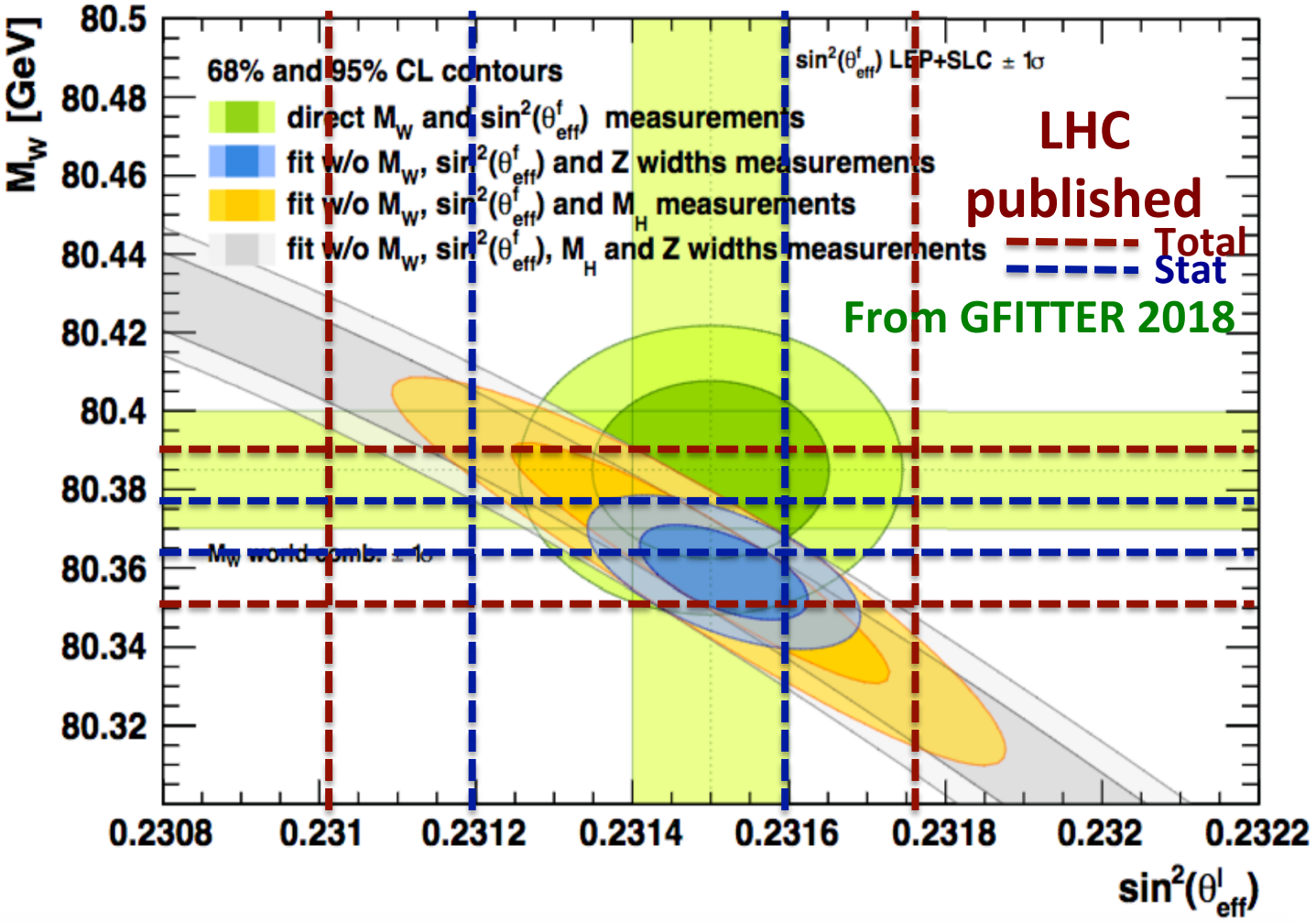
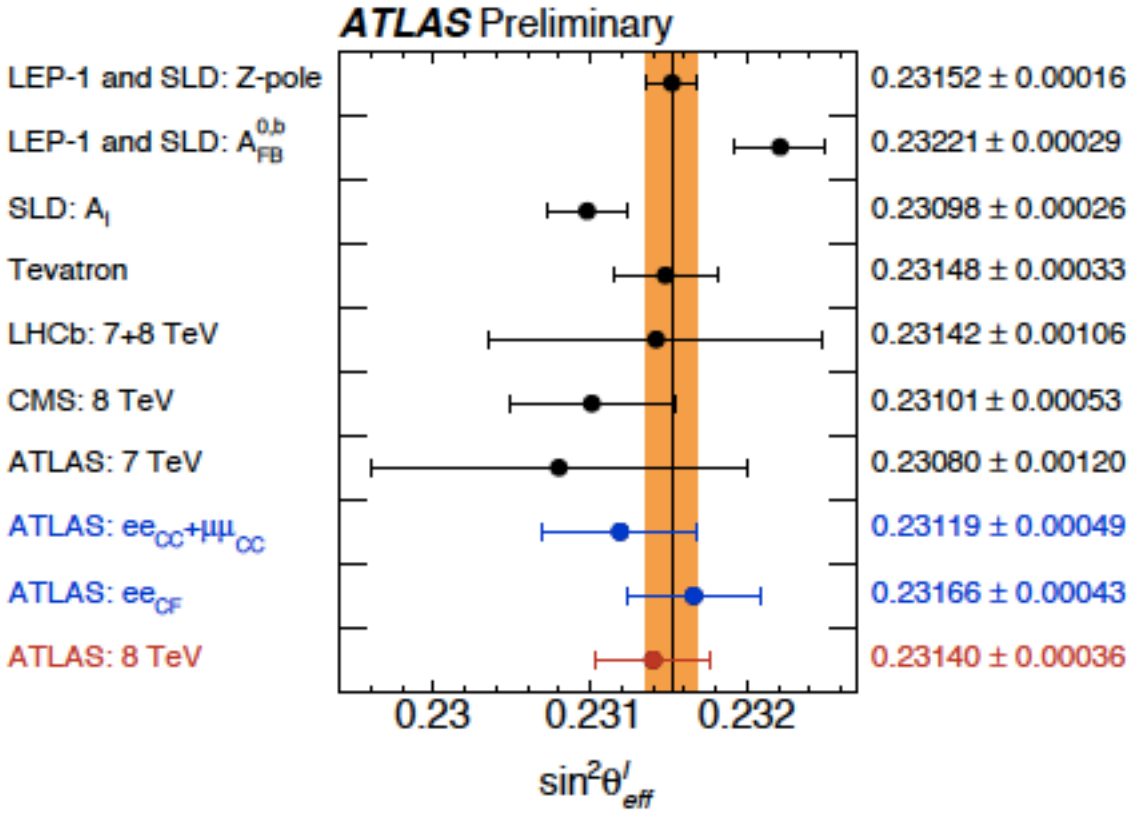
PDF benchmarking proposal for precision DY



- One of the ATLAS goals is to produce "proper" ellipse in this plot
- Currently, direct measurements above are uncorrelated
- LHC measurements are correlated primarily through PDFs

ATLAS measurement of $\sin^2\theta^l_{eff}$ at 8 TeV

$\sin^2\theta^l_{eff} = 0.23140 \pm 0.00021\,(\text{stat.}) \pm 0.00024\,(\text{PDF}) \pm 0.00016\,(\text{syst.})$



Links to papers and to preliminary result on $\sin^2\theta^l_{eff}$:

- Angular coefficients in Z-boson decays at 8 TeV
- Triple-differential measurements of Z-boson decays at 8 TeV
- Measurement of $\sin^2\theta^l_{eff}$ at 8 TeV

$\sin^2\theta^l_{eff}$ results based on reference PDF set (MMHT14)

Channel	ee_{CC}	$\mu\mu_{CC}$	ee_{CF}	$ee_{CC} + \mu\mu_{CC}$	$ee_{CC} + \mu\mu_{CC} + ee_{CF}$
Central value	0.23148	0.23123	0.23166	0.23119	0.23140
	Uncertainties				
Total	68	59	43	49	36
Stat.	48	40	29	31	21
Syst.	48	44	32	38	29
	Uncertainties in measurements				
PDF (meas.)	8	9	7	6	4
p_T^Z modelling	0	0	7	0	5
Lepton scale	4	4	4	4	3
Lepton resolution	6	1	2	2	1
Lepton efficiency	11	3	3	2	4
Electron charge misidentification	2	0	1	1	< 1
Muon sagitta bias	0	5	0	1	2
Background	1	2	1	1	2
MC. stat.	25	22	18	16	12
	Uncertainties in predictions				
PDF (predictions)	37	35	22	33	24
QCD scales	6	8	9	5	6
EW corrections	3	3	3	3	3

- **Fit using MMHT14 provides best overall result, i.e. best fit χ^2 and also smallest uncertainties from PDFs after profiling**

$\sin^2\theta^{l}_{eff}$ results based on reference PDF set (MMHT14)

	CT10	CT14	MMHT14	NNPDF31
$\sin^2\theta^{\ell}_{\rm eff}$	0.23118	0.23141	0.23140	0.23146
	Uncertainties in measurements			
Total	39	37	36	38
Stat.	21	21	21	21
Syst.	32	31	29	31

Table 13: Results for extracted values of $\sin^2\theta^{\ell}_{\rm eff}$ with the global breakdown of their uncertainties, shown for the four PDF sets considered in this note. The uncertainty values are given in units of 10^{-5}.

- **Fit using MMHT14 provides best overall result, i.e. best fit χ^2 and also smallest uncertainties from PDFs after profiling**
- **Results quite close for CT14 and NNPDF31, uncertainties a bit larger.**
- **CT10nnlo also shown since it fits best the ensemble of ATLAS W/Z precision data at 7 TeV used for measurement of m_W.**
- **Overall $\sin^2\theta^{l}_{eff}$ range spanned by all PDF sets is 28 10^{-5}**
- **Will discuss this further with PDF4LHC forum (see later slide)**

PDF benchmarking proposal for precision DY

- **Brief presentation of note sent to PDF4LHC forum experts**
- **Some initial feedback and questions (NNPDF, CTEQ, theory)**
- **Some prioritisation and some concrete suggestions to improve PDF theory**
- **Some resource issues and input from experiments**

- **Short note sent to PDF4LHC forum with proposal for benchmarking exercise can be found attached to the agenda of precision EW sub-group meeting on 13/12/2018 at**
  https://indico.cern.ch/event/775325/contributions/3241729/attachments/1769767/2875062/PDFnote_benchmarking_031218.pdf
- **In back-up, one simple slide explaining how correlations between two PDF sets will be computed for any given observable at any given point in (x,Q^2)**

PDF benchmarking proposal: toy datasets

- **Main idea (credit goes to B. Malaescu ☺) is to measure correlations between structure functions provided by global PDF fits "in situ" through the use of a large number of toy datasets.**
- **We will need a large number (~ 10000?) toys to precisely evaluate correlations which are presumably strong (close to unity) but nevertheless known not to be unity.**
- **Without knowing these correlations, we cannot "average" measurements over different PDF sets and therefore cannot scientifically produce a result for mW or s2w which accounts fully for PDF uncertainties in a way which can be defended when legitimate questions are asked about the meaning/value of LHC (or Tevatron) measurements compared to those of LEP.**
- **The advantage of using toys is that the PDF global fits need not change anything in their internal way of obtaining their fit results.**
- **The correlations can be measured using only the central values of the fit results, there is no need to reevaluate PDF Hessian uncertainties or replicas for such an exercise.**
-

PDF benchmarking proposal (main goal): toy datasets for DIS + LHC run-1

- The correlations between PDF sets would be measured in a first step on a dataset (containing about 5000 data measurements) corresponding roughly to that used by NNPDF3.1 or ABM16 (while awaiting CT18 and MMHT18)
- Since a lot of work has gone into producing these PDF sets already, one could hopefully build on these sets with small adjustments not requiring a full-fledged resource-intensive new set of fits:
  - Reduce somewhat LHC run-1 data to what is relevant for DY production (remove jets, keep perhaps only top cross-section measurements to stabilise gluon) and make sure the DIS data and others are the same (go to largest common denominator)
  - Use the same theory predictions for LHC and Tevatron DY measurements (see later slide)

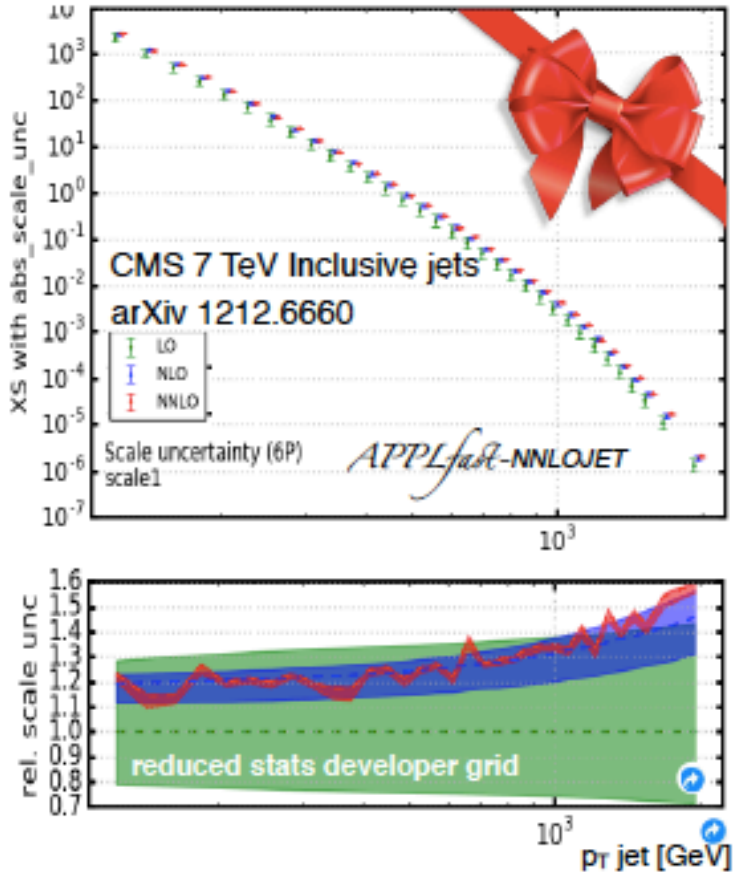
PDF benchmarking proposal: theory

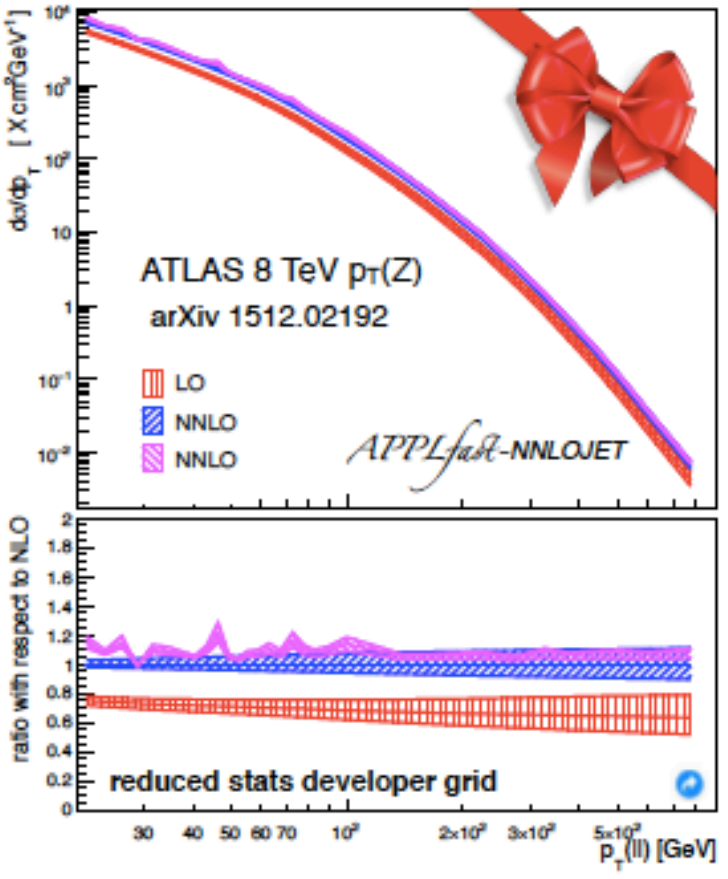
- Through the combination of two fortunate coincidences, we are in position to perform this benchmarking exercise with an improved set of PDF theory tools which are optimal for DY predictions
- NNLOJET with its unique ability to do NNLO pQCD predictions for QCD jet production has been since quite some time a privileged partner in developing a truly NNLO theory for PDF predictions (until now NNLO PDF theory has been in most cases NLO QCD with K factors computed for the relevant observables, usually one-dimensionally)
- It is fortunate that NNLOJET is also unique in its ability to do NNLO pQCD predictions for W/Z pT
- We hope therefore that within a few months, predictions based on NNLOJET could be available for all precision DY measurements available from LHC run-1 data
- This would then be the natural common tool to be used for the benchmarking exercise
- Needs to be prepared ahead of time!

PDF benchmarking proposal: theory

Some example processes

- CMS and ATLAS inclusive jets at 7 TeV
- ATLAS Z(pT) at 8 TeV
  - Available as APPLgrid and fastNLO tables
- **NB: Note the reduced statistics for the NNLO contributions**



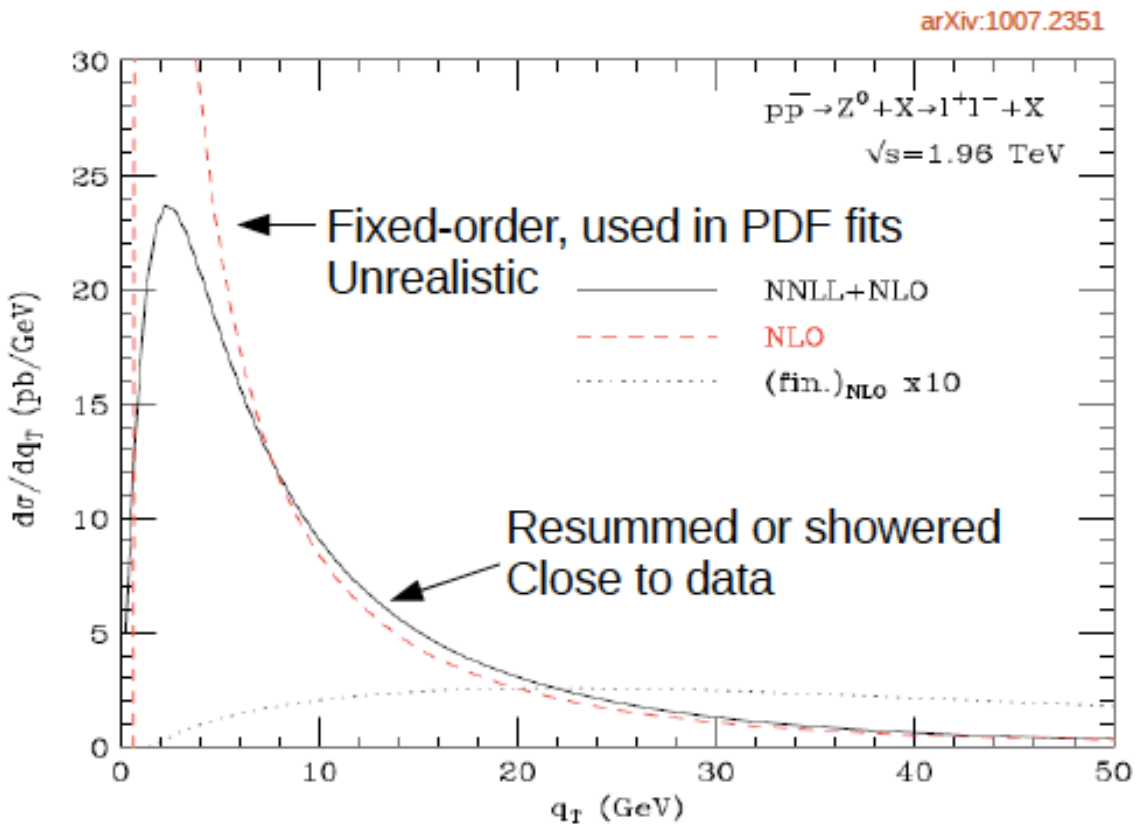


PDF benchmarking proposal: theory

- **Other theory issues are listed in the note and need to be addressed, with perhaps a priority order assumed here to be:**
  1. **Scale variations (already under study in PDF fits, but need to be systematically dealt with somehow in this exercise)**
  2. **Limitations of pQCD versus resummed/parton shower predictions (see next slides for a few examples).**
     Proposal from experiments is to provide correction functions to relevant observables (similar to the NNLO K-factor corrections applied by most global PDF fits to NLO predictions)
     See full talk by M. Boonekamp at precision EW meeting yesterday here
  3. **QED/EW effects which cannot be ignored for precision DY measurements**
  4. **QED PDFs (correlated to point 3 above)**

PDF benchmarking proposal: theory

- **Corrections to pQCD NLO/NNLO predictions based on NLL/NNLL resummation (DYRES/DYTURBO)**



Resummation effects affect the pT distributions, hence the acceptance of fiducial cuts

PDF benchmarking proposal: theory

- **Corrections to pQCD NLO/NNLO predictions based on NLL/NNLL resummation (DYRES/DYTURBO)**

Kinematics

- Main reference : ATLAS 2011 W,Z cross section data, $\sqrt{s} = 7$ TeV

- Z/γ^* : Central – central (CC) : $p_T^l > 20$ GeV; $|\eta_l| < 2.5$

  Central – forward (CF) : $p_T^l > 20$ GeV; $|\eta_C| < 2.5$; $2.5 < |\eta_F| < 4.9$

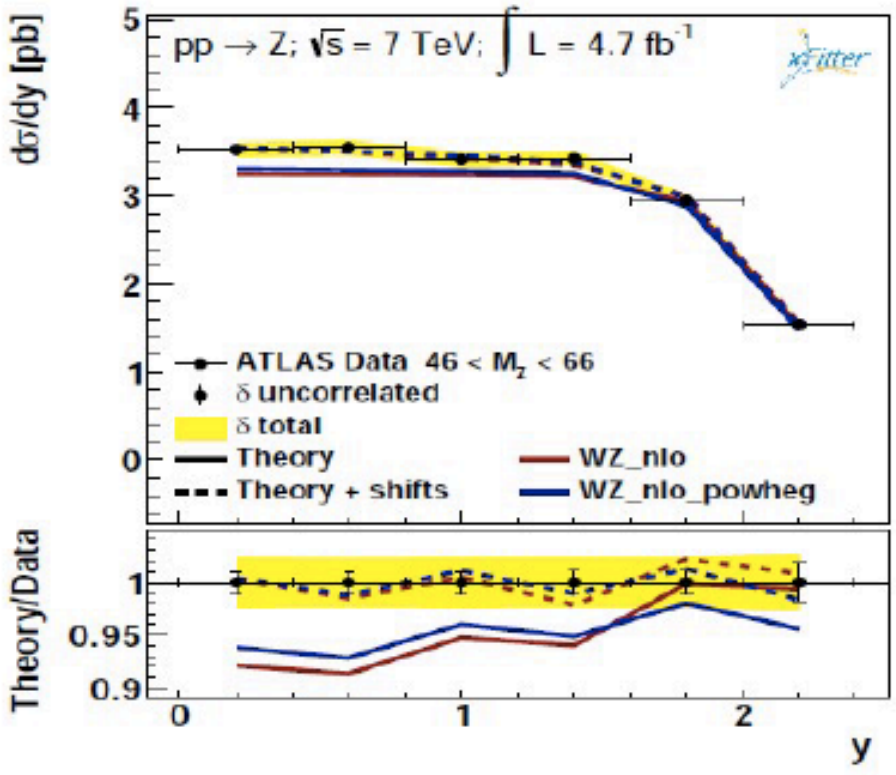
  three mass bins : 46 – 66 ; 66 – 116 ; 116 – 150 GeV

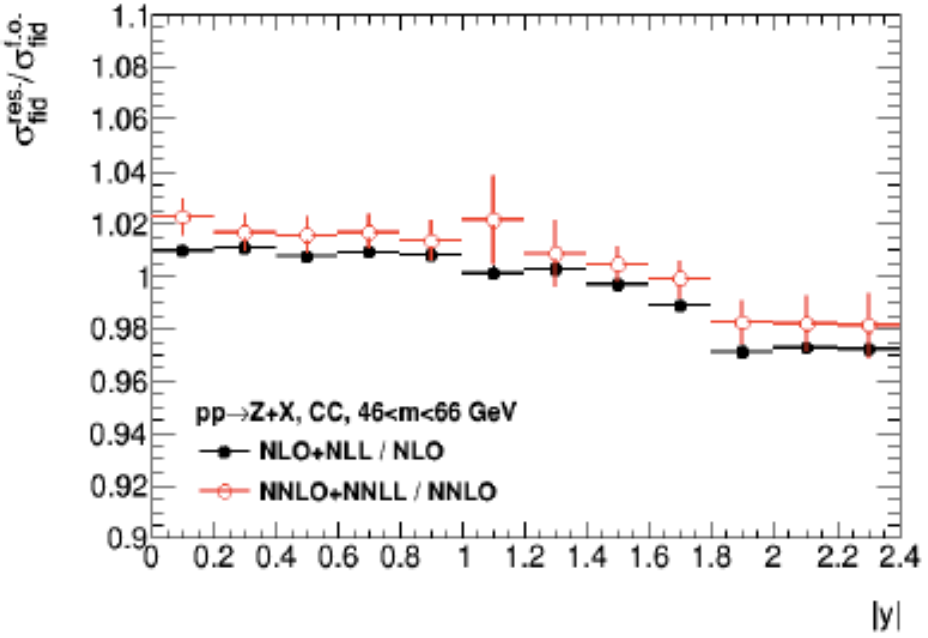
- W +/- : $p_T^l > 25$ GeV; $|\eta_l| < 2.5$; $p_T^\nu > 25$ GeV

  $m_T > 40$ GeV

PDF benchmarking proposal: theory

- **Corrections to pQCD NLO/NNLO predictions based on NLL/NNLL resummation (DYRES/DYTURBO)**

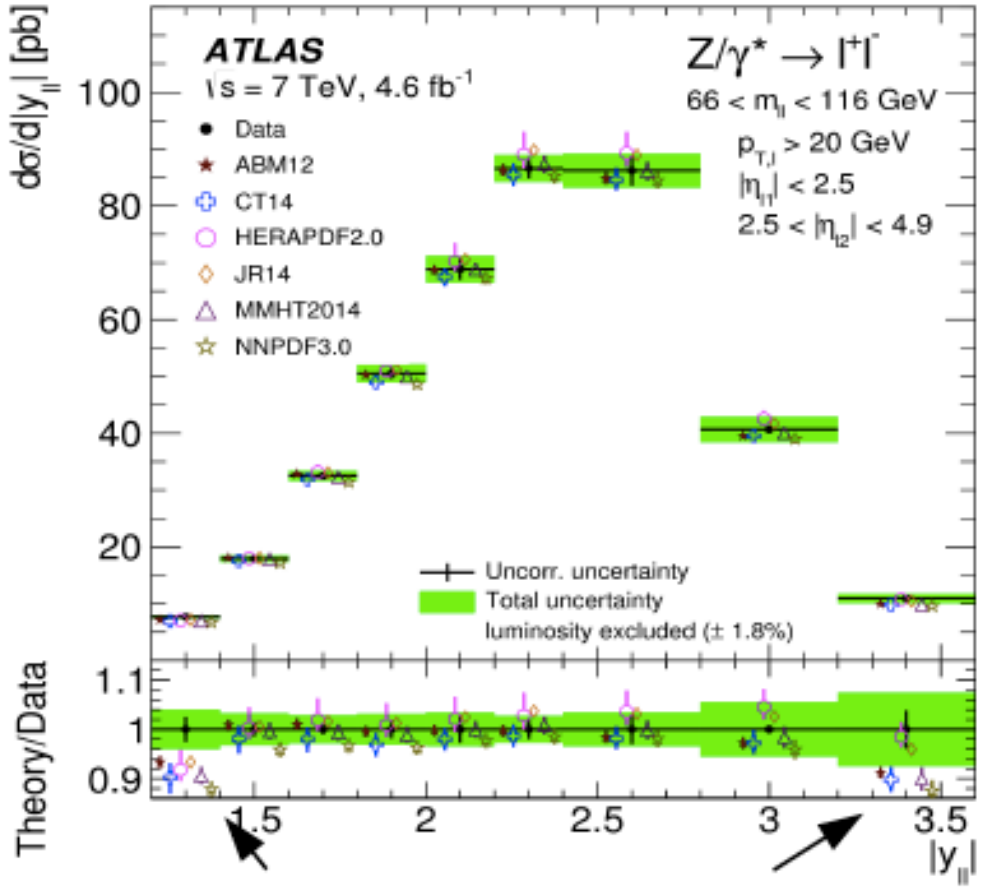


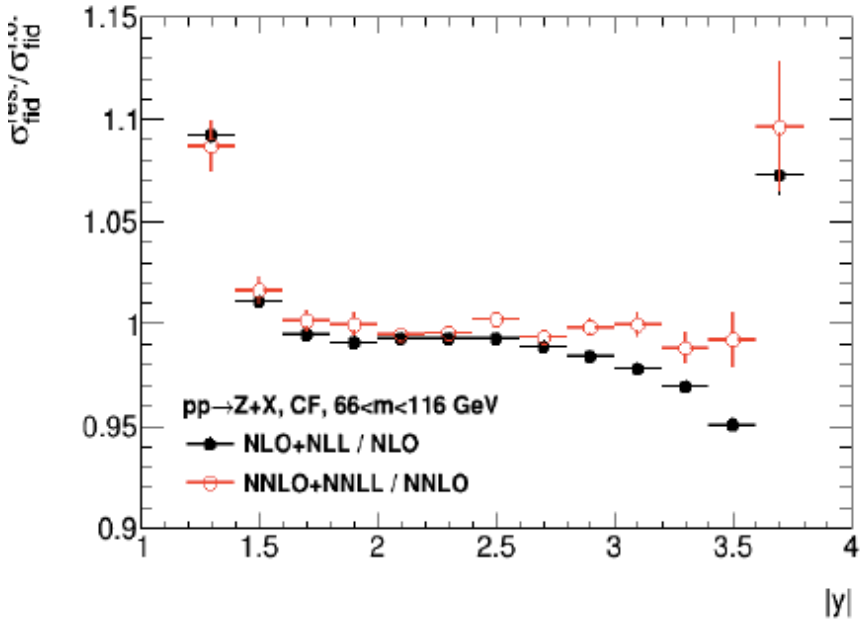


Slope in Theory/data ratio matches the slope of the correction on the right

PDF benchmarking proposal: theory

- **Corrections to pQCD NLO/NNLO predictions based on NLL/NNLL resummation (DYRES/DYTURBO)**

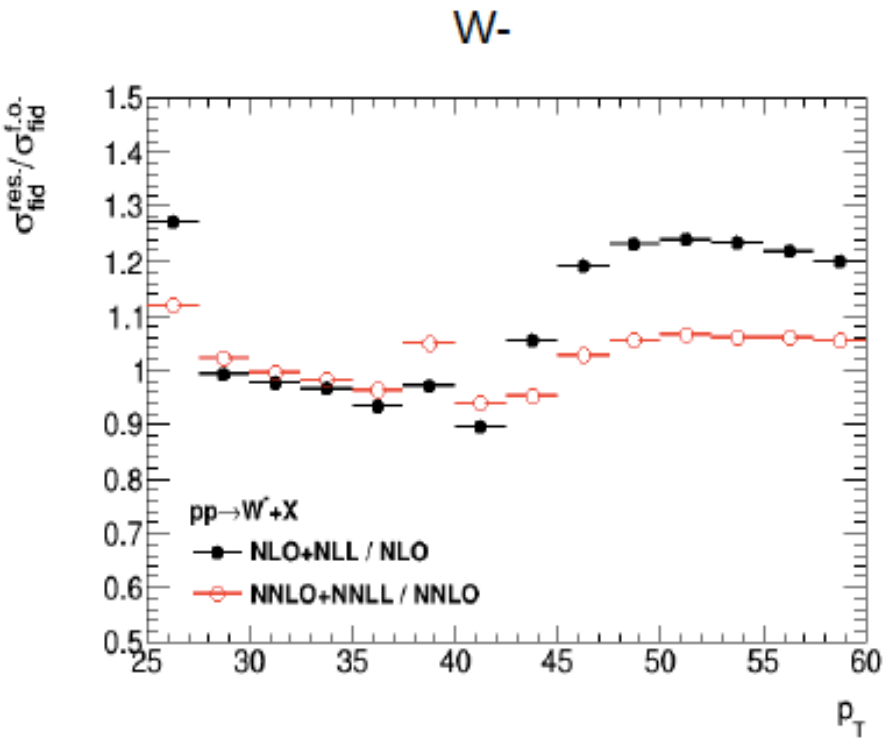


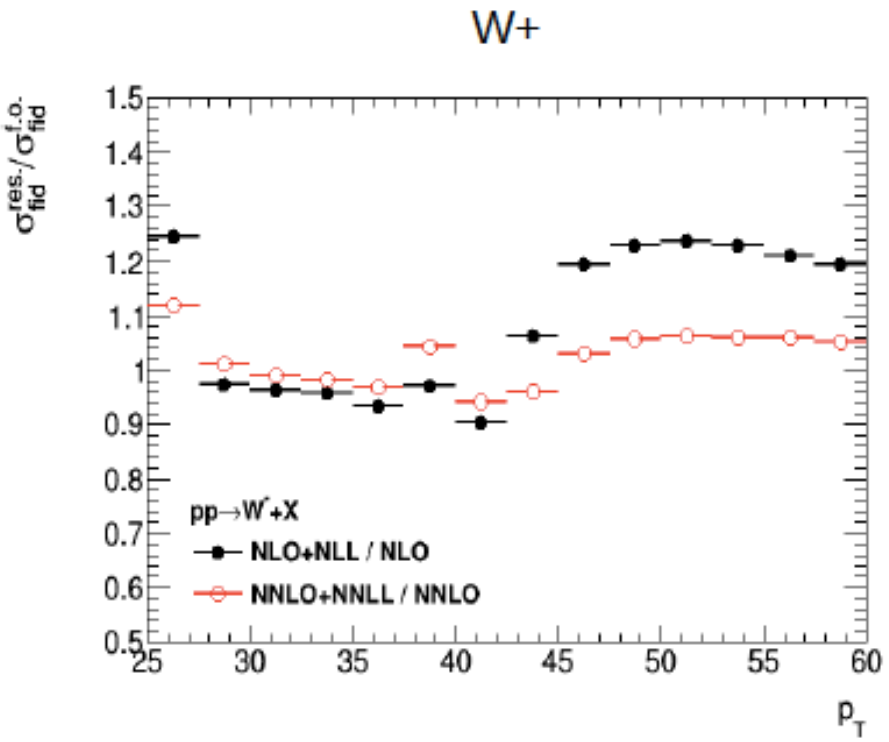


~10% theory deficit near acceptance boundaries (left) matches resummation correction

PDF benchmarking proposal: theory

- **Corrections to pQCD NLO/NNLO predictions based on NLL/NNLL resummation (DYRES/DYTURBO)**





The lepton pT distributions are also sensitive to physics (m_W at the Jacobian peak; PDFs above).
Resummation corrections reach up to ~25% (NLL) and ~10% (NNLL)

PDF benchmarking proposal: resources

- Global PDF fits require CPU, memory and human resources
- CPU resource examples:
  - one NNPDF fit requires one week on 100 cores typically.
  - so, 10000 toy fits require eg the full ATLASor CMS grid computing capacity for three weeks.
  - NNPDF is atypical since it produces its fit results through replicas, so it is not clear whether the CPU time can be decreased for this exercise where toy fits do not require the production of replicas (or of Hessian uncertainties)
  - in the case of CT/MMHT/ABM (?), the fit is 3-4 times faster if the uncertainties are not required
- Memory resources might be larger than what is commonly available on Grid computing single cores (to be understood better)
- Human resources are an issue at least for some PDF groups for the production of a fit like CT14 (public release for use in physics publications), so the benchmarking exercise has surely to be considered as such and not more.
- Human intervention for the 10000 toy fits should in principle not be needed nor even recommended.

PDF benchmarking proposal: resources

- **Global PDF fits require CPU, memory and human resources**
- **The proposed benchmarking exercise needs its own specific human resources:**
  - **If XFITTER could be used as a framework for generating toys, book-keeping, for fitting the toys (after validation of fit results based on comparison with fits performed by global PDF experts within their own framework), some resources could be made available to community (all experiments and all global PDF experts) for any operational issues and/or framework improvements required**
  - **However, some resources (not identified yet) need to be provided by experiments to work run the machinery, provide validation plots in the initial phase, provide toy fit results in format which can be examined by existing tools, and analyse the results, extracting the correlations and defining further iterations**
  - **Some resources are also required, especially in the initial phase, from the global PDF experts (not identified yet)**

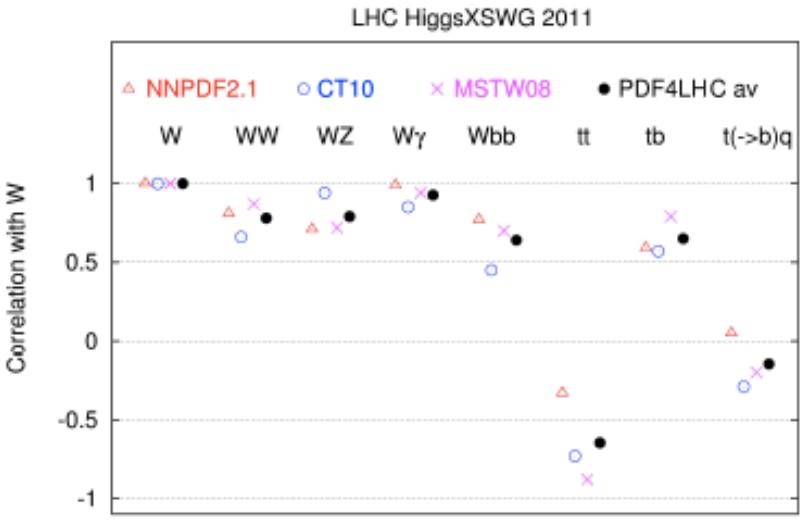
PDF benchmarking proposal: timeline

- **It is desirable to do such a benchmarking exercise if it can be done and documented with its conclusions over ~ one year, such that it is used by the future run-2 publications of the LHC experiments.**
- **Tentative timeline from note is based on this:**
  - **Setting up, validation of toy production and toy fits, dry runs (100 toys?)**
    Jan-April 2019
  - **First production run (as close as feasible to 10000 toy fits**
    April-July 2019
  - **Analysis of results, second iteration**
    July-October 2019
  - **Documentation of results, conclusions**
    September-December 2019
- **This should be in sync with the rest of the precision EW group report, but if it takes a bit longer, we could consider also a different report as a conclusion of this sizable amount of work.**

PDF benchmarking proposal: feedback

CORRELATING PDFS

CORRELATION BETWEEN HIGGS SIGNAL AND BACKGROUND (HXSWG, YR2)



S. Forte

- CORRELATION BETWEEN PROCESSES AND PDFS, PROCESSES AND PROCESSES, PDF AND PDFS TRIVIAL TO COMPUTE ⇒ NO NEED TO RUN DEDICATED FITS
- PREVIOUS EXERCISES SUGGEST VERY LARGE CORRELATION (SHOULD BE 100% FOR SAME DATA)
- IN PDF4LHC15 CORRELATION ASSUMED TO BE 100%: SIMPLE AVERAGE WEIGHTED AVERAGE DUBIOUS AND DANGEROUS
  - PDFS W/ SMALLER UNCERTANITY GET LARGER WEIGHT UNCERTAINTY DOMINATED BY METHODOLOGY ⇒ SMALLER UNCERTAINTY COULD JUST BE BIAS!
  - UNCERTAINTY REDUCED IF CORRELATION LESS THAN 100% CAN WE BELIEVE IT IN THE ABSENCE OF NEW INFORMATION?

PDF benchmarking proposal: feedback

MMHT - Dynamical Tolerance

L. Harland-Lang

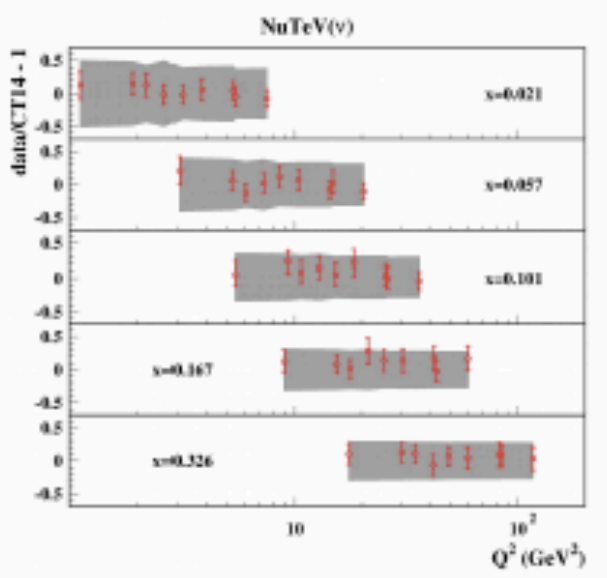
- MMHT approach - 'dynamical' tolerance calculation, i.e. tolerance is determined for each eigenvector direction $\Rightarrow$ 50 tolerances for MMHT14 (25 eigenvectors).
- Values calculated using conservative 'hypothesis testing' criteria. Require every dataset n with N points described within $\Delta\chi^2_n < \sqrt{2N}$.

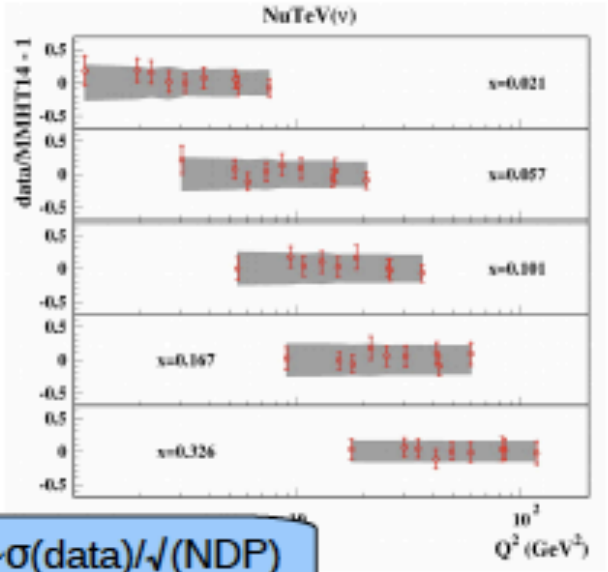
eigen–vector	+ t	T	most constraining data set	− t	T	most constraining data set
1	4.00	3.97	HERA e^+p NC 920 GeV	4.30	4.66	HERA e^+p NC 820 GeV
2	2.50	2.84	HERA e^+p NC 920 GeV	1.80	1.53	NMC μd F_2
3	3.80	4.00	NMC.....HERA F_L	3.70	3.69	NMC μd F_2
4	4.05	4.00	DØ II $W \to \nu e$ asym.	5.00	5.11	DØ II $W \to \nu\mu$ asym.
5	3.40	3.35	DØ II $W \to \nu\mu$ asym.	4.20	4.45	NuTeV $\nu N \to \mu\mu X$
6	1.85	1.88	NuTeV $\nu N \to \mu\mu X$	3.70	3.71	DØ II $W \to \nu\mu$ asym.
7	1.55	1.67	E866/NuSea pd/pp DY	2.15	2.03	E866/NuSea pd/pp DY
8	2.75	2.64	DØ II $W \to \nu\mu$ asym.	1.90	2.01	E866/NuSea pd/pp DY
9	3.40	3.46	E866/NuSea pd/pp DY	3.80	3.78	BCDMS μp F_2
10	3.15	3.47	NuTeV $\nu N \to \mu\mu X$	2.40	2.13	NuTeV νN F_2

- Find range of $T \sim 1.5 - 5$, with average $T \sim 3$ for e.g. MMHT14NNLO.
- Contributions to tolerance? Complex question, but note errors with tolerance broadly compatible with separate NNPDF approach.

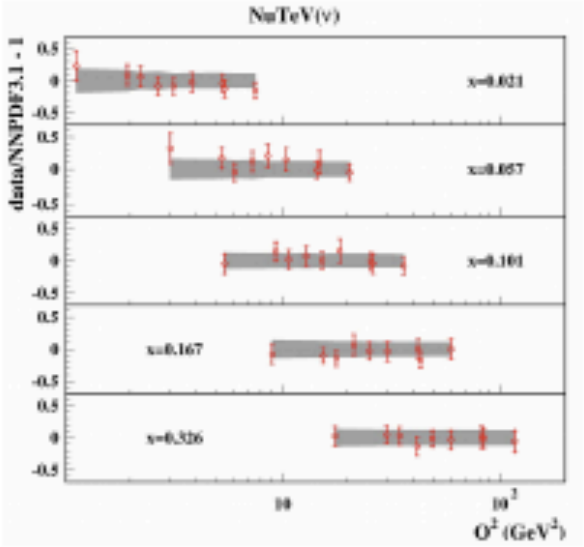
PDF benchmarking proposal: feedback

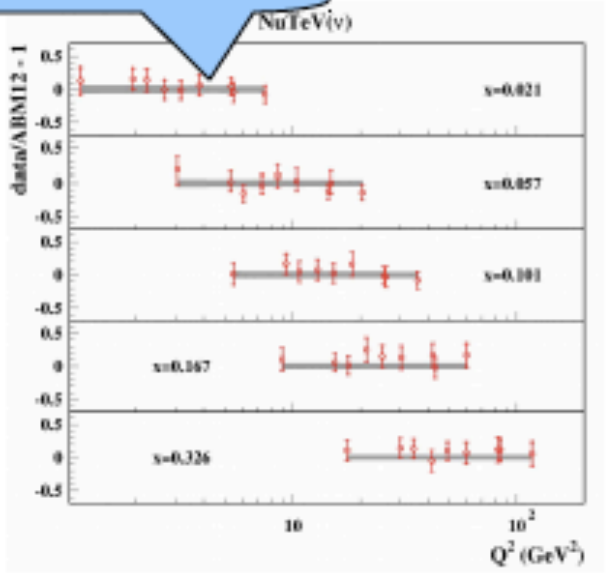
Closure test with the NuTeV data





σ(fit)~σ(data)/√(NDP)





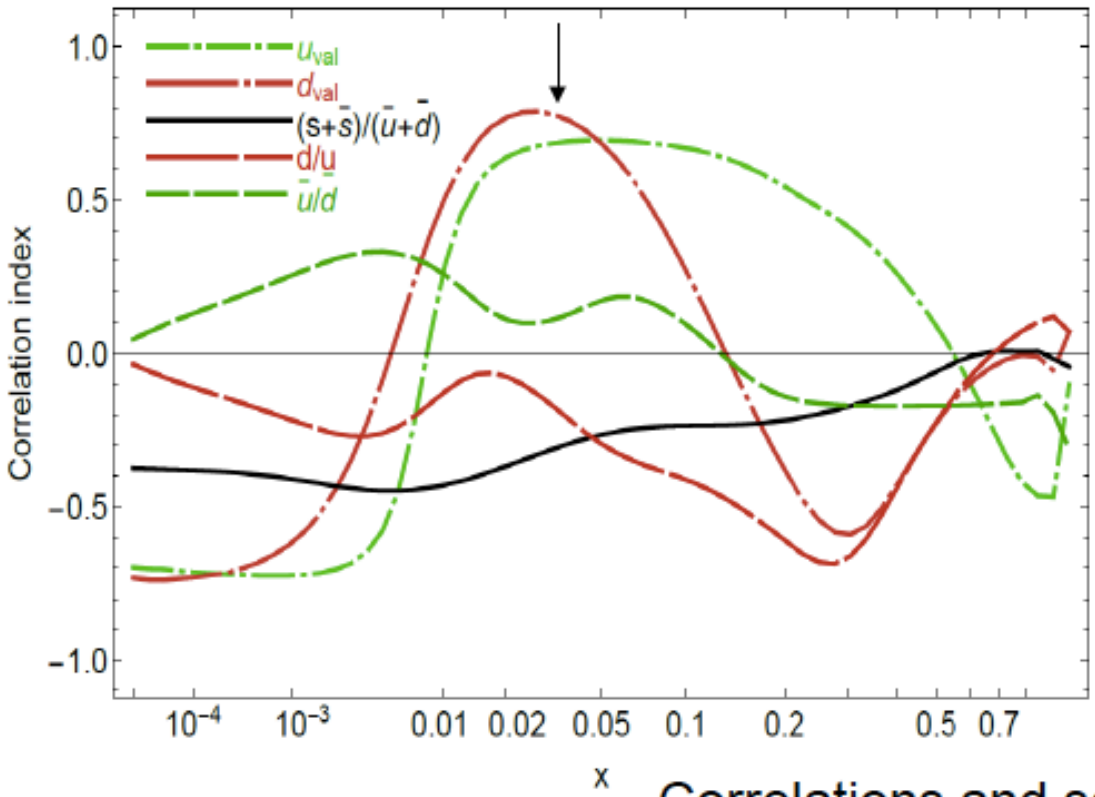
S. Alekhin

PDF benchmarking proposal: feedback

Hessian correlation for $\sin^2\theta_w$ at 8 TeV

Presented at the EW precision subgroup meeting, Nov. 13, 2018

P. Nadolsky



Strongest correlations of s2w with u_{val}, d_{val} at $x = 0.01 - 0.2$

weak correlations with $\bar{u}$, $\bar{\mathrm{d}}$, $\bar{s}$, g

Correlations and sensitivities (next slide) elucidate PDF properties using **published** error PDF sets

PDF benchmarking proposal: next steps

- **Next steps require new document incorporating feedback received yesterday, answers to objections (major ones listed below), and a concrete proposal on how to move forward.**

  This will be done in January

- **Major objections/points of discussion:**
  - **When measuring correlations with toys, won't we only measure noise?**
  - **How to account for tolerances applied by CT and MMHT through toys?**
  - **How do we deal with the fact that different groups use in practice different data?**
  - **Not obvious that NNLOJET grids will be available within next few months**
  - **A large number of toys requires large resources, perhaps one could do this more efficiently by building an umbrella set of eigenvalue-based covariance matrices which would somehow incorporate the correlations between different PDF fits?**

    As far as (not very far) as my personal understanding goes, this should be possible if one would start from the data inputs (about 5000 measurements) and their uncertainties and correlations.

- **Most likely, we will launch on a small scale (100 toys or so, existing theory predictions, reduced set of data points?) a benchmarking test with both approaches to evaluate more realistically what can be done over a year or so**

Back-up slides

Computing correlations between two PDF sets

$$f_{i,A} = f^{PDF\,A}(x_i, Q_i^2)$$

$$Cov(f_{i,A}; f_{j,B}) = \frac{\sum_{t=1}^{N_{toys}} (f_{i,A}^t - \overline{f_{i,A}}) \cdot (f_{j,B}^t - \overline{f_{j,B}})}{N_{toys}}$$

$$Corr(f_{i,A}; f_{j,B}) = \frac{Cov(f_{i,A}; f_{j,B})}{\sqrt{Cov(f_{i,A}; f_{i,A}) \cdot Cov(f_{j,B}; f_{j,B})}}$$

$$O_A^a = H^a(f_A)$$

$$Cov(O_A^a; O_B^b) = \frac{\sum_{t=1}^{N_{toys}} (O_A^{a,t} - \overline{O_A^a}) \cdot (O_B^{b,t} - \overline{O_B^b})}{N_{toys}}$$