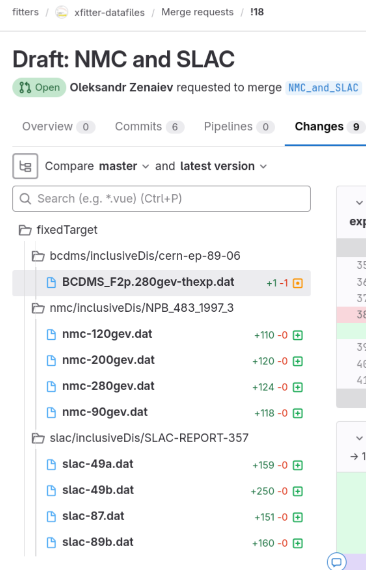
Fixed target DIS data, target mass corrections and higher twist in xfitter

- Fixed target ep (μp)DIS data sets which provide PDF constraints at high x:
  - NMC
  - SLAC
  - (BCDMS) [available in xFitter, however not used in "default" fits]
- Theoretical predictions for these data sets require power corrections:
  - target mass corrections (TMC)
  - higher twist (HT)
- Implemented according to the ABMP16 prescription [Phys.Rev. D96 (2017) no.1, 014011]

Experiment	Beam (E_b) or center-of-mass energy ($\sqrt{s}$)	$\mathcal{L}$ (1/fb)	Process	Kinematic cuts used in the present analysis (cf. orginal references for notations)	Ref.
DIS					
HERA I+II	$\sqrt{s} = 0.225 \div 0.32$ TeV	0.5	$e^{\pm} p \to e^{\pm} X$ $e^{\pm} p \to \overset{(-)}{\nu} X$	$2.5 \leq Q^2 \leq 50000$ GeV^2, $2.5 \cdot 10^{-5} \leq x \leq 0.65$ $200 \leq Q^2 \leq 50000$ GeV^2, $1.3 \cdot 10^{-2} \leq x \leq 0.40$	[4]
BCDMS	E_b =100÷280 GeV		$\mu^+ p \to \mu^+ X$	$7 < Q^2 < 230$ GeV^2, $0.07 \leq x \leq 0.75$	[61]
NMC	E_b =90÷280 GeV		$\mu^+ p \to \mu^+ X$	$2.5 \leq Q^2 < 65$ GeV^2, $0.009 \leq x < 0.5$	[60]
SLAC-49a	E_b =7÷20 GeV		$e^- p \to e^- X$	$2.5 \leq Q^2 < 8$ GeV^2, $0.1 < x < 0.8$, $W \geq 1.8$ GeV	[54] [62]
SLAC-49b	E_b =4.5÷18 GeV		$e^- p \to e^- X$	$2.5 \leq Q^2 < 20$ GeV^2, $0.1 < x < 0.9$, $W \geq 1.8$ GeV	[54] [62]
SLAC-87	E_b =8.7÷20 GeV		$e^- p \to e^- X$	$2.5 \leq Q^2 < 20$ GeV^2, $0.3 < x < 0.9$, $W \geq 1.8$ GeV	[54] [62]
SLAC-89b	E_b =6.5÷19.5 GeV		$e^- p \to e^- X$	$2.5 \leq Q^2 \leq 19$ GeV^2, $0.17 < x < 0.9$, $W \geq 1.8$ GeV	[56] [62]
DIS heavy-quark production					
HERA I+II	$\sqrt{s}$ =0.32 TeV		$e^{\pm} p \to e^{\pm} c X$	$2.5 \leq Q^2 \leq 2000$ GeV^2, $2.5 \cdot 10^{-5} \leq x \leq 0.05$	[63]
H1	$\sqrt{s}$ =0.32 TeV	0.189	$e^{\pm} p \to e^{\pm} b X$	$5 \leq Q^2 \leq 2000$ GeV^2, $2 \cdot 10^{-4} \leq x \leq 0.05$	[15]
ZEUS	$\sqrt{s}$ =0.32 TeV	0.354	$e^{\pm} p \to e^{\pm} b X$	$6.5 \leq Q^2 \leq 600$ GeV^2, $1.5 \cdot 10^{-4} \leq x \leq 0.035$	[16]

New data sets

- MR: `https://gitlab.cern.ch/fitters/xfitter-datafiles/-/merge_requests/18`



- NMC and SLAC data sets come with normalisation factors which can be fitted (via `parameters.yaml`)

Experiment	Process	Beam energy (GeV)	Reference	Normalization
SLAC-49a	$ep \to eX$	$7 \div 20$	[54, 62]	1.001(11)
SLAC-49b	$ep \to eX$	$4.5 \div 18$	[54, 62]	1.010(15)
SLAC-87	$ep \to eX$	$8.7 \div 20$	[54, 62]	1.012(11)
SLAC-89b	$ep \to eX$	$6.5 \div 19.5$	[56, 62]	1.000(11)
BCDMS	$\mu p \to \mu X$	$100 \div 280$	[61]	0.976(7)
NMC	$\mu p \to \mu X$	90	[60]	0.993(13)
		120		1.011(12)
		200		1.022(12)
		280		1.012(12)

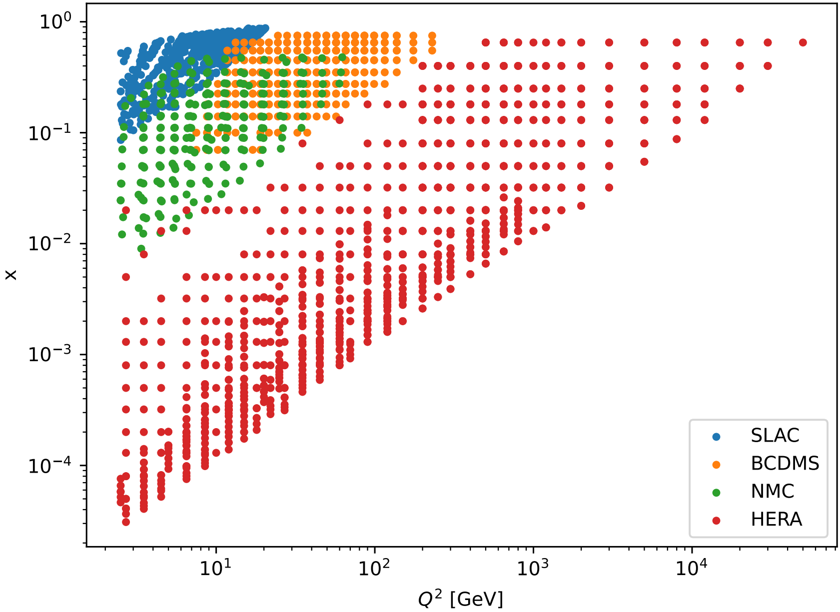
- Data file snapshot:

```
&Data
  Name = 'SLAC 49a'
  IndexDataset = 93
  Reaction = 'NC e+-p'

  TermName = 'R','N'
  TermSource = 'use:hf_scheme_DISNC','KFactor'
  TermInfo = 'type=sigred:echarge=-1:polarity=0','Parameter=DATANORM_SLAC49a'
  TheorExpr = 'R/N'
```

- Snapshot of `parameters.yaml`:

```
Parameters:
  # data normalisation
  DATANORM_NMC90 :   [0.993, 0.01]
  DATANORM_NMC120 :  [1.011, 0.01]
  DATANORM_NMC200 :  [1.022, 0.01]
  DATANORM_NMC280 :  [1.012, 0.01]
  DATANORM_SLAC49a :  [1.001, 0.01]
  DATANORM_SLAC49b :  [1.010, 0.01]
  DATANORM_SLAC87 :  [1.012, 0.01]
```

New data sets: Q^2 vs x



New feature request (or maybe it is already possible?)

- Need to implement such normalisation factor also for BCDMS data
- However, do not want to touch existing data files since it could be analysis specific and will break existing tests
- A more general way could be to override certain fields for some data sets from `parameters.yaml`, e.g.:

```
Override:
  Data:
    'BCDMS F2p 100GeV':
      TermName = 'R','N'
      TermSource = 'use:hf_scheme_DISNC','KFactor'
      TermInfo = 'type=sigred:echarge=-1:polarity=0','Parameter=DATANORM_BCDMS'
      TheorExpr = 'R*N'
```

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

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

  and

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

  respectively, which holds up to $O(M_N^2/Q^2)$. 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 technical details

- Need to compute the integral for each data point (but at LO): time consuming
  - Monte Carlo integration using GSL library: $\sim$ 100 calls of the integration function $\rightarrow$ more than factor 10 slower
  - checked Simpson's 3/8 rule (4 calls): not sufficient accuracy for some data points, however works very well for the vast majority of the data
  - might need further optimisation (currently the code is not cleaned up)

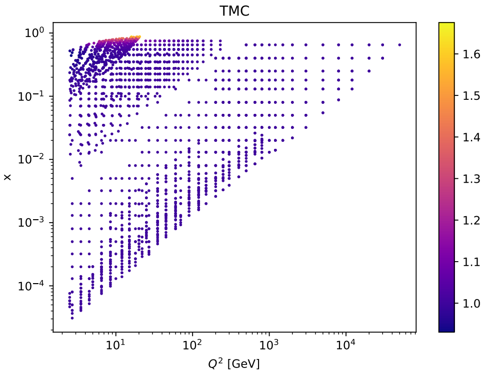
CMakeLists.txt | ReactionHathor.cc | ReactionFFABM_DISNC.cc M × | TheorEval.cc | ReactionBaseDISNC.h | TermData.h | Settings

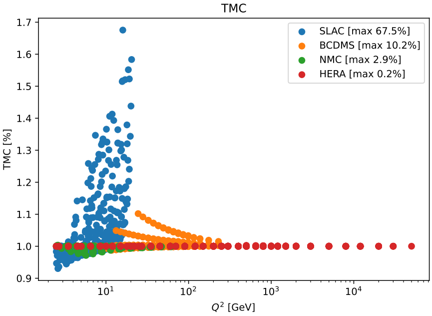
DISNC > ReactionFFABM_DISNC.cc > apply_tmc(double &, double &, double &, const int, const std::valarray<double>&, const std::valarray<double>&, const int, const int, c

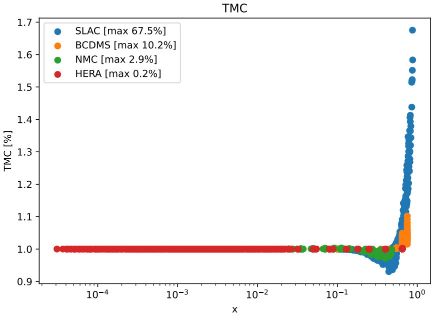
```
pars.flag_flavour = flag_flavour;
pars.order = 0;
// numerical integration
gsl_function F;
F.function = integrate;
F.params = &pars;
size_t alloc_space = 1000;
gsl_integration_workspace * w = gsl_integration_workspace_alloc(alloc_space);
double epsabs = 0;
double epsrel = 1e-3;
int key_param = 6;
double result, error;
gsl_integration_qag (&F, xi, 1.0, epsabs, epsrel, alloc_space, key_param, w, &result, &error);
//gsl_integration_qag (&F, x[i], 1.0, epsabs, epsrel, alloc_space, key_param, w, &result, &error);
gsl_integration_workspace_free (w);
// Simpson 3/8 integration
double a = xi;
//double b = log10(1/xi);
double b = a*5;
if (b > 0.999)
  b = 0.999;
double sim38 = (b-a)/8.*(integrate(a, &pars)+3*integrate((2*a+b)/3., &pars)+3*integrate((a+2*b)/3., &pars)+integrate(b, &pars));
//printf("%f %f %f %f\n", integrate(a, &pars), integrate((2*a+b)/3., &pars), integrate((a+2*b)/3., &pars), integrate(b, &pars));
double I = result;
//I = sim38;
```

- Currently it is implemented in `FFABM` reaction only
  - not so easy to implement in `BaseDISNC` due to variable substitution
  - might need further studies: either it needs to be implemented in each scheme (`RT`, `FONLL`, `ZMVFNS`) separately, or `BaseDISNC` needs to be substantially modified
- Controlled by new parameter `tmc` (0 or 1)

TMC


TMC
x
Q^2 [GeV]
1.0
1.1
1.2
1.3
1.4
1.5
1.6


TMC
SLAC [max 67.5%]
BCDMS [max 10.2%]
NMC [max 2.9%]
HERA [max 0.2%]
TMC [%]
Q^2 [GeV]


TMC
SLAC [max 67.5%]
BCDMS [max 10.2%]
NMC [max 2.9%]
HERA [max 0.2%]
TMC [%]
x

HT

- Description from Phys.Rev. D96 (2017) no.1, 014011 (ABMP16):

$$F_i^{\rm ht}(x,Q^2) = F_i^{\rm TMC}(x,Q^2) + \frac{H_i^{\tau=4}(x)}{Q^2}, \qquad i = 2,T, \tag{6}$$

where $F_i^{\rm TMC}$ is given by the leading twist structure function of Eq. (2) together with the target mass corrections [112], see also [71]. The reference scale in Eq. (2) has been chosen $Q_0^2 = 1\,{\rm GeV}^2$ and the higher twist terms H_i^τ are taken to be independent of Q^2, i.e. to correspond to the central value of Q^2 in the respective x-range being analyzed. The results for H_i^τ will be presented below

We parameterize the coefficients $H_i^{\tau=4}$ for $i = 2,T$ of the higher twist (HT) terms in the DIS nucleon structure functions F_i in Eq. (6) as follows

$$H_i^{\tau=4}(x) = x^{\alpha_i} S_i(x), \qquad i = 2,T, \tag{30}$$

where $S_i(x)$ are the cubic splines defined at the knots $\{[x_k, H_i^{\tau=4}(x_k)] : k = 1,7\}$ and $\{x_k\} = (0, 0.1, 0.3, 0.5, 0.7, 0.9, 1)$. The constraint $H_i^{\tau=4}(1) = 0$ is imposed due to poor coverage of the region of $x \to 1$ by the existing data. The remaining knot values $H_i^{\tau=4}(x_k)$ for $k = 1, \ldots, 6$ are taken as fit parameters and the result of the fit with Eq. (30) is presented in Tab. VIII and Fig. 26.

	$H_2^{\tau=4}(x)/{\rm GeV}^2$	$H_T^{\tau=4}(x)/{\rm GeV}^2$
$x = 0.0$	0.023 ± 0.019	-0.319 ± 0.126
$x = 0.1$	-0.032 ± 0.013	-0.134 ± 0.040
$x = 0.3$	-0.005 ± 0.009	-0.052 ± 0.030
$x = 0.5$	0.025 ± 0.006	0.071 ± 0.025
$x = 0.7$	0.051 ± 0.005	0.030 ± 0.012
$x = 0.9$	0.003 ± 0.004	0.003 ± 0.007
$x = 1$	0	0

HT technical details

- Implemented in `BaseDISNC` such that it works in any scheme:

CMakeLists.txt | ReactionHathor.cc | ReactionFFABM_DISNC.cc M | ReactionBaseDISNC.cc 2

reactions > BaseDISNC > ReactionBaseDISNC.cc > ...

```
void ReactionBaseDISNC::sred BASE_PARS
{
  unsigned termID = td->id;

  auto *yp = GetBinValues(td, "y");
  auto y = *yp;

  valarray<double> f2(_npoints[termID]);
  F2(td, f2, err);
  if (_flag_ht[termID] && 0)
    ApplyHigherTwist(td, 2, f2, err);

  valarray<double> fl(_npoints[termID]);
  FL(td, fl, err);
  if (_flag_ht[termID] && 0)
    ApplyHigherTwist(td, 1, fl, err);
```

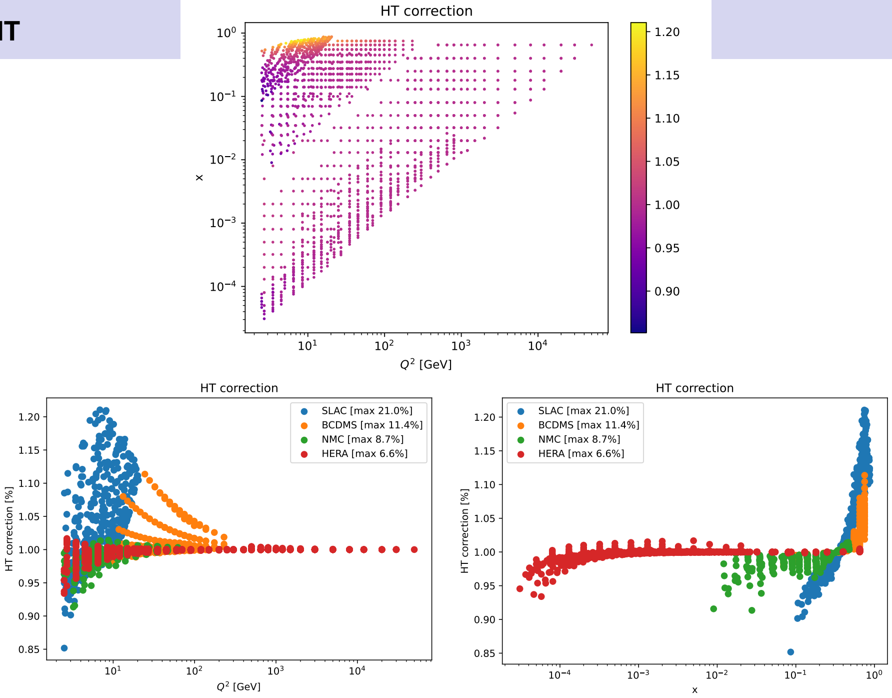
- Controlled by new parameter `ht` (0 or 1)
- Currently the spline knots, their values and α powers are hardcoded (however it is easy to make them fit parameters)

CMakeLists.txt | ReactionHathor.cc | ReactionFFABM_DISNC.cc M | ReactionBaseDISNC.cc 2

reactions > BaseDISNC > ReactionBaseDISNC.cc > initTerm(TermData *)

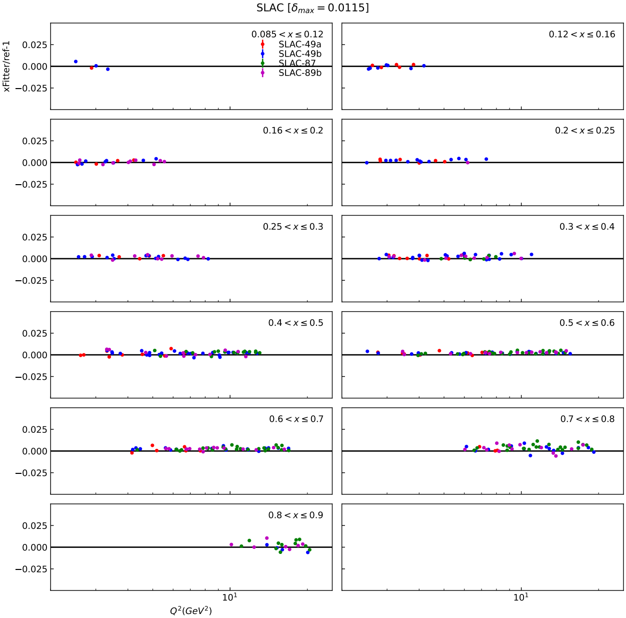
```

  // higher twist spline knots
  _ht_x = {    0.,    0.1,    0.3,    0.5,    0.7,    0.9, 1.};
  _ht_2 = { 0.023, -0.032, -0.005, 0.025, 0.051, 0.003, 0.};
  _ht_t = {-0.319, -0.134, -0.052, 0.071, 0.030, 0.003, 0.};
  _ht_alpha_2 = 0.;
  _ht_alpha_t = 0.05;
}
```

HT


HT correction
x
Q^2 [GeV]
1.20
1.15
1.10
1.05
1.00
0.95
0.90
HT correction
SLAC [max 21.0%]
BCDMS [max 11.4%]
NMC [max 8.7%]
HERA [max 6.6%]
HT correction [%]
Q^2 [GeV]
HT correction
SLAC [max 21.0%]
BCDMS [max 11.4%]
NMC [max 8.7%]
HERA [max 6.6%]
HT correction [%]
x

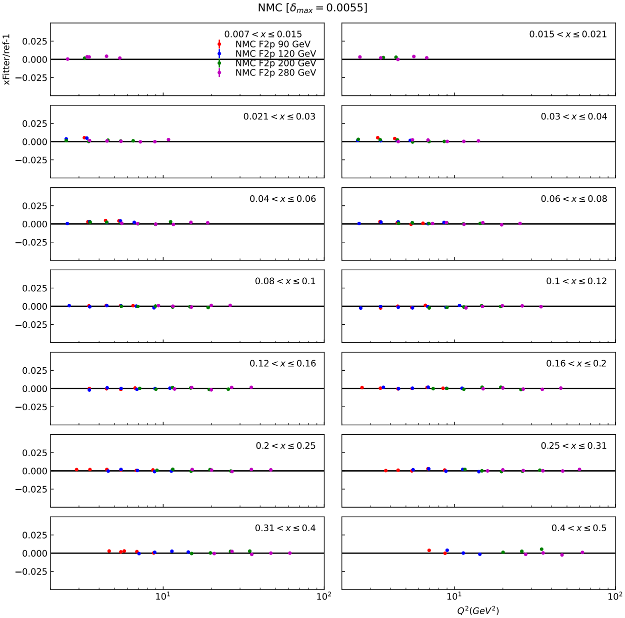
Validation of theoretical predictions vs ABMP16

xfitter vs ABMP16: SLAC



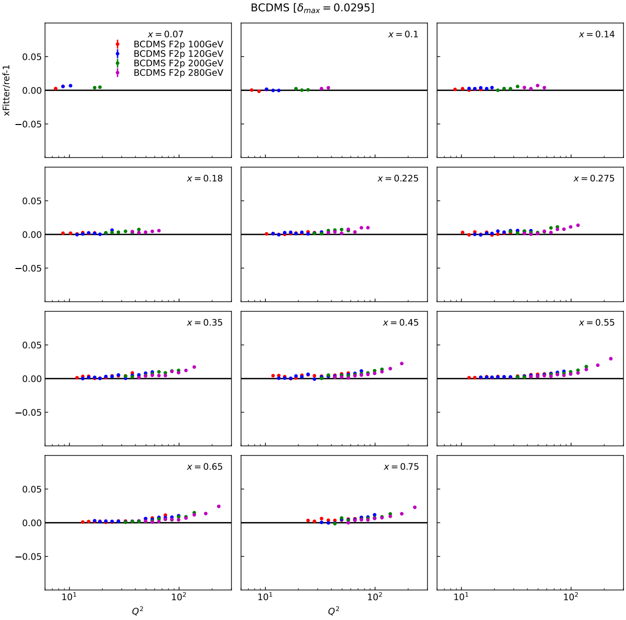
- Differences $\lesssim$ 1%

xfitter vs ABMP16: NMC



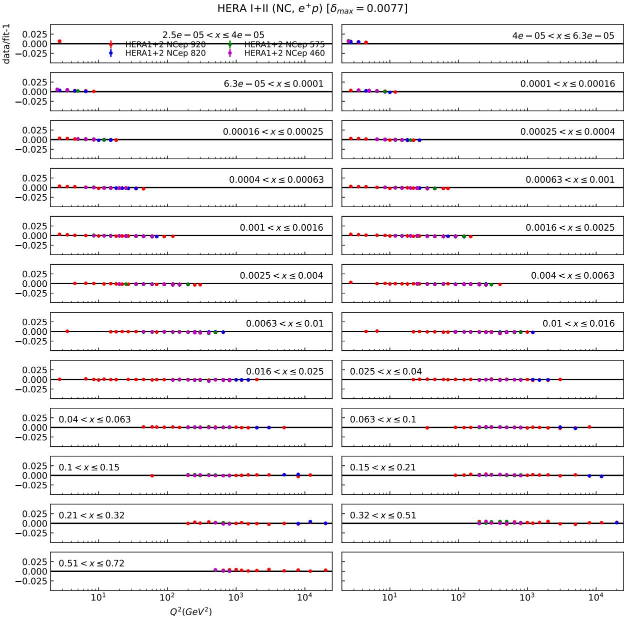
- Differences $< 1\%$

xfitter vs ABMP16: BCDMS

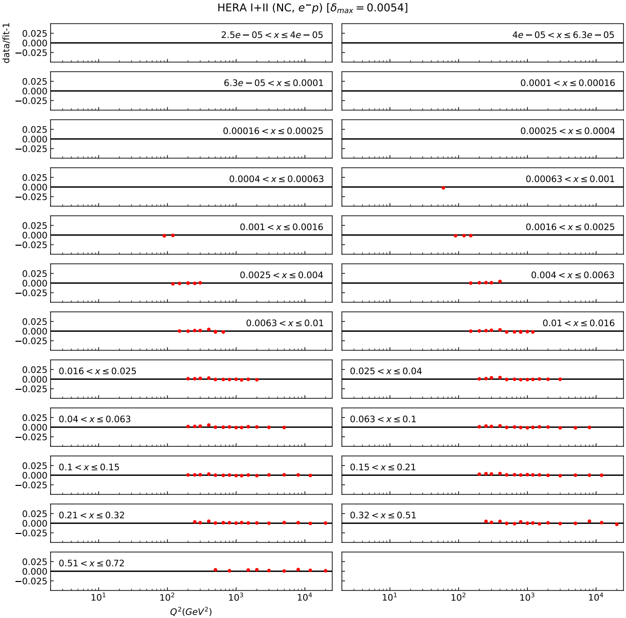


- Known difference up to 3% at large Q^2: ABMP16 does not have F3 for this data set (small compared to data uncertainties)
- Without F3 term in xfitter, the differences reduce to $< 1\%$

xfitter vs ABMP16: HERA NC e^+p

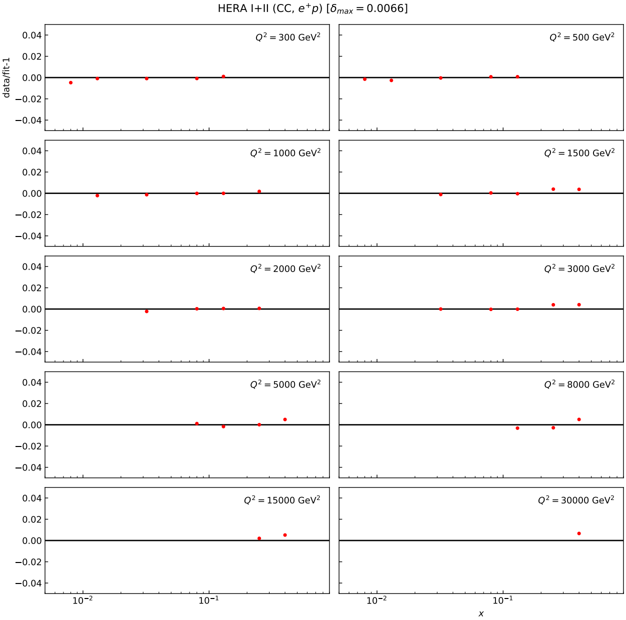


xfitter vs ABMP16: HERA NC $e^- p$



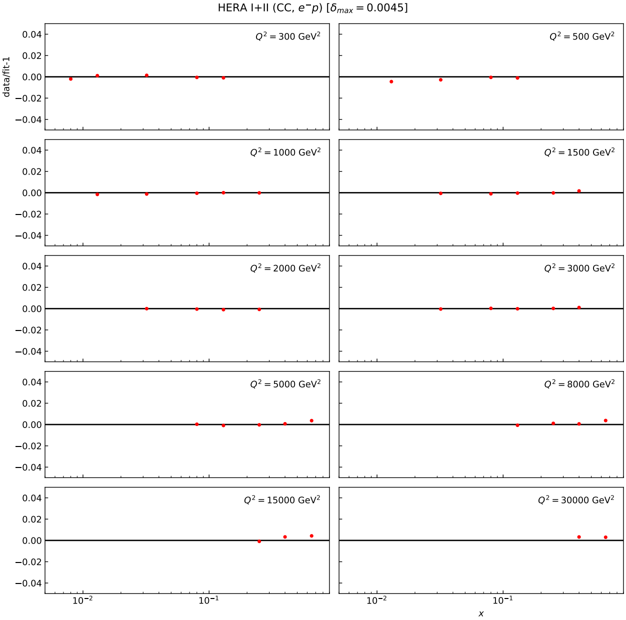
- Differences $< 1\%$

xfitter vs ABMP16: HERA CC e^+p



- Differences $< 1\%$

xfitter vs ABMP16: HERA CC $e^- p$

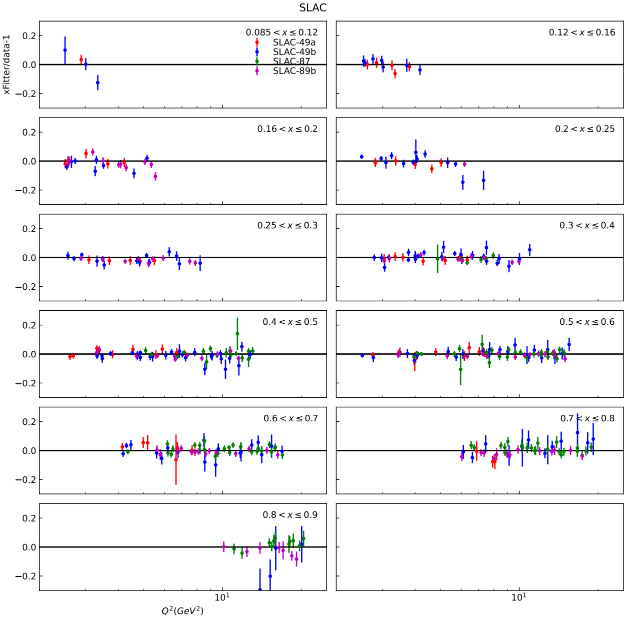


- Differences $< 1\%$

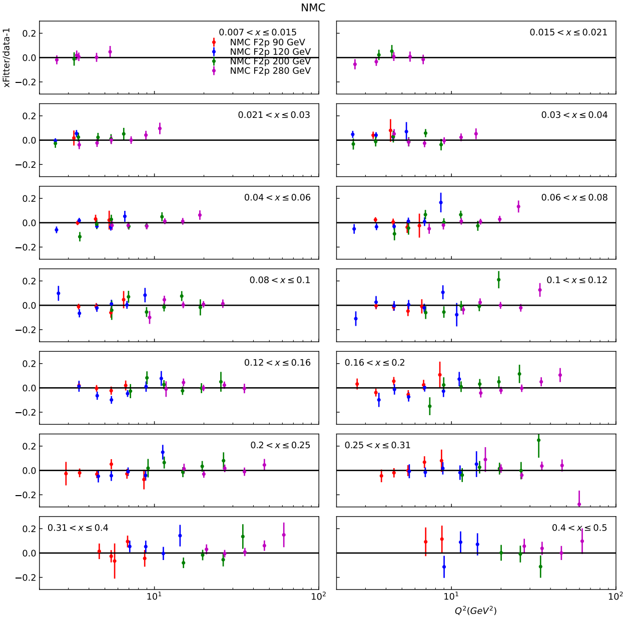
Theory (ABMP16 reproduced with xfitter) vs data

Dataset	# of points	χ^2 (ABMP16)	χ^2 (xFitter)
BCDMS	352	411	411
NMC	245	343	343
SLAC-49a	59	38	39
SLAC-49b	154	171	172
SLAC-87	109	103	102
SLAC-89b	90	79	83
HERA I+II	1168	1510	1513
HERA I+II *c*	52	66	67
H1 *b*	12	5	5
ZEUS *b*	17	16	15

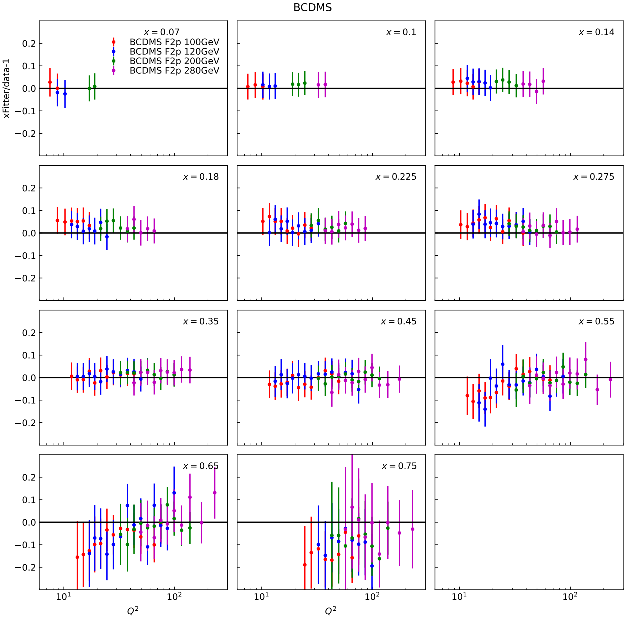
xfitter[ABMP16] vs data: SLAC



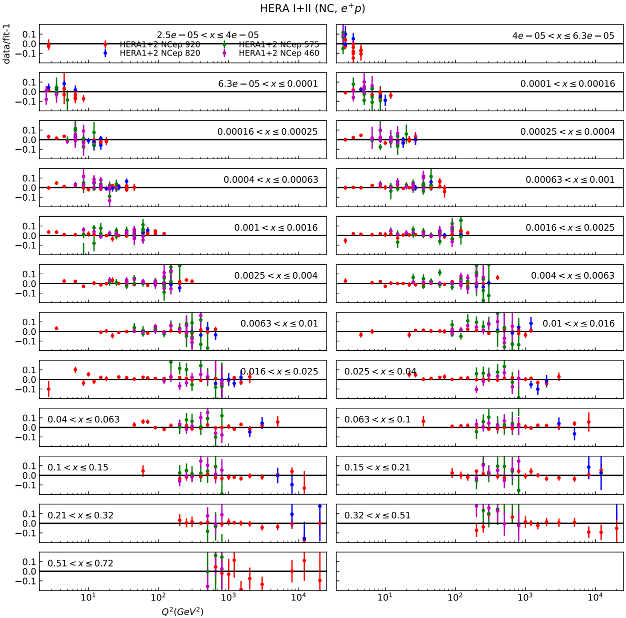
xfitter[ABMP16] vs data: NMC



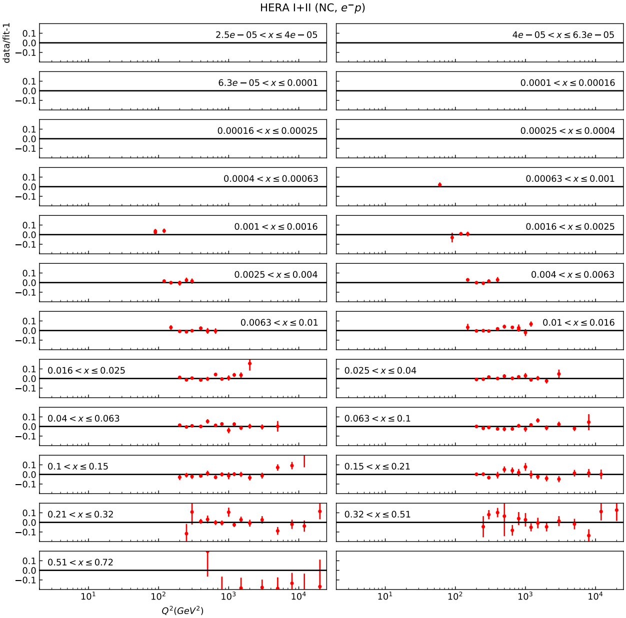
xfitter[ABMP16] vs data: BCDMS

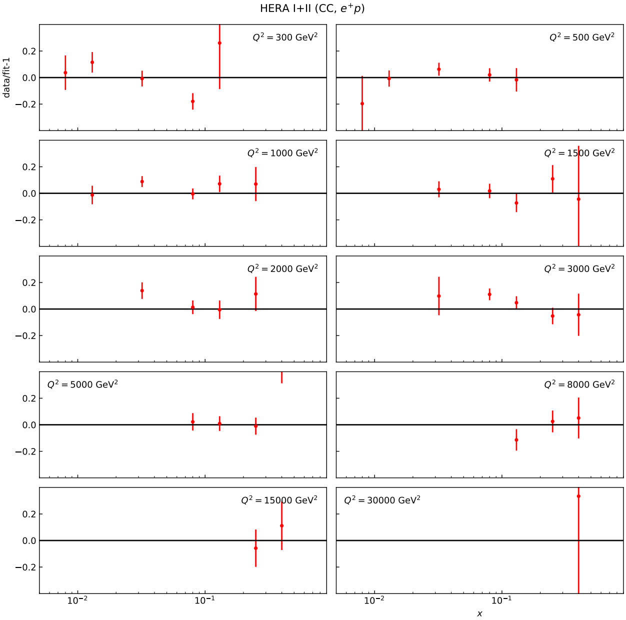


xfitter[ABMP16] vs data: HERA NC $e^{+}p$

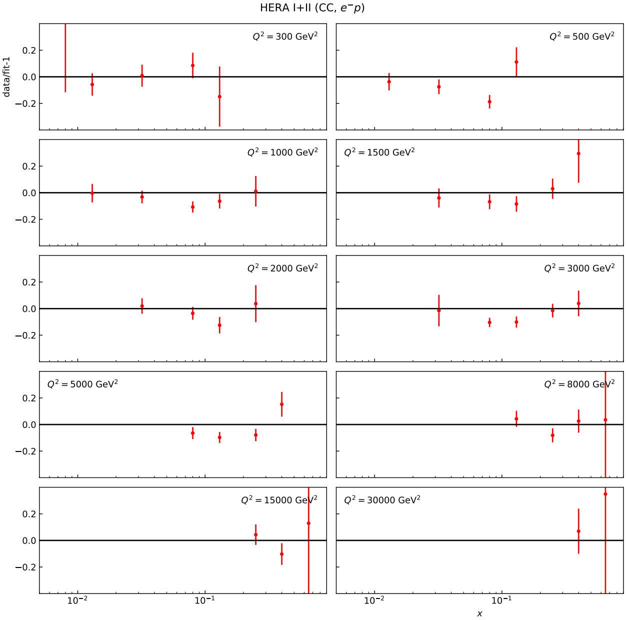


xfitter[ABMP16] vs data: HERA CC $e^- p$




HERA I+II (CC, e^+p)
data/fit-1
$Q^2 = 300$ GeV^2
$Q^2 = 500$ GeV^2
$Q^2 = 1000$ GeV^2
$Q^2 = 1500$ GeV^2
$Q^2 = 2000$ GeV^2
$Q^2 = 3000$ GeV^2
$Q^2 = 5000$ GeV^2
$Q^2 = 8000$ GeV^2
$Q^2 = 15000$ GeV^2
$Q^2 = 30000$ GeV^2
0.2
0.0
−0.2
10^{-2}
10^{-1}
x

xfitter[ABMP16] vs data: HERA CC $e^- p$



Summary

- New code for TMC and HT corrrections:
  - MR: `https://gitlab.cern.ch/fitters/xfitter/-/merge_requests/347`
  - TMC does not have any tunable parameters, but available for `FFABM` scheme only
  - HT available for any scheme, but be aware that its parametristion is taken from ABMP16 $n_f = 3$ fit and might need to be adjusted for other HF schemes
  - new test (example) `ABMP16-DIS`
  - possible further improvements:
    - ★ optimize TMC calculation
    - ★ TMC in other HF schemes
    - ★ possibility to override data file entries from `parameters.yaml`
- New NMC and SLAC data
  - MR: `https://gitlab.cern.ch/fitters/xfitter-datafiles/-/merge_requests/18`
  - breaks `ALLDATA` test due to the extra BCDMS data point (`BCDMS_F2p.280gev-thexp.dat` contains NDATA = 75 but 76 data entries) - should we update `ALLDATA` reference output?
- DIS (ep,μp) part of ABMP16 PDF fit is now available in xfitter (PDF parametrisation, theory predictions, data) and fully validated
  - this setup (or its variants) can be used as a baseline fit in future studies
  - if you use this TMC or HT implementetaion, please cite ABMP16 paper [Phys.Rev. D96 (2017) no.1, 014011] and original TMC papers [Phys.Rev. D14, 1829 (1976), Nucl.Phys. B63, 237 (1973)]