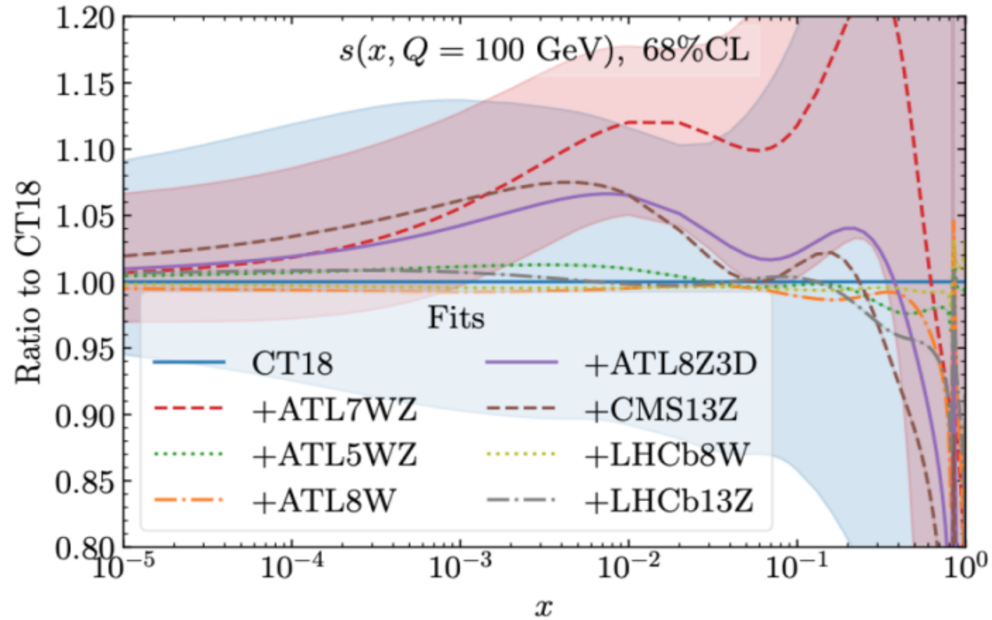
- **increases** for: 108 (CDHSW); 124 (NuTeV); 203 (E866 D/p), 225 and 227 (CDF A_{ch}^{lep}); 266 (CMS 7 TeV A_{ch}^{μ})
- **decreases** for: 204 (E866 pp); 234 (D0 A_{ch}^{lep}), 261 (CDF y_Z); 267 (CMS 7 TeV A_{ch}^{e}), 245 (LHCb 7 TeV $y_{W/Z}$), 246 (LHCb 8 TeV y_Z), 248 (ATLAS 7 TeV W/Z), 249 (CMS 8 TeV A_{ch}^{μ}), 250 (LHCb 8 TeV W/Z)

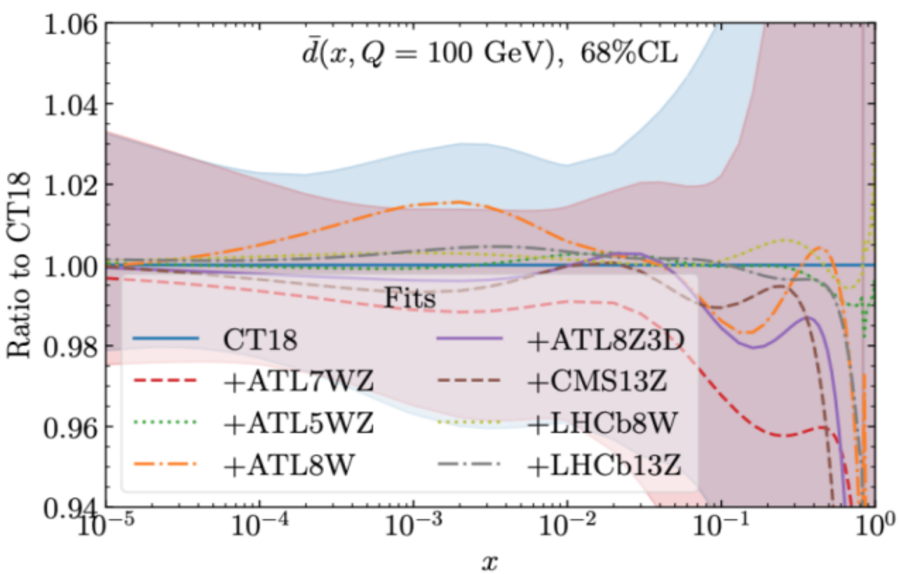
- The post-CT18 Drell-Yan data from the LHC experiments at 8 TeV and 13 TeV are found to generally impose similar constraints on s-PDF (and $\bar{d}$-PDF) as the ATLAS data from the 7 TeV W/Z processes. However, the constraints from the Drell-Yan data are significantly weaker.
- Thus, CT25 starts from CT18A rather than CT18
  - but basically, we just wanted to make Mandy happy

Post-CT18 Drell-Yan data's impact

2305.10733 (PRD23')

ID	Experiment	N_{pt}	CT18	CT18A	CT18As
215	ATLAS 5.02 TeV W, Z	27	0.81	0.71	0.71
211	ATLAS 8 TeV W	22	2.45	2.63	2.51
214	ATLAS 8 TeV Z 3D†	188	1.12	1.14	1.18
212	CMS 13 TeV Z	12	2.38	2.03	2.71
216	LHCb 8 TeV W	14	1.34	1.36	1.43
213	LHCb 13 TeV Z	16	1.10	0.98	0.83
248	ATLAS 7 TeV W, Z	34	2.52	2.50	2.30
Total 3994/3953/3959 points			1.20	1.20	1.19



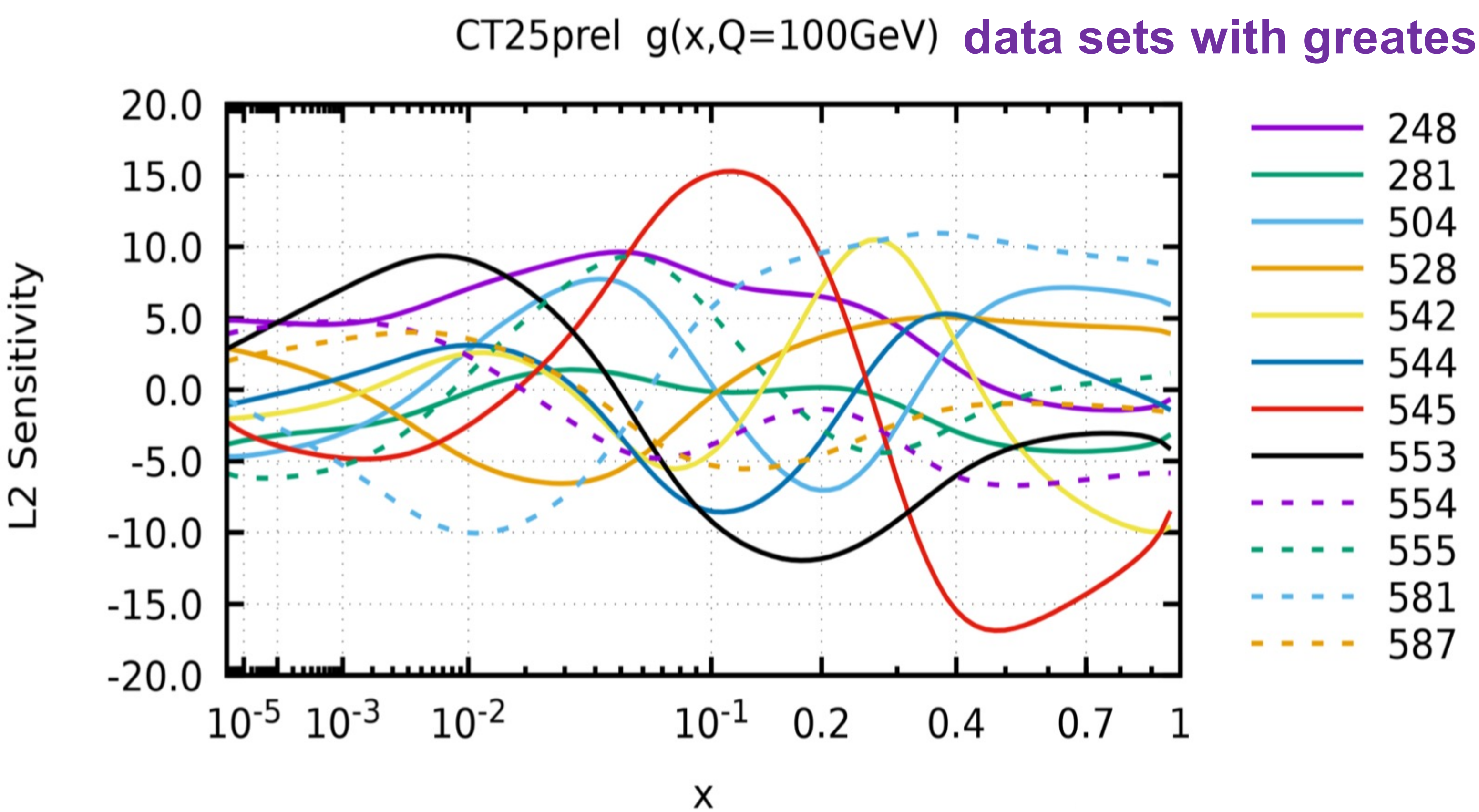


- Many new Drell-Yan (nDY) data came out after the release of CT18 PDFs.
- We found that most of the nDY data sets are consistent with the ATLAS 7 WZ precision data (16') and prefer enhanced strangeness at $x \sim 0.02$
- Only one exception: ATL8W has an opposite pull on $d, \bar{d}$
- CMS13Z and ATL8W have a similar χ^2/N_{pt} as ATL7WZ
- The more flexible strangeness parameterization in CT18As can relax the tension, but not completely resolve it.

L2 sensitivity of new data

There is a strong correlation between the value of $\alpha_s(m_Z)$ and the gluon distribution. So it's also important to understand how the gluon is determined.

data sets with greatest L2 sensitivity



200's are DY data

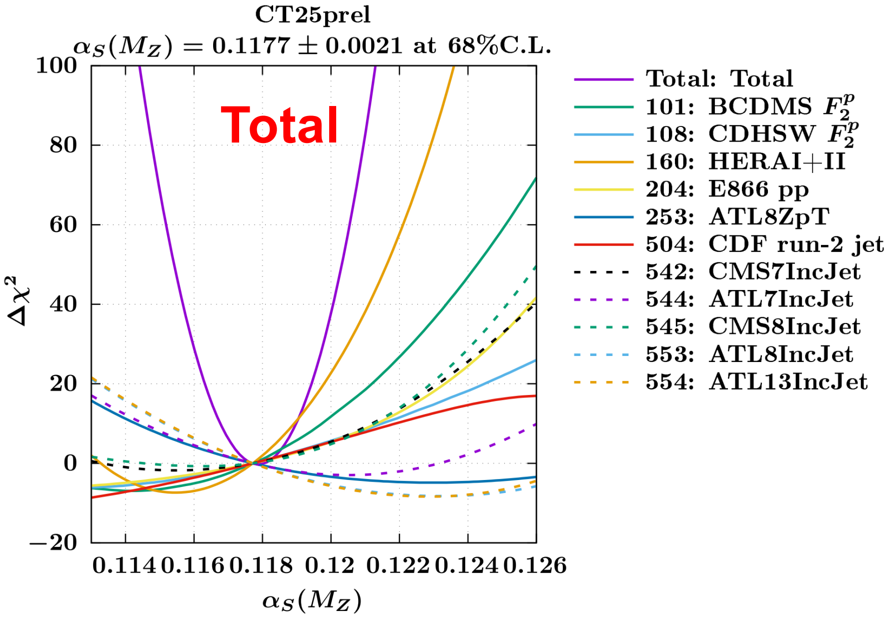
504-555 are jet data

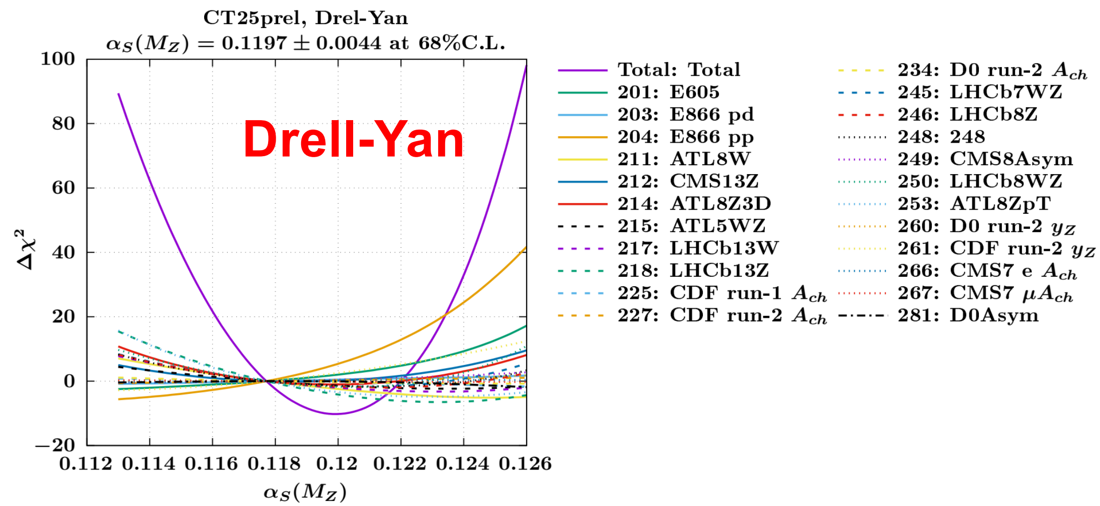
581-587 tT data

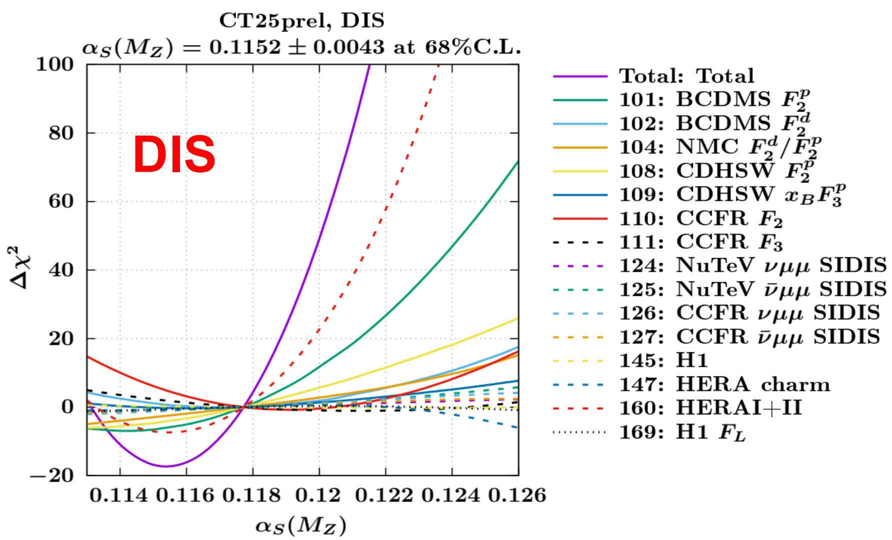
By definition, there is tension among the data included in the global PDF fit. A positive value of L2 sensitivity means the data is trying to pull the gluon down, and vice versa. The sum of all L2 sensitivity values at a particular x value should be close to zero.

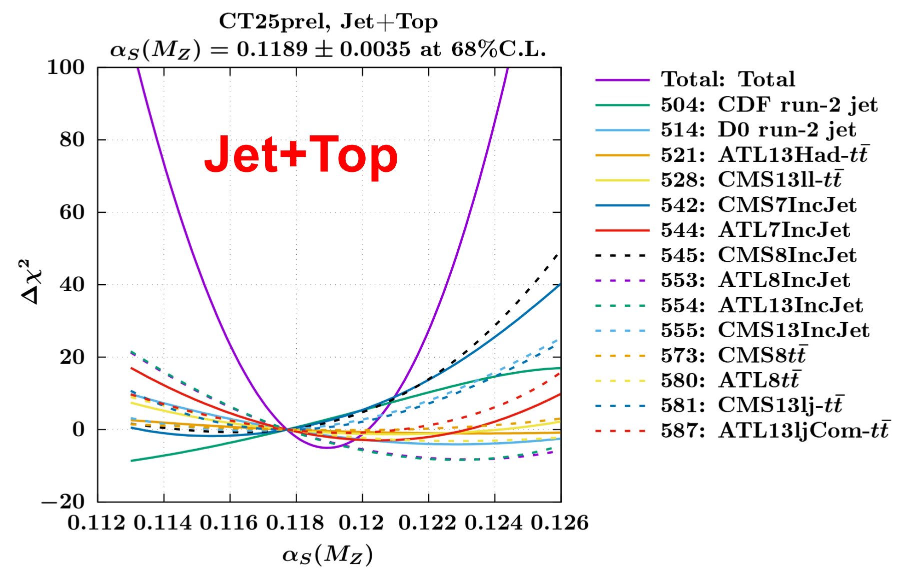
•*Phys.Rev.D* 108 (2023) 3, 034029 e-Print: 2306.03918 (L2 sensitivity paper)

CT25prel PDF+α_s fit (Lagrange Multiplier scans)

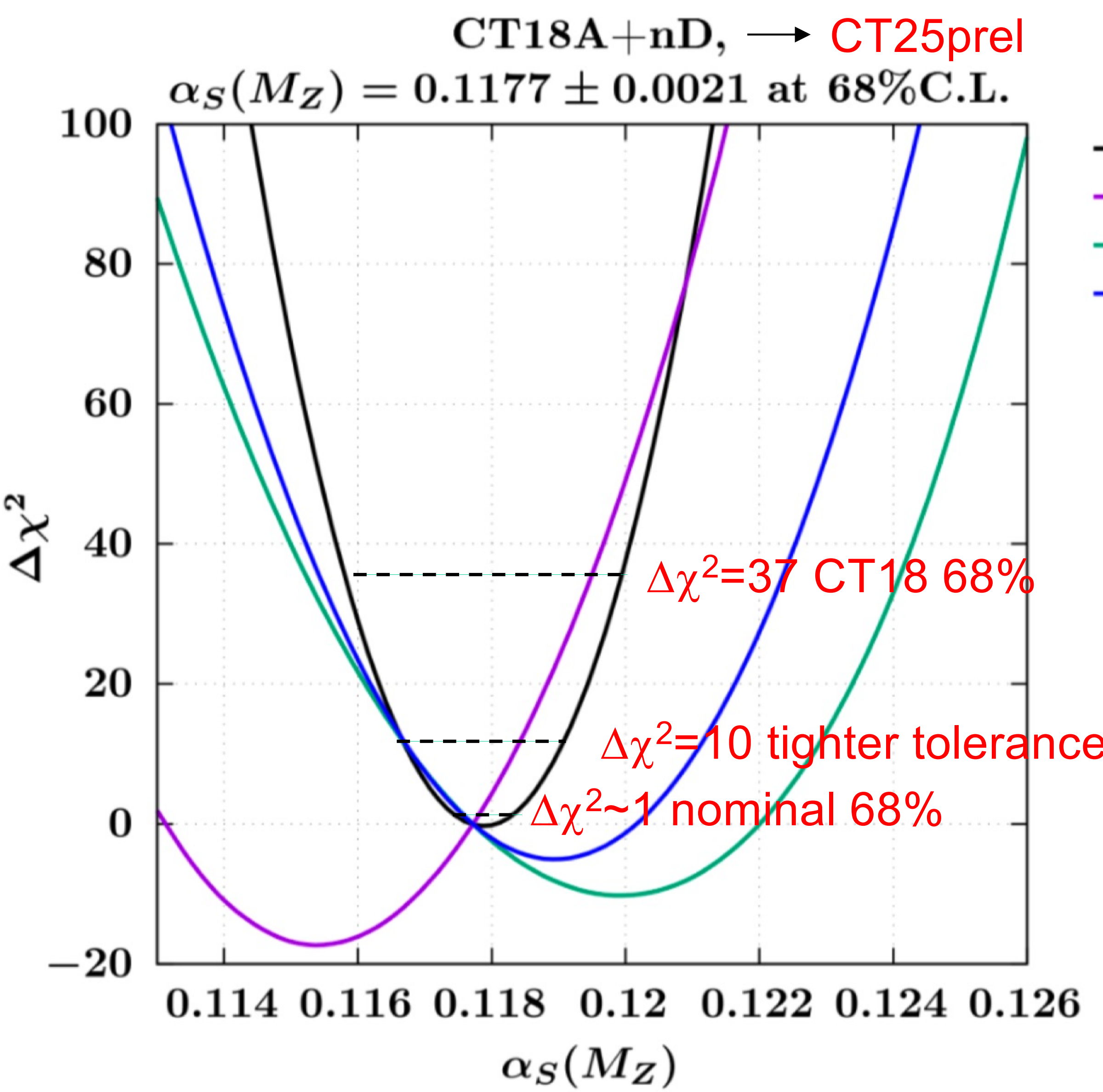




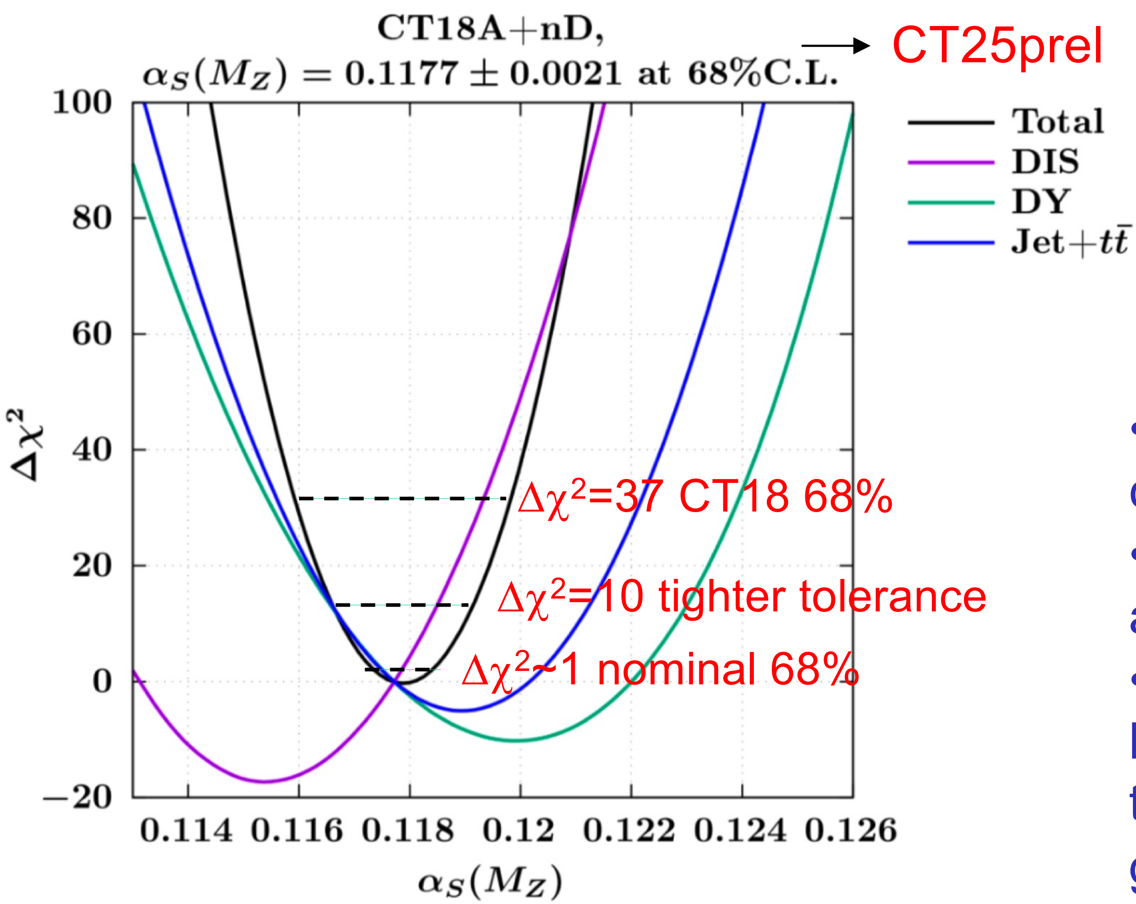




Towards CT25 $\alpha_s(m_Z)$



- Tension between DIS data and collider data
- Smaller tension between jet+tT and DY
- Result is parabolic (by definition), but $\Delta\chi^2$ near one does not reflect the spread of the information that goes into the determination

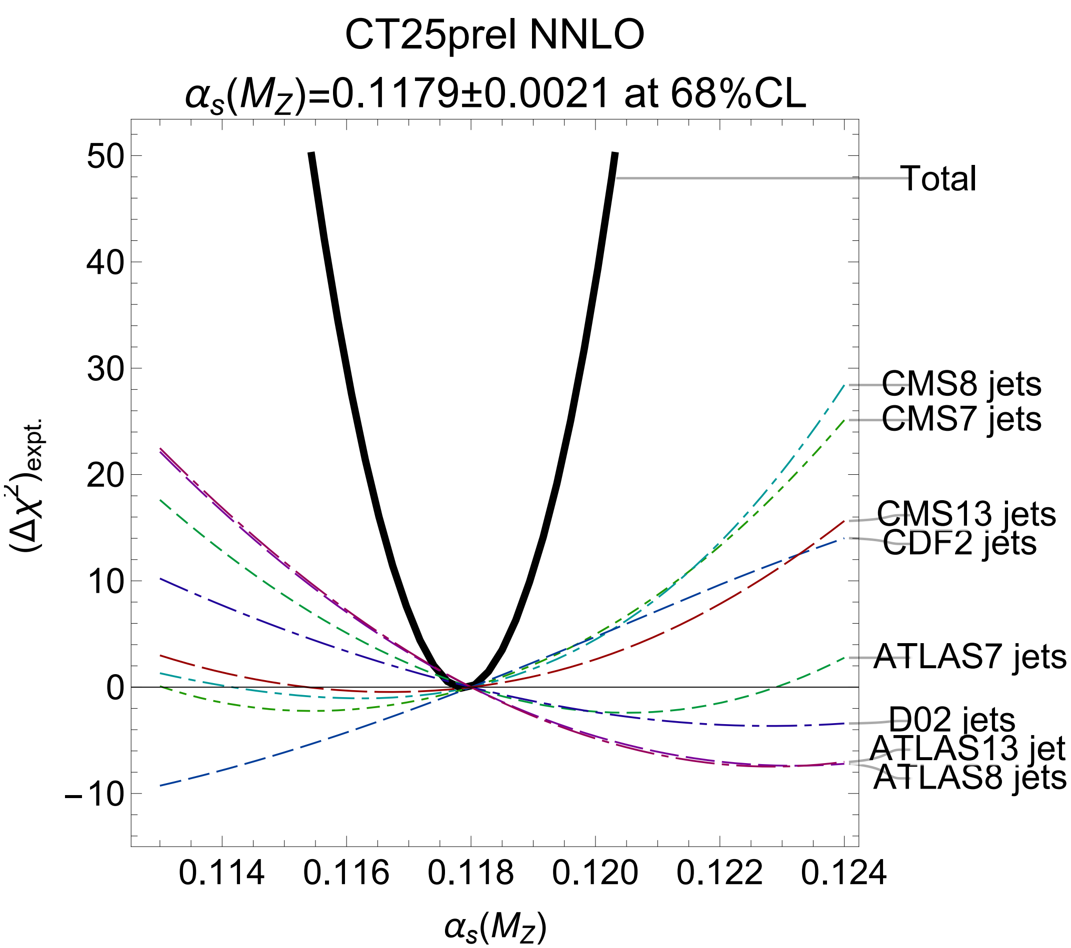


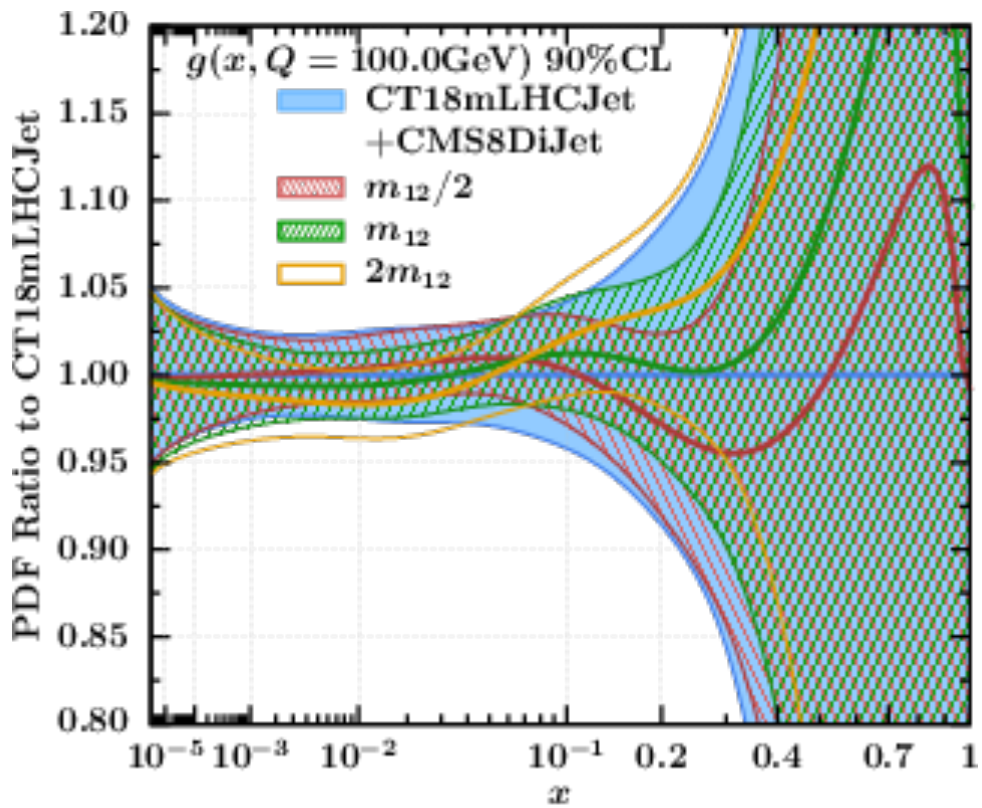
- Tension between DIS data and collider data
- Smaller tension between jet+tT and DY
- Result is parabolic (by definition), but $\Delta\chi^2$ near one does not reflect the spread of the information that goes into the determination

CT25prel All data with weight=1	Total	DIS	DY	Jet+tt
# of data points	4450	2381	876	1193
Minimal χ^2 value	5346	2714	1100	1531
$(\alpha_s, \Delta\alpha_s)$ with $\Delta\chi^2 = 37$	0.1177, 0.0021	0.1152 , 0.0043	0.1197, 0.0044	0.1189, 0.0035
$(\alpha_s, \Delta\alpha_s)$ with $\Delta\chi^2 = 10$	0.1177, 0.0012	0.1152, 0.0030	0.1197, 0.0031	0.1189, 0.0021

Consider jet data alone

- dijet data (not in CT25pre)
- scale dependence seems larger than for inclusive jet
- under investigation



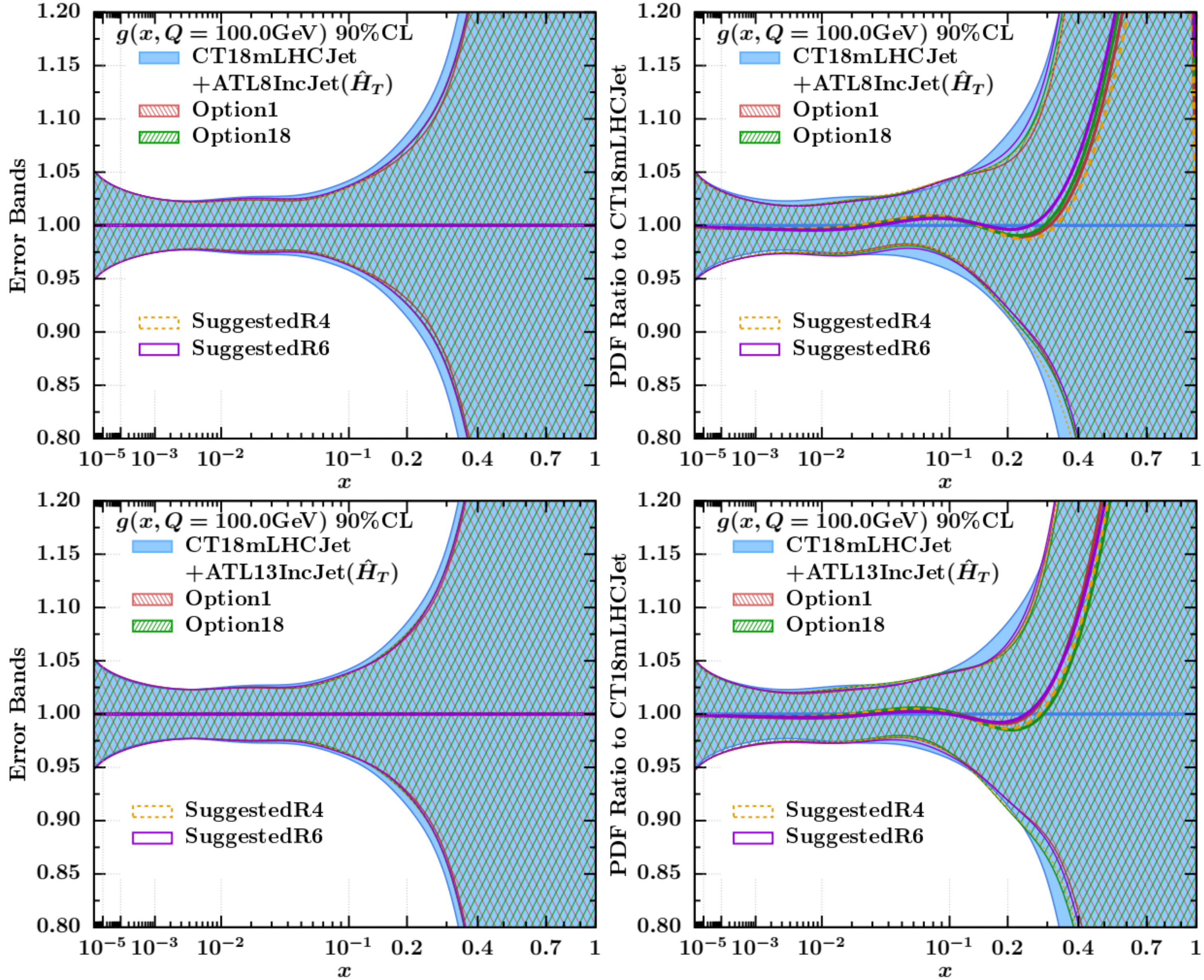


Aside: Decorrelating errors

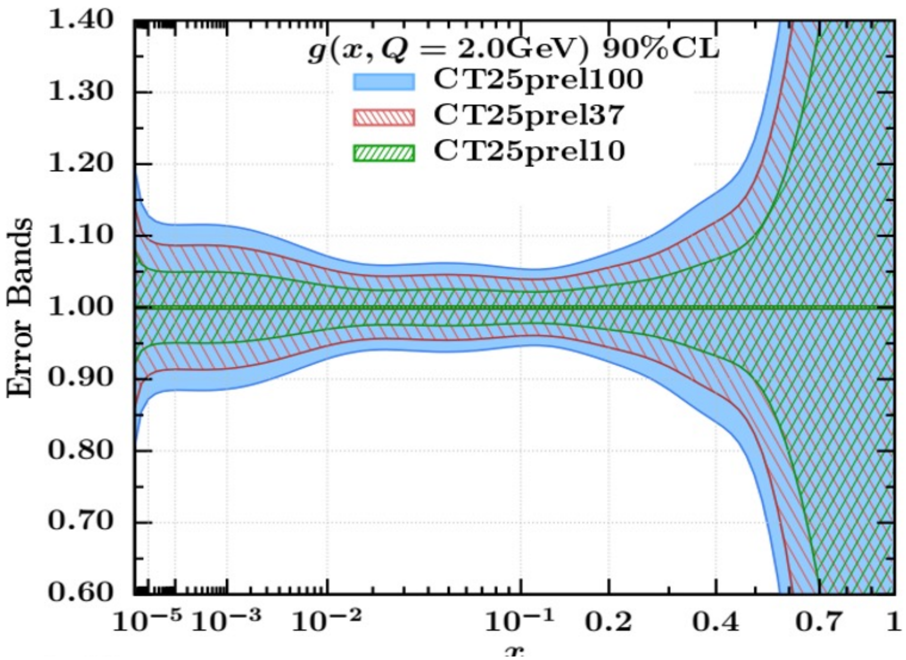
The table lists $\chi 2$ from different decorrelation options applied to ATLAS 8 TeV and 13 TeV inclusive jet data sets using ePump profiling. Decorrelation improves the goodness-of-fit $\chi 2$

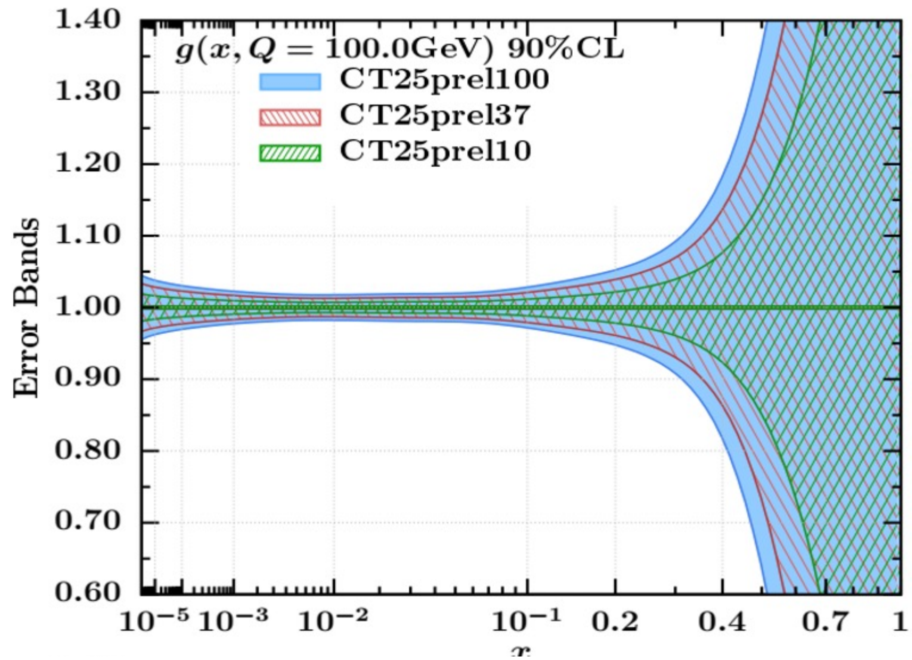
Option	ATLAS 8 TeV IncJet		ATLAS 13 TeV IncJet	
Scale	$\hat{H}_T$	p_T^j	$\hat{H}_T$	p_T^j
Fully correlated	3.04	3.77	2.87	4.10
Option 1	3.03	3.76	2.87	4.10
Option 2	3.04	3.77	2.87	4.10
Option 3	3.04	3.77	2.87	4.10
Option 4	3.04	3.77	2.87	4.10
Option 5	3.03	3.76	2.82	4.05
Option 6	3.03	3.76	2.86	4.09
Option 7	2.99	3.68	2.82	4.03
Option 8	3.03	3.75	2.85	4.07
Option 9	3.04	3.77	2.87	4.10
Option 10	3.01	3.74	2.87	4.10
Option 11	3.04	3.77	2.87	4.10
Option 12	3.03	3.76	2.87	4.10
Option 13	2.84	3.54	2.85	4.08
Option 14	2.85	3.57	2.81	4.03
Option 15	2.94	3.66	2.85	4.08
Option 16	2.84	3.56	2.79	4.01
Option 17	2.90	3.61	2.79	4.00
Option 18	2.83	3.52	2.76	3.98
SuggestedR4	2.76	3.45	2.78	3.99
SuggestedR6	2.75	3.47	2.79	4.01

Decorrelating errors has little impact on gluon distribution



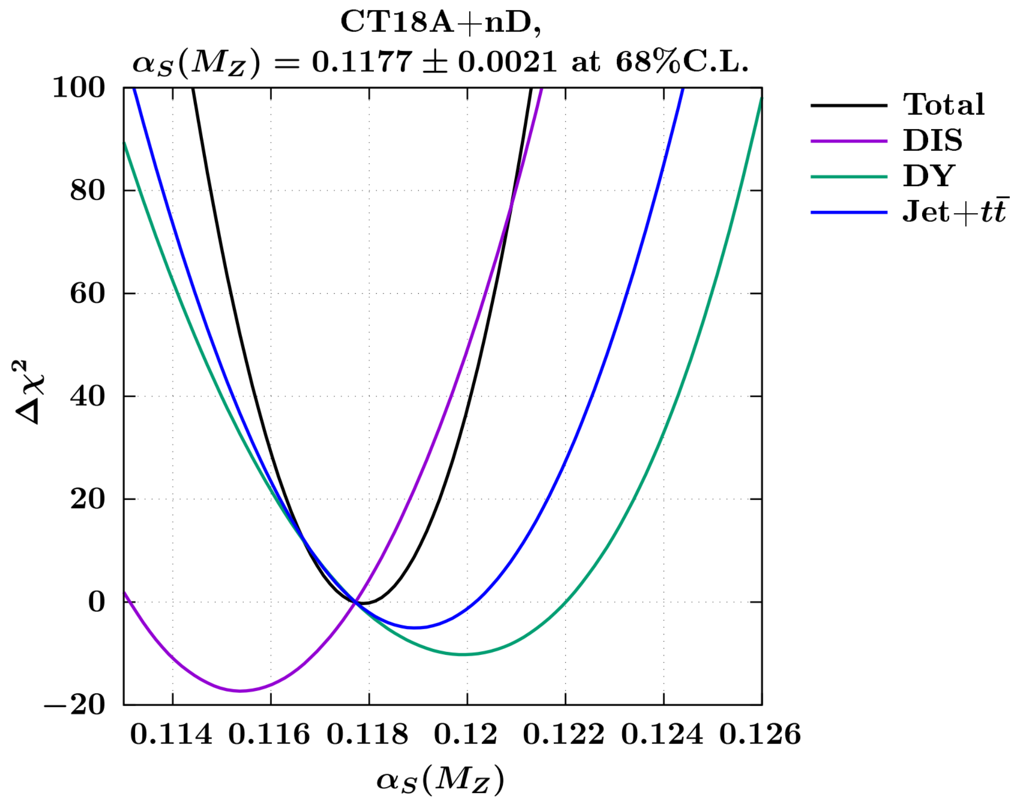
Gluon-PDF error bands with various Tolerance values





Combining three measurements

- Consider the fit shown in the plot
- It is easy to see that there is a large tension between DIS and other data sets.
- Treat each of the sets (DIS, DY, Jet +$t\bar{t}$) as independent and identically distributed measurements of α_s
- $\chi^2_{tot} = \sum_i \chi^2_i$
- Mean and Error given by minimizing
- $\tilde{\chi}^2 = \sum_i \frac{(\alpha_{s_i} - \bar{\alpha}_s)}{\sigma_i^2}$
- $\bar{\alpha}_s = \sigma^2_{tot} \sum_i \frac{\alpha_{s_i}}{\sigma_i^2}$, $\frac{1}{\sigma^2_{tot}} = \sum_i \frac{1}{\sigma_i^2}$
- Large Tension: $\frac{\tilde{\chi}^2}{dof} \simeq 17$
- **Yet small uncertainty: $\boldsymbol{\bar{\alpha}_s = 0.1179 \pm 0.000329}$**
- The small difference with α_s given in plot shows good agreement with quadratic approximation.



	DIS	DY	Jet +$t\bar{t}$	$\bar{\alpha}_s$
$\alpha_s \times 10^3$ at min of χ_i^2	115.9	119.9	118.9	**117.9**
Error ($\sigma_i \times 10^3$) ($\Delta\chi_i^2 = 1$)	0.553	0.655	0.539	**0.329**

How to handle this situation?

GMM and Scaling Errors

Gaussian Mixture Model

- PDG proposal: scale errors by a factor e_s to make fits more consistent, i.e. each $\sigma_i \to e_s\, \sigma_i$
- $e_{s_{PDG}} = \sqrt{\frac{\tilde{\chi}^2}{dof}} \simeq 4.1$ so that each $\sigma_i \to 4.1 \times \sigma_i$ and $\frac{\tilde{\chi}^2}{dof} \to 1$
- Caveat: For very large $\sqrt{\frac{\tilde{\chi}^2}{dof}}$, PDG recommends making an educated guess of the uncertainty rather than scaling the errors.
- Alternate: use GMM and Information criteria to determine if the posterior (probability of theory given data) is better modeled with a uni-modal (usual χ^2) or by a multi-modal distribution. Then use the best distribution (uni-modal or multi-modal) to determine the uncertainty.

K=1

K>=2

arXiv: 2406.01664

How many Gaussians? How do we determine K?

Akaike Information Criterion (AIC)
(Akaike, 1974)
Bayesian Information Criterion (BIC)
Schwarz (Ann Stat 1978, 6:461–464)

$$\begin{aligned} \mathrm{AIC} &= N_{\mathrm{parm}} \log N_{\mathrm{pt}} - 2\log L\big|_{\theta=\hat{\theta}}, \\ \mathrm{BIC} &= 2N_{\mathrm{parm}} - 2\log L\big|_{\theta=\hat{\theta}}. \end{aligned}$$

$$N_{\mathrm{parm}} = 2K + (K-1).$$

Use the lowest values of AIC & BIC to determine the best value of K and avoids over-fitting.

Strong tension

Weak tension due to large uncertainty

Consistent but data fluctuated

Consistent - No fluctuation

		$K=1$	$K=2$	$K=3$	$K=4$
case-1	AIC	-102.2	**-203.6**	-194.9	-187.9
	BIC	-106.1	**-211.2**	-206.4	-203.2
N_{pt}=100	$-\log L$	-55.0	**-109.6**	-109.2	**-109.6**
case-2	AIC	**-21.2**	-15.4	-7.9	-0.2
	BIC	**-25.0**	-23.0	-19.3	-15.5
N_{pt}=100	$-\log L$	-14.5	-15.5	**-15.7**	**-15.7**
case-3	AIC	-219.3	**-220.2**	-212.8	-205.0
	BIC	-223.2	**-227.8**	-224.3	-220.3
N_{pt}=100	$-\log L$	-113.6	**-117.9**	**-117.9**	-118.1
case-4	AIC	**-117.8**	-109.9	-102.1	-94.3
	BIC	**-121.6**	-117.6	-113.6	-109.6
N_{pt}=50	$-\log L$	**-62.8**	**-62.8**	**-62.8**	**-62.8**
case-5	AIC	**-169.3**	-161.5	-153.6	-145.8
	BIC	**-173.1**	-169.1	-165.1	-161.1
N_{pt}=50	$-\log L$	**-88.6**	**-88.6**	**-88.6**	**-88.6**

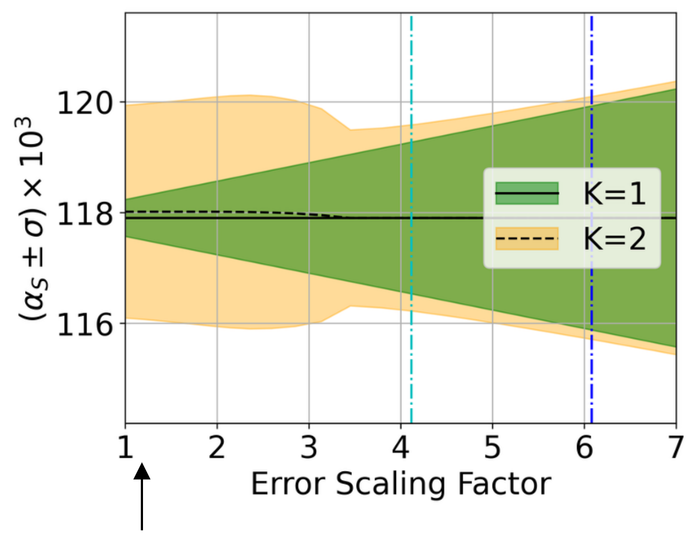
$$\pi(Y|\vec{\theta}) = \prod_{j=1}^{N_{\mathrm{pt}}} \pi(y_j, \Delta y_j|\vec{\theta}) = \prod_{j=1}^{N_{\mathrm{pt}}} \sum_{i=1}^{K} \omega_i \mathcal{N}(y_j, \Delta y_j|\theta_i),$$

$$0 \le \omega_k \le 1 \quad \text{and} \quad \sum_k \omega_k = 1,$$

Preliminary results

# of modes	Loss	AIC	BIC
K=1	18.09	37.27	38.17
K=2	**3.69**	**10.67**	**13.37**
K=3	14.29	21.82	26.33
K=4	14.8	24.02	30.33

- GMM (K=2) (Yellow shaded)
  - $\alpha_s \pm \sigma = 0.11801 \pm 0.00192$
- $e_{s_{PDG}} \sim 4.1$ *(Green shaded, Cyan line)*
  - $\alpha_s \pm \sigma = 0.11795 \pm 0.00135$
- $e_s \sim 6.1$ (Green shaded, Blue line)
  - $\alpha_s \pm \sigma = 0.11795 \pm 0.0020$
- Caveat: This is just preliminary. How we partition the data sets does have an impact on uncertainty determination. More complete study is underway



Stay tuned to this channel

using a nominal tolerance for results with tensions as seen in the α_s determinations would result in a large underestimate of the uncertainty

Dynamic tolerances

Dependence on **tolerance** (uncertainty definition)

PRELIMINARY

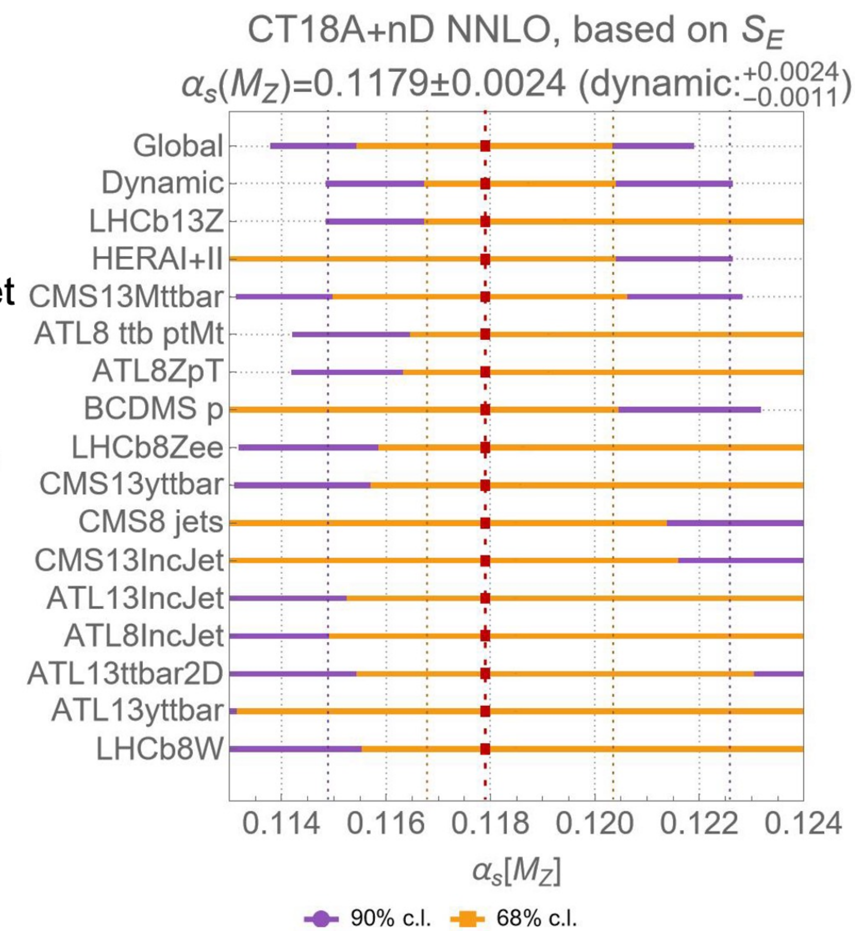
PDF uncertainties are commonly defined based on two hypothesis-testing criteria:

1. Global (CTEQ, 2001):
   here derived from a χ^2/N_{pt} quantile of the whole data set not considering PDF parametrization dependence
2. Dynamic (MRST, 2008):
   derived from $\chi_i^2/N_{pt,i}$ quantiles of individual experiments

We also have a dynamic tolerance implemented

In the CT25prel NNLO analysis, global and dynamic prescriptions predict comparable $\alpha_s(M_Z)$ uncertainty

The dynamic uncertainty is positively skewed due to preferences of multiple LHC experiments



Dynamic tolerance and clustering of data sets

Interesting to examine the dependence of dynamic T^2 on the clustering/partition of experimental data sets and credibility level (c.l.)

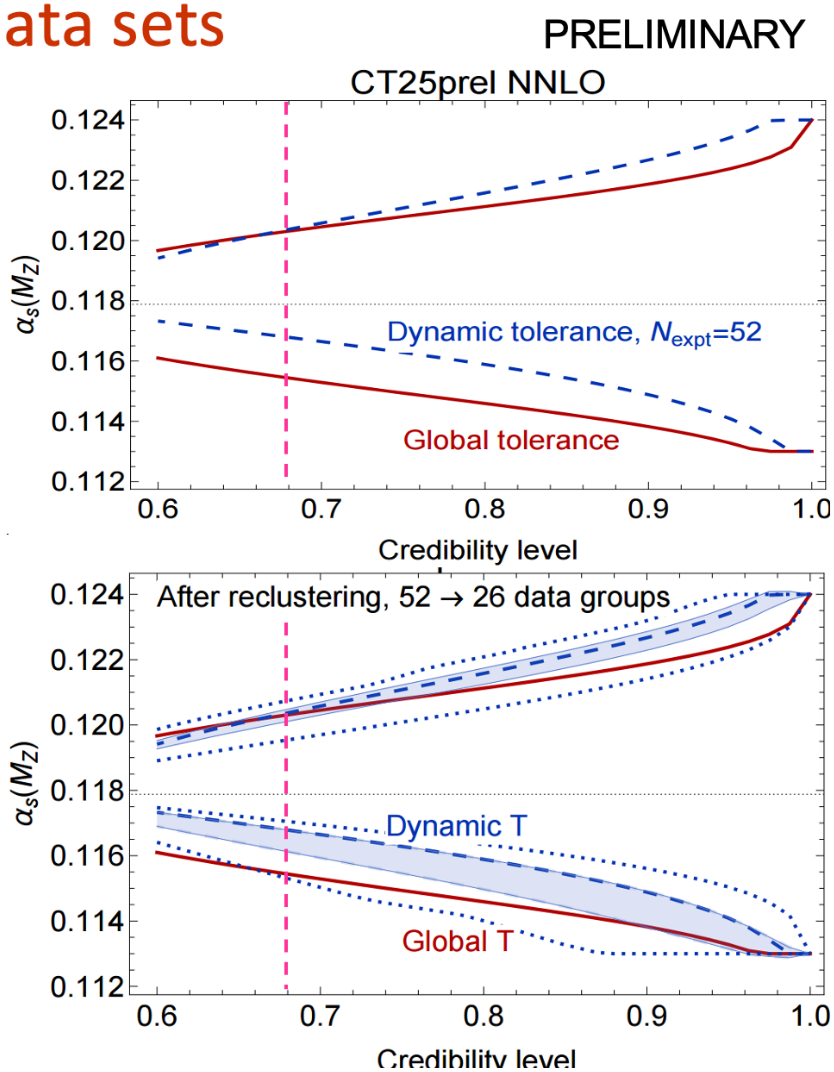
The default dynamic T^2 is for the whole set of 52 fitted experiments

Combinations of data sets generally bring dynamic T^2 closer to the global T^2, which does not require clustering

Here illustrated by comparing the default dynamic T^2 with the one for 1000 random clusterings into 26 pseudoexperiments

Dashed: default dynamic T^2
Blue band: for 68% of 1000 26-experiment clusters
Dotted: for the envelope of 26-experiment clusters

Grouping of data sets can substantially change the dynamic uncertainty



A few words about PDG

- A non-lattice result was determined from sub-fields 1-6 using a χ^2-averaging method

$$\alpha_s(M_Z^2) = 0.1175 \pm 0.0010\,, \qquad \text{(without lattice)}$$

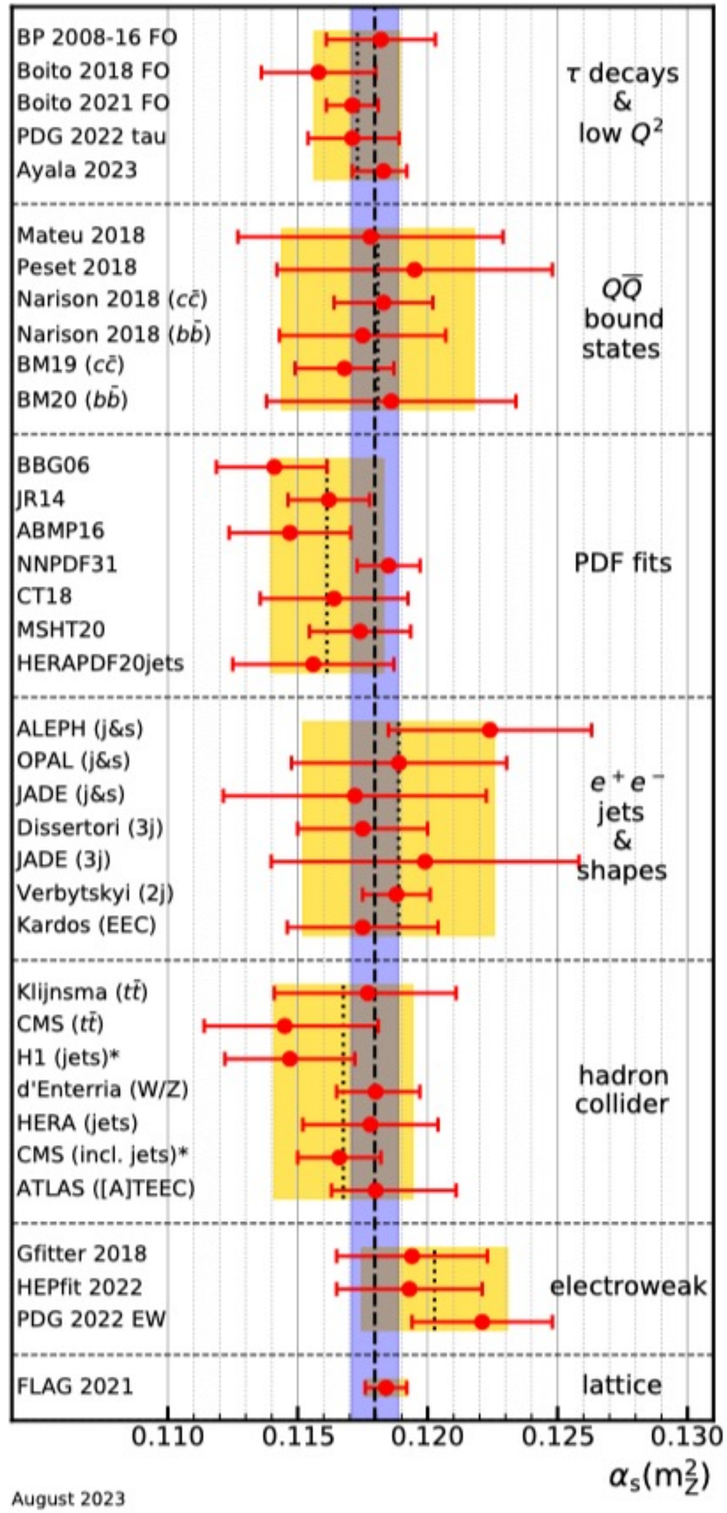
$$\overline{0.1178 \pm 0.0005} \qquad \text{weighted}$$

- FLAG result itself is an average and is taken as is

$$\alpha_s(m_Z^2) = 0.1184 \pm 0.0008 \qquad \text{(lattice)}$$

- **Note that the uncertainty for the data-driven determination is similar to that from lattice; lattice error will come down faster than non-lattice**
- Combine two numbers in un-weighted average, and take uncertainty as an average of the two uncertainties (conservative)

$$\alpha_s(m_Z^2) = 0.1180 \pm 0.0009 \qquad \text{(PDG 2023 average)}$$



PDG value of $\alpha_s(m_Z)$

- There was discussion on the grouping of the measurements used in the determination
- We've been considering the philosophy shown below

from 2023 review

In this combination, as in past combinations, we have considered lattice QCD calculations of α_s independently of experimental/phenomenological determinations. In the future, when the lattice continuum extrapolations are under better control, it may be useful to group lattice QCD determinations of α_s with experimental determinations of α_s that have systematics of similar origin, in a similar manner as we currently group, for example, hadron collider results together [723].

[723] L. Del Debbio and A. Ramos, Physics Reports **920**, 1 (2021), ISSN 0370-1573, [arXiv:2101.04762].

PDG value of $\alpha_s(m_Z)$

- Every two years, the QCD section in the Particle Data Book is updated; part of that update is a review of the world average of $\alpha_s(m_Z)$, revising it to include the impact of new measurements and calculations
- The last revision was in 2023; which means I'm going to have a busy summer this year
- The selection of results to include in the α_s averaging are restricted by the following considerations:
  - published in a peer-reviewed paper at the time of the report (or is based on a summary of results that have been published in a peer-reviewed journal, such as the FLAG report)
  - based on the most complete perturbative predictions of at least NNLO accuracy, accompanied by reliable estimates of all experimental and theoretical uncertainties

Summary

- The LHC has been very productive with high statistics precision data becoming available at 13 TeV, useful for PDF and $\alpha_s(m_Z)$ determination
- We have made use of much of this new data in our development of CT25
  - some choices have to be made, for example inclusive jets over dijets
- Tensions exist between and within data sets requiring a good understanding of the tolerance needed to provide a robust estimate of the PDF uncertainty
  - we are pursuing multiple paths for this exploration
- There are opportunities to involve professional statisticians, e.g. PDF4LHC in conjunction with PHYSTAT

Extras

Effects of different nonperturbative parametrization of g-PDF

As an alternative parametrization of the g-PDF, we adopted a polynomial in $y = x^c$, with a fixed value of $c = 0.23$, and

Case (4983)

$$g(x, Q = Q_0) = a_0\, x^{a_1 - 1} (1 - x)^{a_2}\, P_a^g(y), \tag{A9}$$

$$P_a^g(y) = e^{a_3}(1 - y)^5 + \sinh[a_4] 5y(1 - y)^4 + \sinh[a_5] 10y^2(1 - y)^3 + \sinh[a_6] 10y^3(1 - y)^2 + a_7 5y^4(1 - y) + y^5, \tag{A10}$$

in which the a_7 parameter is related to a_1 and c as

$$a_7 = 1 + (a_1 - 1)/(5c). \tag{A11}$$

CT18NNLO

Case (848)

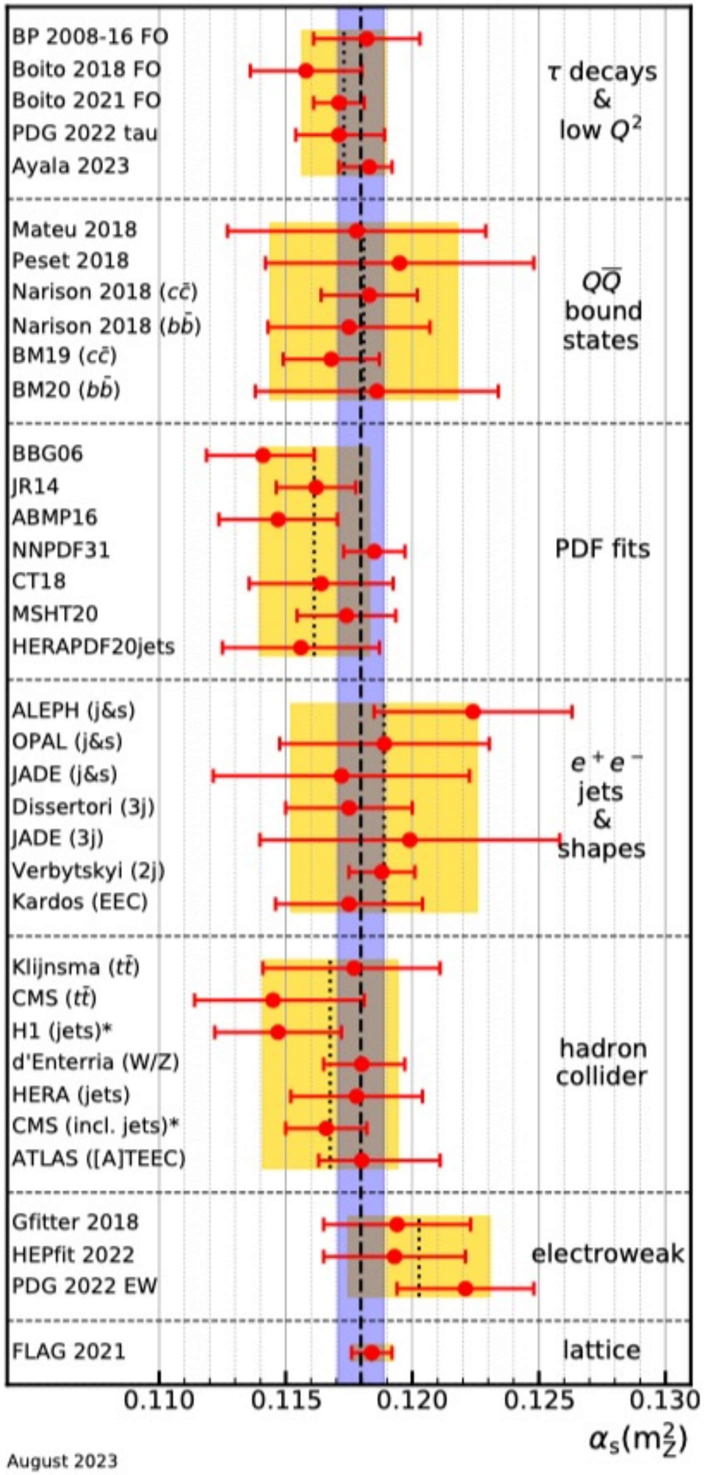
$$g(x, Q = Q_0) = a_0\, x^{a_1 - 1} (1 - x)^{a_2}\, P_a^g(y), \tag{A5}$$

$$P_a^g(y) = \sinh[a_3](1 - y)^3 + \sinh[a_4] 3y(1 - y)^2 + a_5 3y^2(1 - y) + y^3,$$

PDF fits often do not have explicit estimate of theory uncertainty

issue of whether necessary to do simultaneous PDF fit; * == fit

new version of ePump will allow simultaneous fit of $\alpha_s(m_Z)$ with new data



$$\alpha_s(m_Z^2) = 0.1173 \pm 0.0017,$$

$$\alpha_s(m_Z^2) = 0.1181 \pm 0.0037,$$

unweighted averages of central value and uncertainties within sub-fields; impact of correlations reduced

$$\alpha_s(M_Z^2) = 0.1161 \pm 0.0022,$$

$$\alpha_s(m_Z^2) = 0.1189 \pm 0.0037.$$

some results removed due to underestimate of NP uncertainties

$$\alpha_s(m_Z^2) = 0.1157 \pm 0.0021$$

$$\alpha_s(m_Z^2) = 0.1168 \pm 0.0027,$$

if use only H1 and CMS, which used PDF fit

$$\alpha_s(m_Z^2) = 0.1203 \pm 0.0028,$$

combination of first 6 pre-averages using χ^2 averaging gives

$$0.1175 \pm 0.0010$$