ATLAS data to be added to ATLASpdf21 in future—

Things to be implemented in xFitter

A M Cooper-Sarkar

xFitter 2023 meeting

1. The 13 TeV $W^{\pm}D^{(*)-}$ data arXIV:2302.00338
2. The 13 TeV direct photon data arXIV:2302.00510
3. The NEW 13 TeV jet data, and inclusion of previous 7,8,13TeV jet data sets simultaneously
4. The NEW13 TeV t-tbar data
5. The NEW13 Tev Drell-Yan data– but perhaps we also need to reconsider the 7 and 8 TeV Drell-Yan data we have already in light of revised predictions
9. And think about Zpt fits? (5,8,13 and low μ)

Then there is data coming in Run 3?---

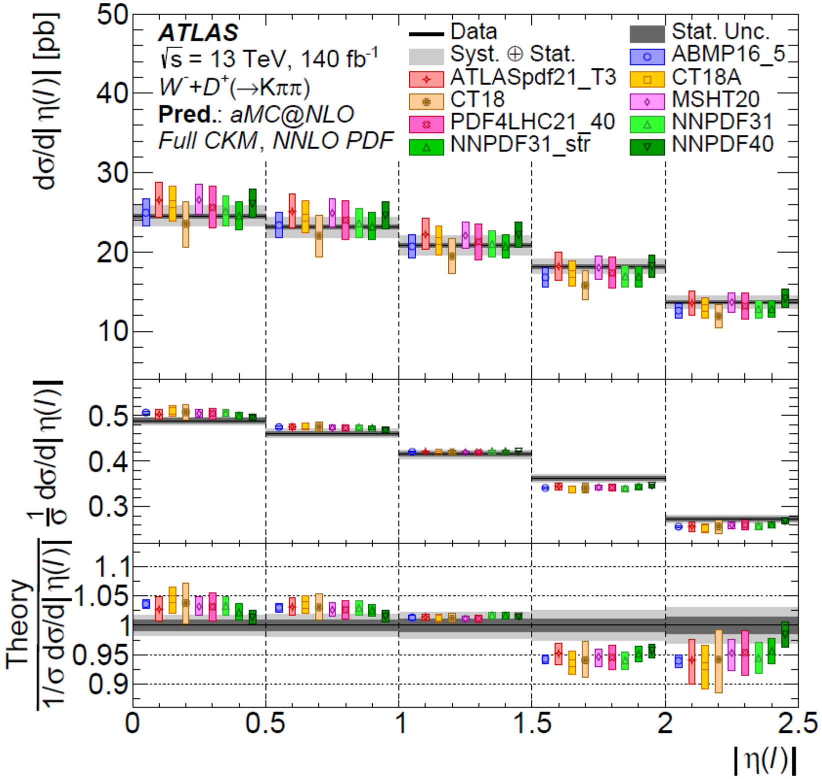
And what about data that is relevant to intrinsinc charm? Z + c /Z +jets at high Z rapidity

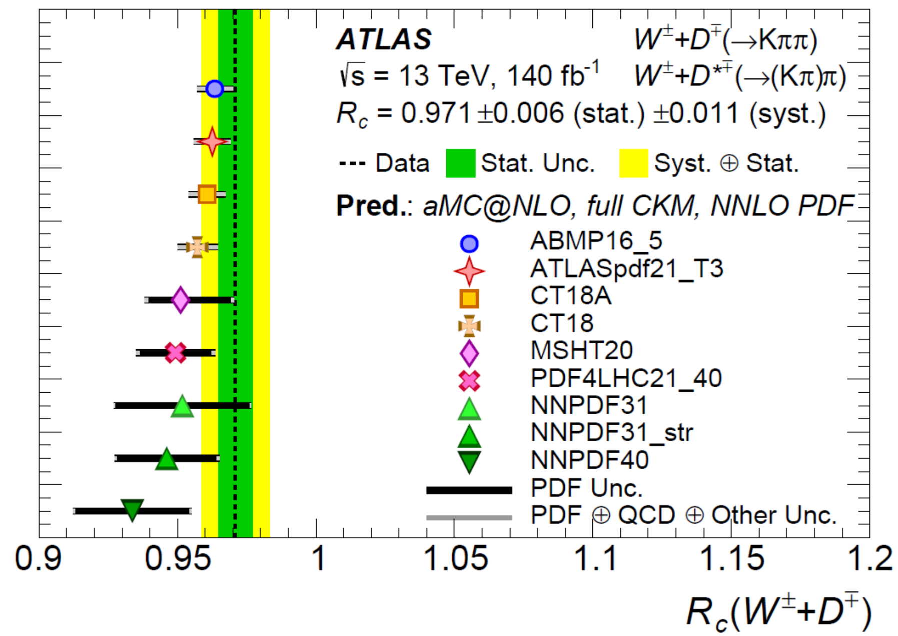
The 13 TeV $W^{\pm}D^{(*)-}$ data ArXIV:2303.00338

Analysers have provided data in our preferred format with information on correlated systematic uncertainties

We have checked this out using ‘toy’ predictions from current PDFs

What we lack is full NNLO predictions which can be put on grids or used for k-factors. Time-scale –a few more months??





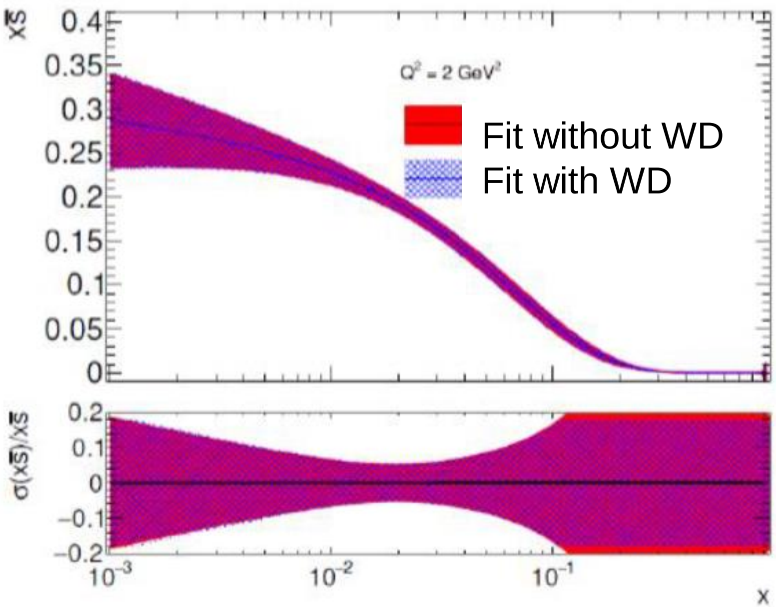
Of course W+c is also on its way, but I have no plots

Partial Chisq are very good
13TeVW-D+ = 5 3.6033804454720277
13TeVW+D- = 5 3.4537515561879295
13TeVW-D*+ = 5 3.2180920173295924
13TeVW+D*- = 5 1.6836593249830210
wd13 11.958883343972570
Sum over nuisance parameters squared
sumsqwd= 24.428910530465984

Toy exercise with ATLASpdf21 predictions treated as truth, no grid predictions

So the total chisq for the WD data partial+sumsqwd = 36.35 for 20 points
Whereas this is not so great I am not sure one can do much better than a sum over nuisance parameters squared of 24 for 270 sources of nuisance parameter!!

Recent tendency to provide HUGE numbers of correlated systematics, probably better if confined to those which are large



It fits very well and has little effect because the W+D results are fully compatible with the ATLAS strangeness

To DO: Expand predictions to accommodate s ≠ sbar

The 13 TeV direct photon data

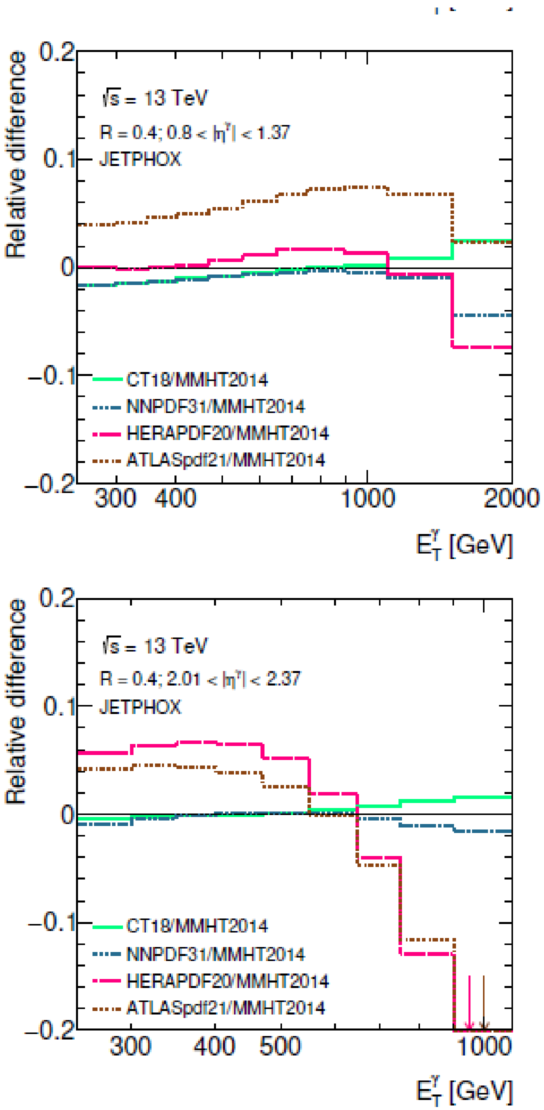
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2022-065/

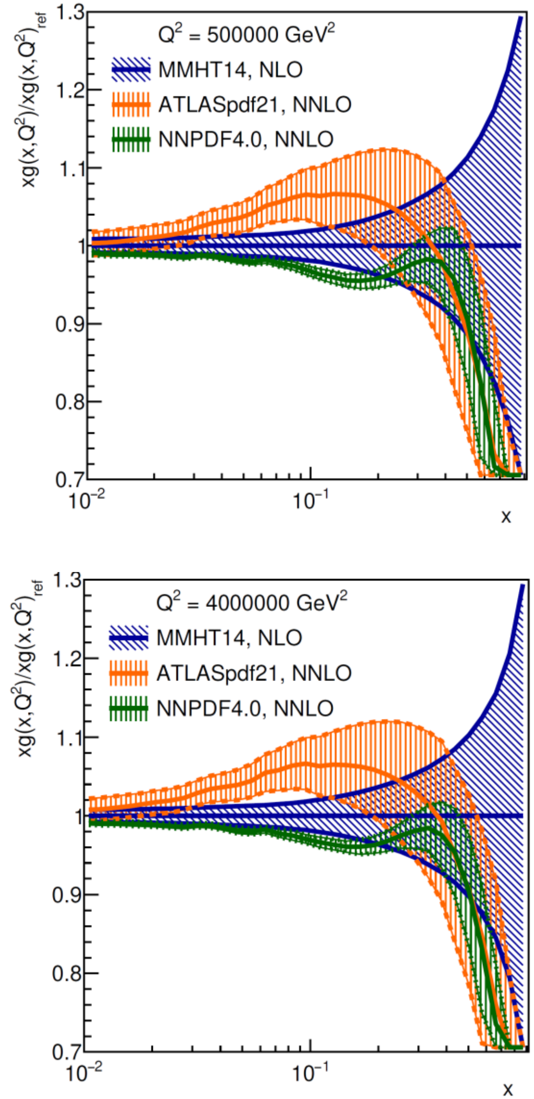
It is clear that these data as ~5% different from ATLASpdf21 predictions at x~0.1—this is a region where we have little other data to influence our fit.

The analysers have not yet provided HEPDATA format tables, but they are aware of what information is needed

We also need predictions- since the bins have changed since the earlier data (they are split-up more). These could come from
EITHER NLO*k-factors as last time– we have already contacted J.Rojo on this—but it is unlikely to come before data are published.
OR direct NNLO grids from NNLOjet group—again they have been contacted but this is not at the top of their ‘to-do’ list

Time scale this year





Comparisons of PDF predictions from the CONF NOTE

Explanation in terms of our high-x gluon PDF at high scale– we always did admit to 2σ differences in this region

The new 2017/2018 13 TeV jet data– this is coming along 'soon'

So far on the jet data we already **published 7 TeV inclusive and dijet, 8 TeV inclusive (used in ATLASpdf21) and earlier 13 TeV inclusive and dijet from 2015/2016**

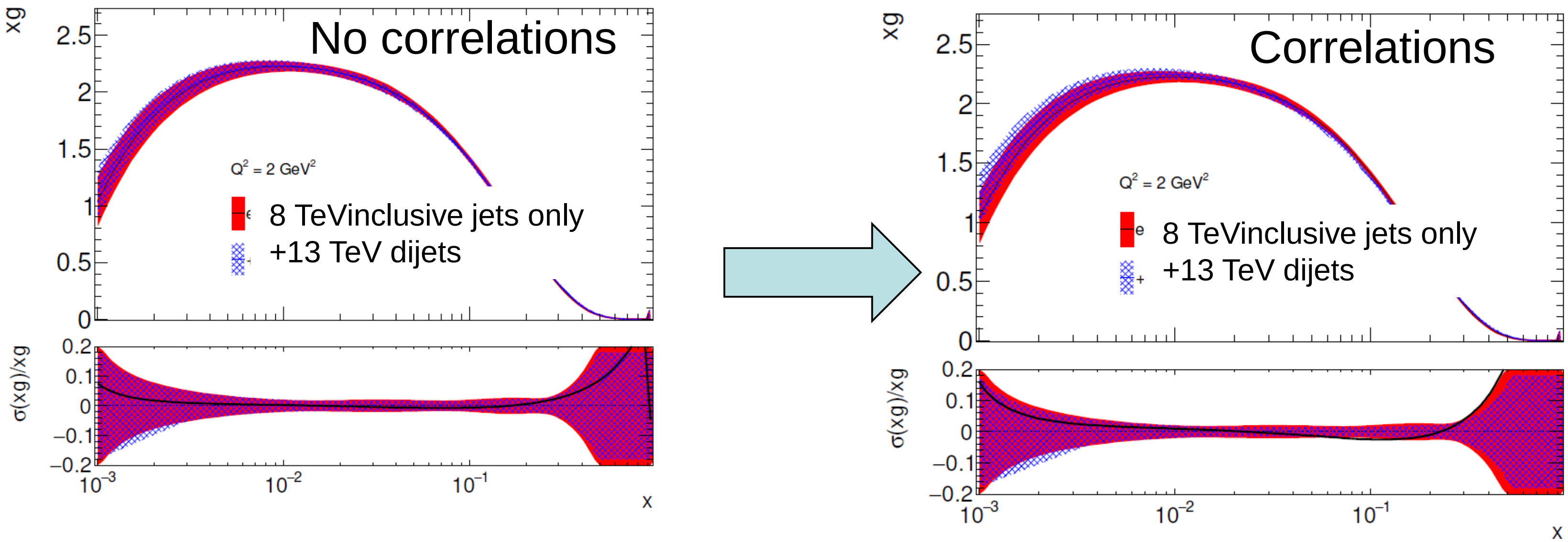
The issue was that there are systematic as well as statistical correlations between these data (ie for inclusive and dijet at the same cm energy) such that only one data set could be used at once.
This has been clarified internally for the 8 and 13 TeV jet data sets and indeed a toy study of 8 TeV inclusive and 13 TeV dijets has established that there would be some gain in precision of the high-x gluon from using both
This study has used direct NNLOgrids from NNLOjet but these are not yet of sufficient precision, work to improve this is ongoing.

When the new 13 TeV data are available (this year) we should also use them– **awareness of the need for clarity in the systematic correlations is well developed in this sub-group.**

But there is a warning. To implement systematic correlations between the data sets requires combination of the matrices which encapsulate this in formation in the fit. These are becoming very large and slowing down fits

The Effect of having included the systematic correlations between 8 TeV inclusive and 13 TeV dijets is best seen on the gluon

Plots show 8 TeV inclusive alone in red
To 8 TeV inclusive +13 TeV dijet in blue, without (left) and with (right) systematic correlations



Gluon is a bit harder and has reduced high-x uncertainty
Interim conclusion: systematic correlations are worth considering

To explain my point about the size of the matrices, straight from the xFitter manual…

From the statistical point of view a measurement result, μ_i, at point i is a random variable which can be modelled as

$$\mu_i = m_i(\boldsymbol{p}) + r_i\sigma_i + \sum_{\alpha=1}^{N_{\text{syst}}} \Gamma^i_\alpha b_\alpha \qquad (47)$$

where

$m_i(\boldsymbol{p})$ is the 'true', physical model value depending on parameters $\boldsymbol{p} = (p_1, p_2, \dots)$,

Moreover, the relative rather than absolute uncertainties are used, e.g. $\Gamma^i_\alpha = \gamma^i_\alpha \mu_i$ or $\sigma_i = \sqrt{\delta^2_{i,\text{stat}} + \delta^2_{i,\text{uncor}}}\,\mu^i$.

$$\chi^2_{\text{exp}}(\boldsymbol{m}, \boldsymbol{b}) = \sum_i \frac{\left[m^i - \sum_\alpha \gamma^i_\alpha \mu^i b_\alpha - \mu^i\right]^2}{\left(\delta_{i,\text{stat}}\,\mu^i\right)^2 + \left(\delta_{i,\text{uncor}}\,\mu^i\right)^2} + \sum_\alpha b^2_\alpha. \qquad (50)$$

Two ways to go to implement this

In the first method a minimisation of Eq. 50 wrt. b_α is used to define the covariance matrix for the systematic uncertainties, which is determined as

$$C_{\text{syst}\ ij} = \sum_\alpha \Gamma^i_\alpha \Gamma^j_\alpha\,. \qquad (51)$$

Add this to your statistical correlation matrix (diagonal in eq 50) to get $C_{tot\ I,j}$ and minimise

$$\chi^2(\boldsymbol{m}) = \sum_{ij} (m^i - \mu^i)\, C^{-1}_{\text{tot}\ ij}\, (m^j - \mu^j)\,. \qquad (53)$$

The point is that the C matrix to be inverted is never larger than the number of data points (squared)

The second method is used to determine optimal shifts of the nuisance parameters at each iteration. The minimisation wrt. $\boldsymbol{b}$ leads to a system of linear equations

$$\sum_{\beta}\sum_{ij} C^{-1}_{\mathrm{stat}\ ij}\Gamma^i_\alpha\Gamma^j_\beta b_\beta + b_\alpha = \sum_{ij} C^{-1}_{\mathrm{stat}\ ij}\Gamma^i_\alpha(m^j - \mu^j)\,, \tag{54}$$

where, $1 \leq \alpha \leq N_{\mathrm{syst}}$, the total number of correlated systematic uncertainties. The methods given by Eq. 53 and Eq. 54 are equivalent algebraically but Eq. 54 is more efficient numerically when the number of nuisance parameters is smaller than the number of measurements.

To implement this you will need to invert a matrix of the size number of systematic uncertainties (squared). **When this is getting larger than the number of data points this is NOT efficient, in fact it is making jobs take several days!!**

BUT WHY DO IT THIS WAY?

Well, we have found for several data sets –jets, t-tbar– that the degree of correlation between data points has been overestimated– it is either 100% or 0% and the fashion is for 100%. We have introduced some decorrelations in fitting (NOT just us but all the global fitters) ATLAS have been more sophisticated and not gone from 100 to 0% immediately but have often introduced a sliding scale of correlation according to rapidity or pt for example.

The upshot is that this alters the corresponding systematic covariance matrix.

If all we have been given is such a matrix then we cannot alter it meaningfully if there are problems. The separate systematic sources are needed to get the information as to what to do.

In future, we may need to implement the fit using the resulting new covariance matrices

The 2017/2018 13 TeV t-tbar data— coming soon—this year

Considerable work has been done to reduce the larger systematic uncertainties
e.g. parton shower systematics for lepton+jets
Such that these data should have a larger impact than the previous t-tbar data

Extended kinematic range and 2-D and 3-D distributions are also coming

The group are aware of the need for full information on systematic correlations
(and backward compatibility if the data are to be used together with earlier t-tbar data)

The analysers are in touch with the theoreticians (Mitov et al) who provide the predictions—they are also using MATRIX– so this part of the analysis should be quicker—see Joey's recent ATLAS talk…..

NNLO Grids with MATRIX and NNLOJET

ATLAS PDF forum meeting 24/02/2023

Yuval Kay
University College London,
Max Planck Institute for Physics

Fazila Ahmadova
Max Planck Institute for Physics, Munich
LMU Munich

Daniel Britzger
Max Planck Institute for Physics

We stimulated (and tested) the implementation of MATRIX PineAPPL grids as well as implementing NNLOJET grids.

Task to be done:

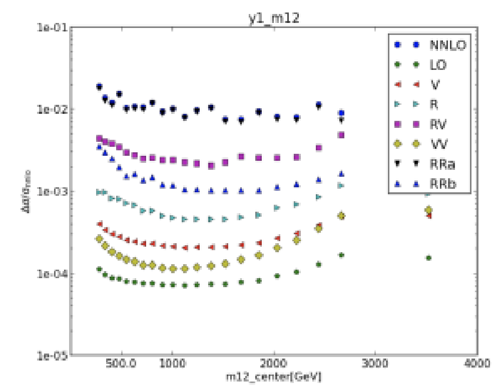
Interfacing MATRIX to fast-interpolation tools i.e. APPLgrid

...with STA Alex Huss, have been generating NNLOJET grids for 13 TeV inclusive jet and dijet production to fully exploit ATLAS jet data, understand better issues with scale choice and systematic errors for the two related processes

NNLOJET v2

- Grid interface (APPLgrid, fastNLO) for NNLOJET v2
  - Updated grid interface was developed and is currently tested (M. Sutton, C. Gwenlan, D. Britzger, F. Ahmadova, L. Kunz, K. Rabbertz, X. Chen, G. Heinrich, A. Huss. J. Pires)
  - First dijet grids (with full color) were produced (right) → Promising performance (CPU time, closure, stat. uncertainties, etc...)
  - Shown on the right: relative stat. uncertainties of individual contributions to NNLO cross section of the first grid production.

ATLAS 7TeV, dijet, R=0.6, $d\sigma/dm_{12}$



The 13 TeV Drell-Yan data—coming soon

For high-mass 13 TeV Drell-Yan

beware we would need photon PDF within proton formalism OR is this NOW considered sufficiently well known now that is can be reliably subtracted?

For low-mass 13 TeV Drell-Yan

The data are differential in mll and yll and also in ptll

The latter can only be fitted with q_T resummed calculations

Mll and yll can in principle use fixed order QCD, where NNLO is essential but may not be enough…

The data start to probe such low Mll values that I would expect ln(1/x) resummation calculations would be needed—do these exist for Drell-Yan?

If they do it would make sense to marry them to ln(1/x) resummation for the HERA data. We used to cut Q2>10 GeV2 to avoid the problematic region, but it is not clear that this is a good idea!

The fit makes an even worse approximation (to the effect of adding higher-order or ln(1/x) resummation) when it does not have the information ! Fitting down to Q2=3.5 GeV^2 could actually be better.

There is also the question of using N3LO here…

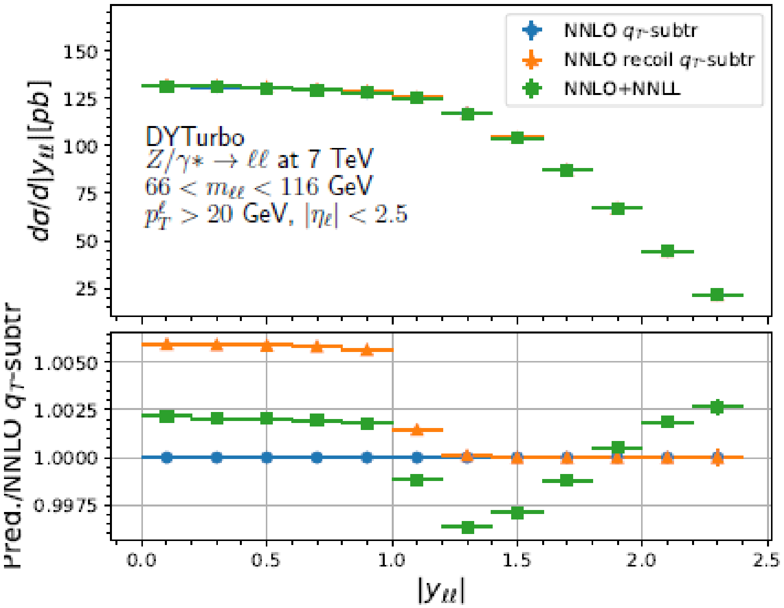
But then only part of the data can be fitted to N3LO right now…

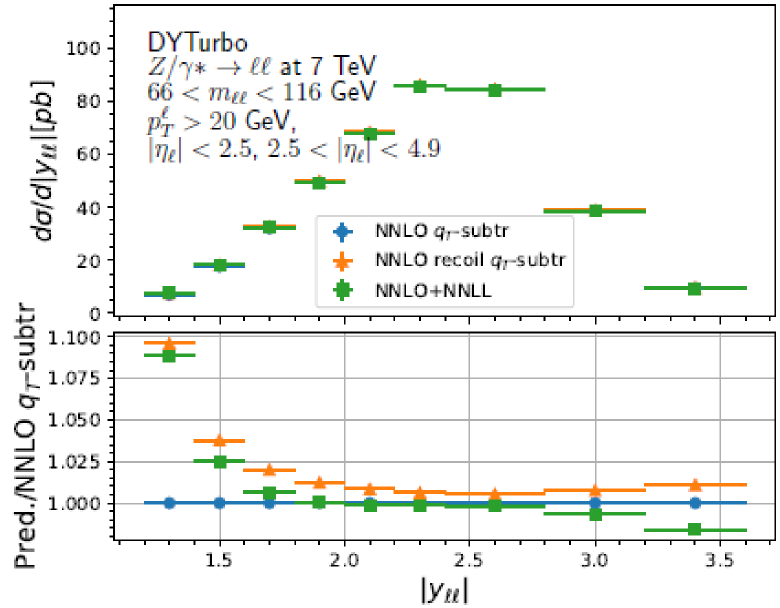
Drell-Yan data are the most accurate LHC data and also those for which the most sophisticated predictions exist
In ATLASpdf21 we accounted for scale variations for the 7 and 8 TeV W and Z data for this reason. But we know N3LO does not lie in these bounds.
Could we go further JUST for these data sets?
AND
There is not just N3LO, there is N4LO +N3LL resummation predictions

We have long been told that we could improve on the predictions for the precision W,Z 7 TeV data by using k-factors for parton shower resummation
We are still waiting fo these k-factors!

This seems to affect 7 TeV much more than 8 TeV because 7 TeV had particularly unfortunate symmetric fiducial cuts.
The highly differential data of the 'Z3D' are only used in regions of very high fiducial acceptance for which bins do not wander into a region for which NNLO predictions are effectively only NLO (because leading order is zero).

Let's look at possible improvements for the 7 TeV Z rapidity spectra Central-central and central-forward….





CT14nnlo 68%CL

Dataset	NNLO q_T-subtr.	NNLO recoil q_T-subtr.	NNLO+ NNLL
W^+ lepton rapidity	9.4/11	8.8/11	8.8/11
W^- lepton rapidity	8.2/11	8.7/11	8.2/11
Low mass, Z rapidity	11/6	7.2/6	7.5/6
Mass peak, central Z rapidity	15/12	10/12	7.7/12
Mass peak, forward Z rapidity	9.6/9	5.3/9	6.4/9
High mass, central Z rapidity	6.0/6	6.5/6	5.8/6
High mass, forward Z rapidity	5.2/6	5.6/6	5.3/6
Correlated χ^2	40	40	31
Log penalty χ^2	-4.33	-3.39	-4.20
Total χ^2/dof	99/61	88/61	77/61
χ^2 p-value	0.00	0.01	0.08

NNLL refers to q_T resummation

There is also new Zpt data

At 5 TeV, at 13 TeV with low μ as well as the 8 TeV data that we already have.

We chose to fit Z+jets rather than Zpt at 8 TeV. Would considering 8 TeV Zpt bring more? –

In fact I think not …

I have fitted Zpt instead of Zjets 8 TeV in the ATLASpdf21 fit

There are choices as to how much Zpt data to fit

CT chose 45 < pt < 150 (low pt non perturbative effects? High pt EW corrections?)

NNPDF chose 30 < pt <150

MSHT chose 30<pt< 550

They also differ as to their estimates of uncertainty on NNLO/NLO k-factors

They also differ as to how much off-mass-peak data to use—but this does not matter much, off peak data is well fitted

I have mimicked these fits using the smallest NNLO/NLO k-factor uncertainties from MSHT and MSHT generated NLO grids

I get χ2/ndp ~1.0, 1.5, 1.5 for CT, NNPDF, MSHT style fits

As compared to 1.0,1.0, 1.9 in their own global fits

The big jump in chisq comes from the 30< pt < 45 bins for on-mass-peak data

These are poorly fitted at NNLO. It turns out that these are well fitted for MSHTaN3LO—perhaps not surprising?

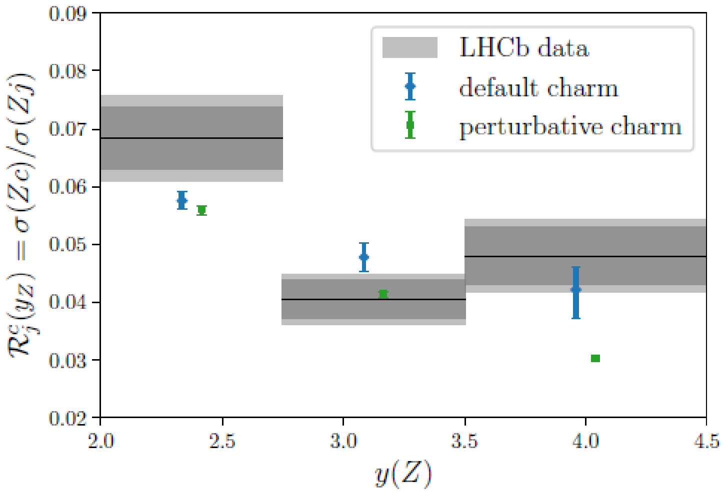
For the good fit to Zpt --CT style-- the resulting PDFs are ~ the same as those

Obtained when using Z +jets--- no motivation to change

And what about instrinsic charm

Clearly you need Z+c or γ+c

It's probably smart to take ratios like Z+c/Z+jets and do it at high rapidity to reach high x



NNPDF say 3σ evidence for ~0.5%...CTEQ says NO because PDF uncertainties are larger

But there is trouble on the horizon. This paper suggests that we do not really know what we are doing with charm jets

NNLO QCD predictions for Z-boson production in association with a charm jet within the LHCb fiducial region arXiv:2302.12844v1 [hep-ph] 24 Feb 2023

R. Gauld,[a] A. Gehrmann–De Ridder,[b,c] E. W. N. Glover,[d,e] A. Huss,[f] A. Rodriguez Garcia,[b] G. Stagnitto[c]

IRC sensitivity of common jet algorithms (antikT) requires the use of new flavour jet algorithms for predictions

Flavour jet algorithms not directly useable at the LHC; effort needed to unfold from one definition to the other

->large program planned for this year's Les Houches

Concluding remarks

All in all it seems to me that we could have quite a significant update within ~the next year. We need an xFitter implementation of everything

I can maybe see it for jets and t-tbar………NNLOjet/MATRIX

I don't see it for most of the rest..

Predictions are a nightmare, have to beg for them from those who have done them

Ploughshare: ploughshare.web.cern.ch
has a LOT of jets, some W+/- and Z, some W+jets, not much t-tbar
What about direct photon, Zpt, W+D, Z+c, D-Y beyond NNLO??

HighTea (High-energy **Theory Event Analyser) f**rom Mitov et al—not many processes yet.. www.precision.hep.phy.cam.ac.uk/hightea/ (arXiv:2304.05993)
But here you have to do the work yourself using their tools