Easy visualization of experimental sensitivity to parton distributions

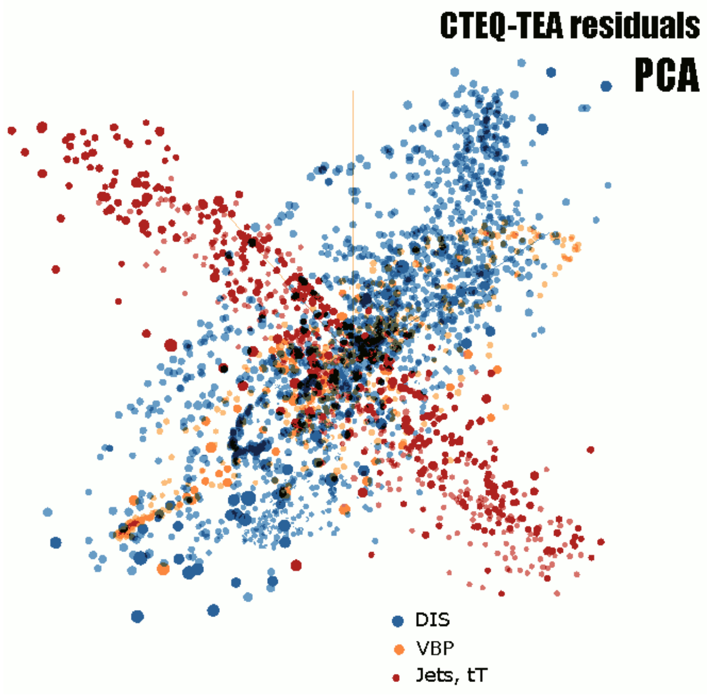
Pavel Nadolsky

Southern Methodist University

In collaboration with Bo Ting Wang, T. J. Hobbs, S. Doyle, J. Gao, T.-J. Hou, F. Olness

arXiv:1803.02777

PDFSense program, article, and figures:
http://tinyurl.com/PDFSense



Beyond CT14 parton distributions

- **Ongoing work toward the new CTEQ-TEA PDF analysis**

  [T.-J, Hou, DIS'2018 workshop]
- **CT14 QEDinc PDFs**: constraints on photon PDFs in the nucleon **[1509.02905]**
- **CT14 MC PDFs:** Monte-Carlo replicas for certain applications **[1607.06066]**
- **CT14 HERA2 PDFs:** effects of combined HERA1+2 data **[1609.07968]**
- **CT14 IC PDFs:** intrinsic/fitted charm component **[1706.00657]**

- NLO calculation for c, b production at LHCb, ATLAS in the S-ACOT-χ
- scheme using MCFM/Applgrid **[Campbell, P. N., Xie, arXiv:1803.xxxxx]**
- NNLO fast interface for charged-current DIS **[Gao et al.]**
- Programs for fast survey **[PDFSense, this talk]** and Hessian reweighting of the data **[ePump]**

How sensitive is an experiment to a PDF? Can we know it **before** doing the global fit?

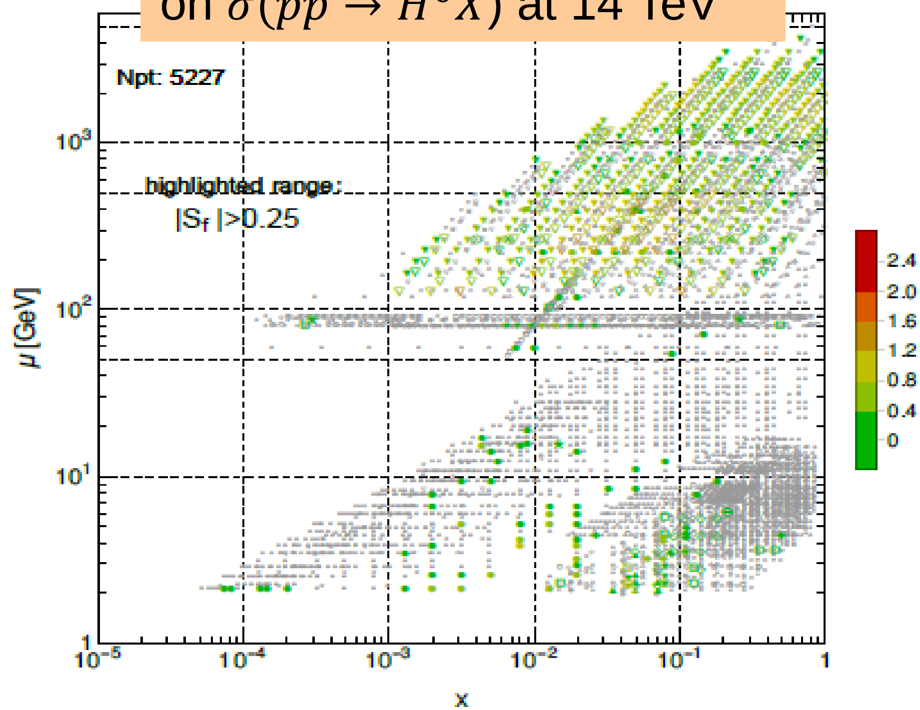
PDFSense estimates…

…ranking of strength of sensitivities of experimental data sets to PDF flavors without (re-)doing the full global fit

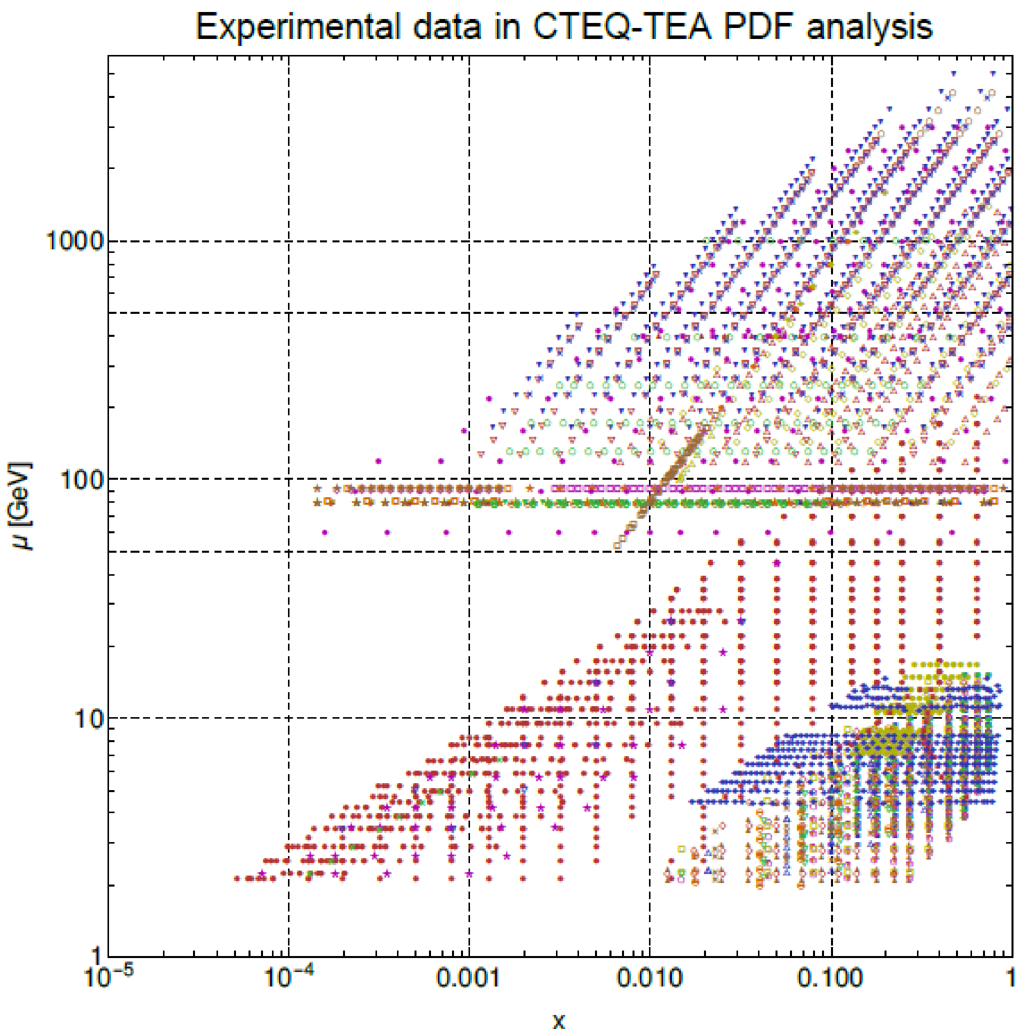
					Rankings											
No.	Exp. ID	N_{pt}	$\sum_f \lvert S_f\rvert$	$\sum_f \langle\lvert S_f\rvert\rangle/N_f$	$\lvert S_{\bar u}\rvert$	$\langle\lvert S_{\bar u}\rvert\rangle$	$\lvert S_{\bar d}\rvert$	$\langle\lvert S_{\bar d}\rvert\rangle$	$\lvert S_g\rvert$	$\langle\lvert S_g\rvert\rangle$	$\lvert S_u\rvert$	$\langle\lvert S_u\rvert\rangle$	$\lvert S_d\rvert$	$\langle\lvert S_d\rvert\rangle$	$\lvert S_s\rvert$	$\langle\lvert S_s\rvert\rangle$
1	160	1120.	620.	0.0922	**A**	3	B		**A**	3	**A**	3	B		C	
2	**545**	288	397.	0.234	B	3	B	3	**A**	**1**	C		C	3	B	3
3	**542**	280	359.	0.217	B	3	B	3	**A**	**1**	C		C		B	3
4	201	238	225.	0.158	B	2	B	2			C	3				
5	111	86	218.	0.423	C	**1**	C	**1**		3	B	**1**	C	2		
6	204	368	206.	0.0942	B	3	C	3	C		C	3				
7	101	337	184.	0.0909	C				C		B	3	C			
8	104	123	169.	0.229			C	2			C	2	B	2		
9	102	250	141.	0.0938			C		C	3	C	3	C	3		
10	109	96	115.	0.199	C	2	C	2		3	C	2	C	3		
11	538	222	109.	0.0834					C	3						
12	110	69	89.3	0.216		3		3	C	2		3		2		3
13	**250**	84	82.9	0.165		3	C	3		3		3	C	2		
14	108	85	82.4	0.161		3		3		3		3	C	3		
15	**544**	236	79.8	0.0573					C	3						
16	268	82	79.3	0.161		3		3				3		3		3
17	**249**	66	78.3	0.198		3		2				3		2		3
18	**252**	94	68.5	0.121		3		3				3				3
19	203	30	66.6	0.37	C	**1**	C	**1**				3		2		
20	**245**	66	60.3	0.152		3		3		3		3		3		
21	124	38	58.9	0.258		3		3				3		3	C	**1**

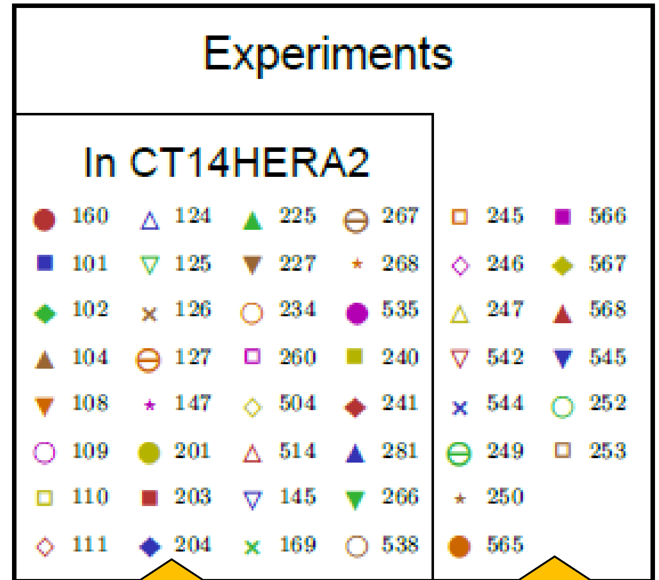
…kinematical distributions of sensitivities to the PDFs in the $\{x, \mu\}$ plane

Sensitivity to the PDF error on $\sigma(pp \to H^0 X)$ at 14 TeV



PDFSense: motivation





in the CT14 HERA2 NNLO fit (N_d =3250 points)

considered for the CTEQ-TEA fit (N_d=734 points from ATLAS, CMS, LHCb)

Experiments are labeled according to the experimental ID’s in the CT fit

1xx: DIS, **2xx**: vector boson production, **5xx**: jet and $t\bar{t}$ production

Experiments in the CT14 HERA2 fit

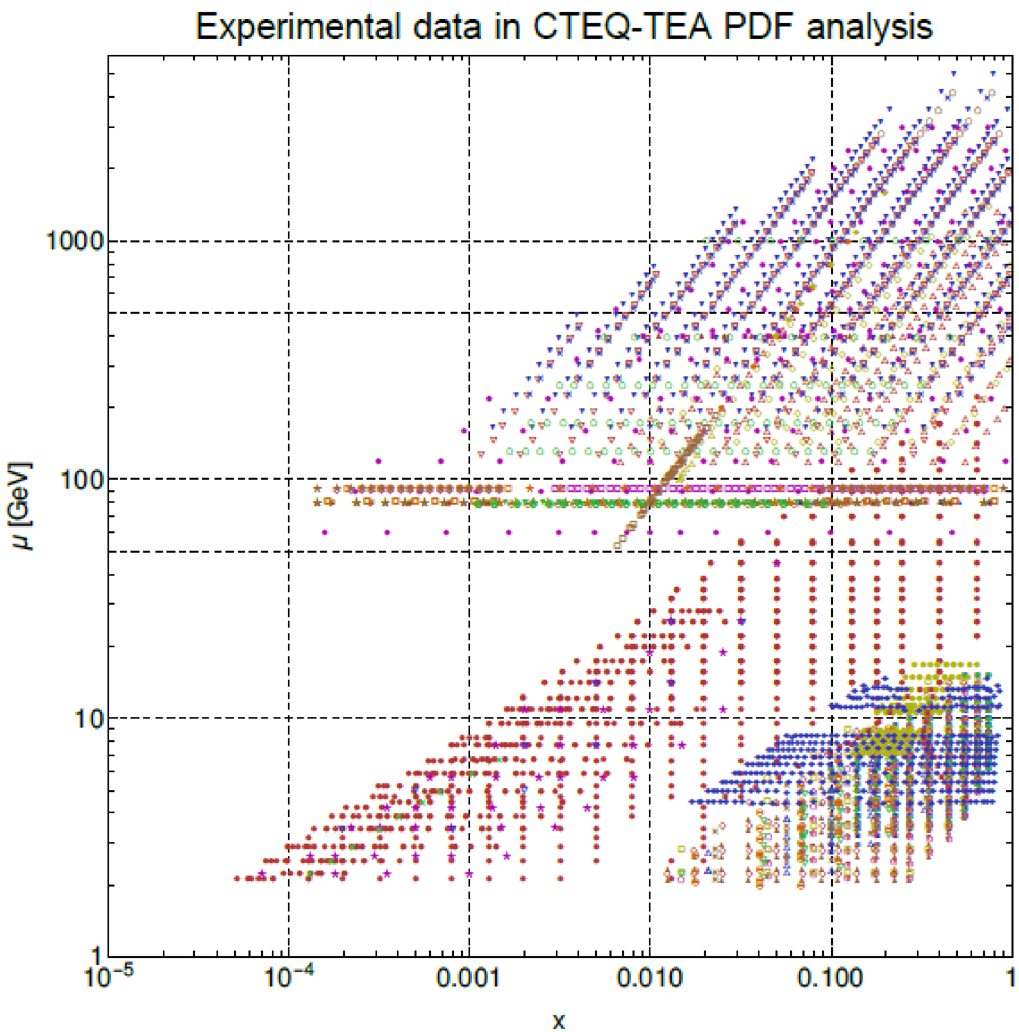
ID#	Experimental dataset		N_d
101	BCDMS F_2^p	[47]	337
102	BCDMS F_2^d	[48]	250
104	NMC F_2^d/F_2^p	[49]	123
108	CDHSW F_2^p	[50]	85
109	CDHSW F_3^p	[50]	96
110	CCFR F_2^p	[51]	69
111	CCFR xF_3^p	[52]	86
124	NuTeV $\nu\mu\mu$ SIDIS	[40]	38
125	NuTeV $\bar{\nu}\mu\mu$ SIDIS	[40]	33
126	CCFR $\nu\mu\mu$ SIDIS	[41]	40
127	CCFR $\bar{\nu}\mu\mu$ SIDIS	[41]	38
145	H1 σ_r^b (57.4 pb^{-1})	[53][54]	10
147	Combined HERA charm production (1.504 fb^{-1})	[39]	47
160	HERA1+2 Combined NC and CC DIS (1 fb^{-1})	[6]	1120
169	H1 F_L (121.6 pb^{-1})	[55]	9

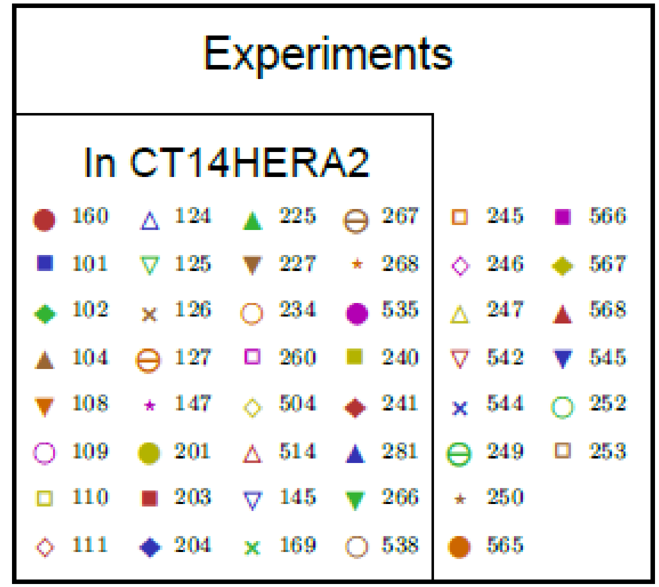
ID#	Experimental dataset		N_d
201	E605 DY	[56]	119
203	E866 DY, $\sigma_{pd}/(2\sigma_{pp})$	[57]	15
204	E866 DY, $Q^3 d^2\sigma_{pp}/(dQdx_F)$	[58]	184
225	CDF Run-1 $A_e(\eta^e)$ (110 pb^{-1})	[59]	11
227	CDF Run-2 $A_e(\eta^e)$ (170 pb^{-1})	[60]	11
234	DØ Run-2 $A_\mu(\eta^\mu)$ (0.3 fb^{-1})	[61]	9
240	LHCb 7 TeV W/Z muon forward-η Xsec (35 pb^{-1})	[62]	14
241	LHCb 7 TeV W $A_\mu(\eta^\mu)$ (35 pb^{-1})	[62]	5
260	DØ Run-2 Z $d\sigma/dy_Z$ (0.4 fb^{-1})	[63]	28
266	CMS 7 TeV $A_\mu(\eta)$ (4.7 fb^{-1})	[64]	11
267	CMS 7 TeV $A_e(\eta)$ (0.840 fb^{-1})	[65]	11
268	ATLAS 7 TeV W/Z Xsec, $A_\mu(\eta)$ (35 pb^{-1})	[66]	41
281	DØ Run-2 $A_e(\eta)$ (9.7 fb^{-1})	[67]	13
504	CDF Run-2 incl. jet $(d^2\sigma/dp_T^j dy_j)$ (1.13 fb^{-1})	[36]	72
514	DØ Run-2 incl. jet $(d^2\sigma/dp_T^j dy_j)$ (0.7 fb^{-1})	[37]	110
535	ATLAS 7 TeV incl. jet $(d^2\sigma/dp_T^j dy_j)$ (35 pb^{-1})	[68]	90
538	CMS 7 TeV incl. jet $(d^2\sigma/dp_T^j dy_j)$ (5 fb^{-1})	[69]	133

Candidate experiments in the CTEQ-TEA fit

ID#	Experimental dataset		N_d
245	LHCb 7 TeV Z/W muon forward-η Xsec (1.0 fb^{-1})	[70]	33
246	LHCb 8 TeV Z electron forward-η $d\sigma/dy_Z$ (2.0 fb^{-1})	[71]	17
247	ATLAS 7 TeV $d\sigma/dp_T^Z$ (4.7 fb^{-1})	[72]	8
249	CMS 8 TeV W muon, Xsec, $A_\mu(\eta^\mu)$ (18.8 fb^{-1})	[73]	33
250	LHCb 8 TeV W/Z muon, Xsec, $A_\mu(\eta^\mu)$ (2.0 fb^{-1})	[74]	42
252	ATLAS 8 TeV Z $(d^2\sigma/d\lvert y\rvert_{ll} dm_{ll})$ (20.3 fb^{-1})	[75]	48
253	ATLAS 8 TeV $(d^2\sigma/dp_T^Z dm_{ll})$ (20.3 fb^{-1})	[76]	45
542	CMS 7 TeV incl. jet, R=0.7, $(d^2\sigma/dp_T^j dy_j)$ (5 fb^{-1})	[34]	158
544	ATLAS 7 TeV incl. jet, R=0.6, $(d^2\sigma/dp_T^j dy_j)$ (4.5 fb^{-1})	[33]	140
545	CMS 8 TeV incl. jet, R=0.7, $(d^2\sigma/dp_T^j dy_j)$ (19.7 fb^{-1})	[35]	185
565	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dp_T^t$ (20.3 fb^{-1})	[38]	8
566	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dy_{<t/\bar{t}>}$ (20.3 fb^{-1})	[38]	5
567	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dm_{t\bar{t}}$ (20.3 fb^{-1})	[38]	7
568	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dy_{t\bar{t}}$ (20.3 fb^{-1})	[38]	5

N_d is the number of data points



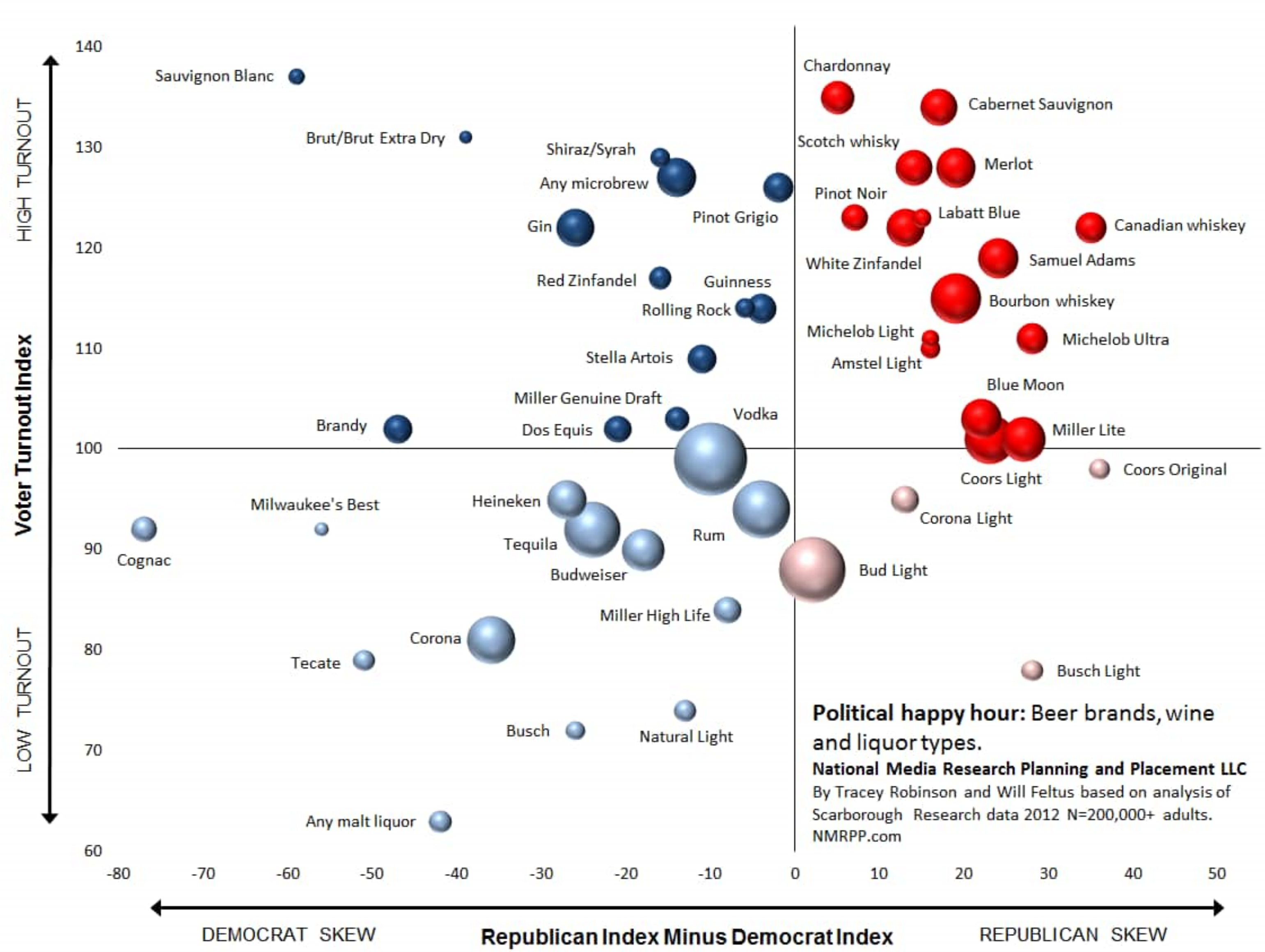


Including up to 48 experiments at NNLO QCD accuracy with many free PDF parameters incurs significant costs in complexity and CPU time.

The CTEQ-TEA fitting code underwent significant upgrades to solve this challenge

We also developed a simple test to identify high-value experiments that must be included in the global fit as the first priority.

Example: a predictive simple test in the US political science

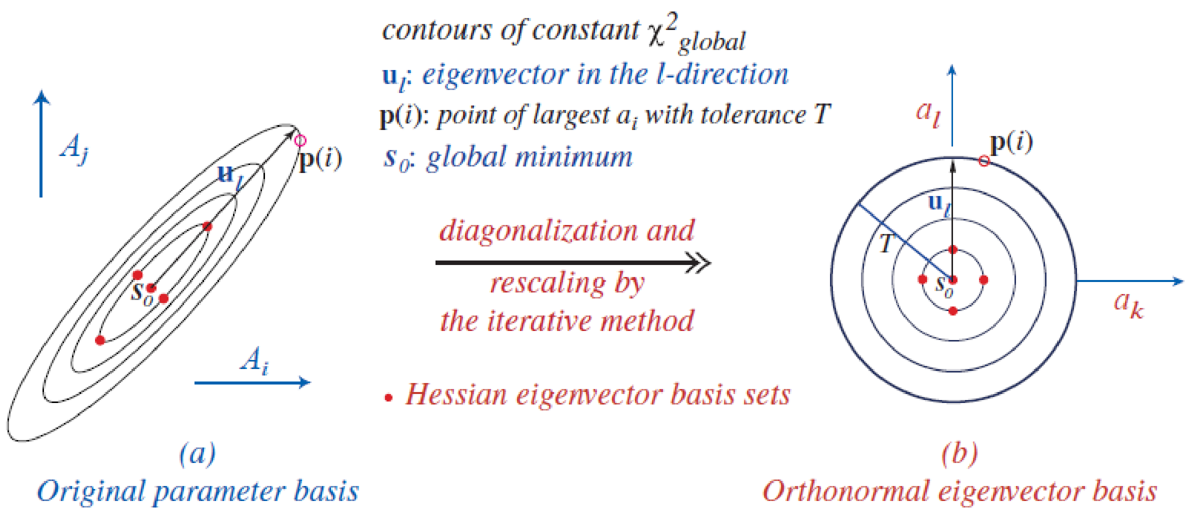


PDFSense: operating principles

- ~~PDF reweighting~~
- ~~Hessian profiling~~
- Generalized correlations (**sensitivities S_f**) comparing experimental and CT14HERA2 PDF uncertainties
- Based on the output from the CT14HERA2 fit in a new format (**shifted residuals for Hessian PDFs**)

Tolerance hypersphere in the PDF space

2-dim (i,j) rendition of N-dim (26) PDF parameter space

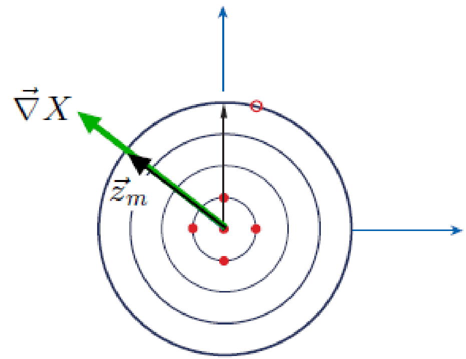


A hyperellipse $\Delta\chi^2 \leq T^{\,2}$ in space of N physical PDF parameters $\{\mathrm{A}_i\}$ is mapped onto a filled hypersphere of radius T in space of N orthonormal PDF parameters $\{\mathrm{a}_i\}$

Hessian method: Pumplin et al., 2001

Tolerance hypersphere in the PDF space

2-dim (i,j) rendition of N-dim (26) PDF parameter space



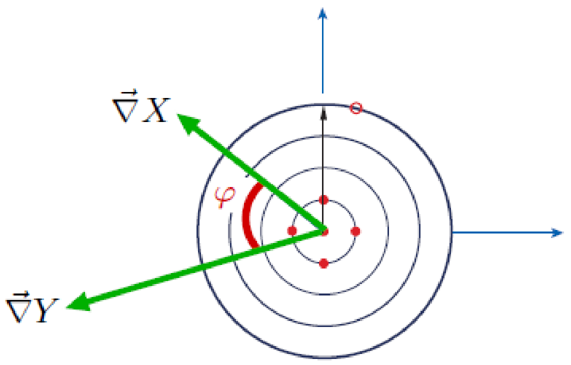
(b)
Orthonormal eigenvector basis

A symmetric PDF error for a physical observable X is given by

$$\Delta X = \vec{\nabla} X \cdot \vec{z}_m = \left|\vec{\nabla} X\right| = \frac{1}{2}\sqrt{\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)^2}$$

Tolerance hypersphere in the PDF space

2-dim (i,j) rendition of N-dim (26) PDF parameter space



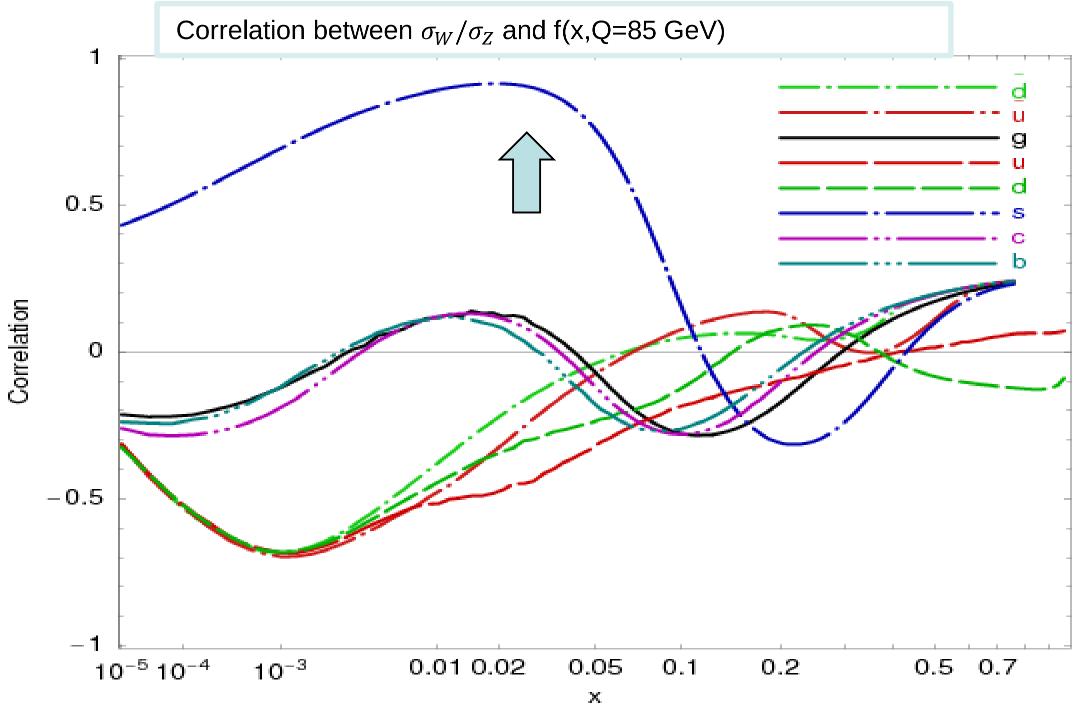
(b)
Orthonormal eigenvector basis

Correlation cosine for observables X and Y:

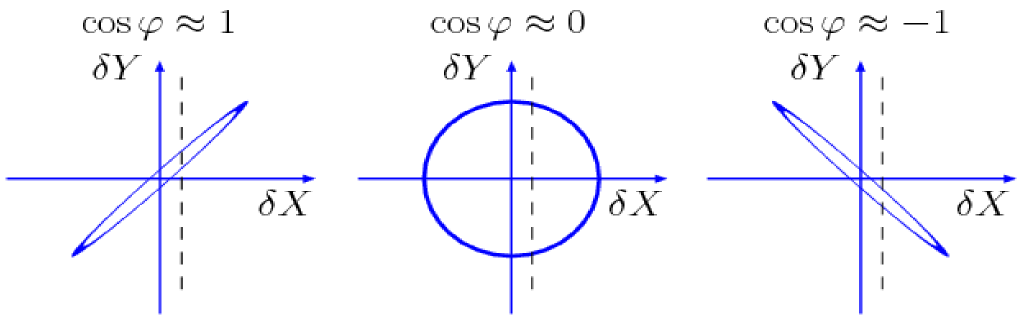
$$\cos\varphi = \frac{\vec{\nabla}X\cdot\vec{\nabla}Y}{\Delta X \Delta Y} = \frac{1}{4\Delta X\,\Delta Y}\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)\left(Y_i^{(+)} - Y_i^{(-)}\right)$$

$cos\varphi \equiv Corr[X,Y]$ -- realization of the Pearson correlation coefficient in the Hessian method

Correlations carry useful, but limited information



CTEQ6.6 [arXiv:0802.0007]:
$\cos\varphi > 0.7$ shows that the ratio σ_W/σ_Z at the LHC must be sensitive to the strange PDF $s(x,Q)$



$\cos\varphi \approx \pm 1$ suggests that a measurement of X **may** impose tight constraints on Y

But, Corr[X,Y] between theory cross sections X and Y does not tell us how the experimental uncertainty on X affects Y

Solution: choose X = a shifted residual r_i

The $\chi^2(\vec{a})$ for experiment E is

$$\chi^2(\vec{a}) = \sum_{i=1}^{N_d} r_i^2\,(\vec{a}) + \sum_{\alpha=1}^{N_\lambda} \bar{\lambda}_\alpha^2\,(\vec{a}) \approx \sum_{i=1}^{N_d} r_i^2\,(\vec{a})$$

$r_i(\vec{a}) = \frac{T_i(\vec{a}) - D_i^{sh}(\vec{a})}{s_i}$ is a **shifted residual** for point i, PDF parameters $\vec{a}$

$\bar{\lambda}_\alpha(\vec{a})$ are N_λ **optimized nuisance parameters** (dependent on $\vec{a}$)

$T_i(\vec{a})$ is the theory prediction for PDF parameters $\vec{a}$
D_i^{sh} is the data value including the optimal systematic shift

$$D_i^{sh}(\vec{a}) = D_i - \sum_{\alpha=1}^{N_\lambda} \beta_{i\alpha}\bar{\lambda}_\alpha(\vec{a})$$

s_i is the uncorrelated error

Solution: choose X = a shifted residual r_i

The CTEQ-TEA fit returns tables of $r_i(\vec{a})$ and $\bar{\lambda}_\alpha(\vec{a})$ for every i and α

Alternatively, they can be found from the covariance matrix:

$$r_i(\vec{a}) \; = \; s_i \sum_{j=1}^{N_d} (\mathrm{cov}^{-1})_{ij} \; (T_j(\vec{a}) - D_j)\, , \qquad \overline{\lambda}_\alpha(\vec{a}) = \sum_{i,j=1}^{N_d} (\mathrm{cov}^{-1})_{ij} \frac{\beta_{i\alpha}}{s_i} \frac{(T_j(\vec{a}) - D_j)}{s_j}$$

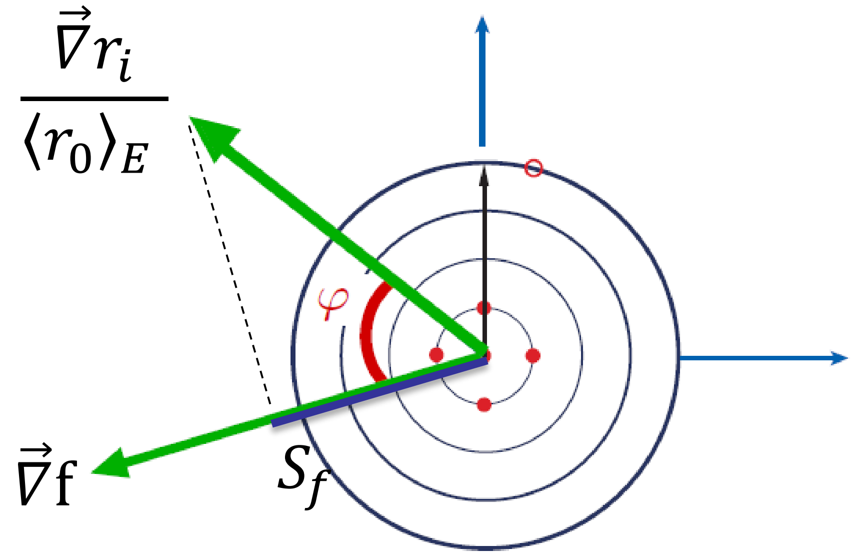
N_{pt}

N_{pt}

Correlation C_f and sensitivity S_f

Given $2N+1$ Hessian eigenvectors for the shifted residual r_i and PDF-dependent quantity f, **define**:

- $\vec{\rho}_i \equiv \vec{\nabla} r_i / \langle r_0 \rangle_E$ -- conveniently normalized to $\langle r_0 \rangle_E$, the root-mean-squared residual for the experiment E

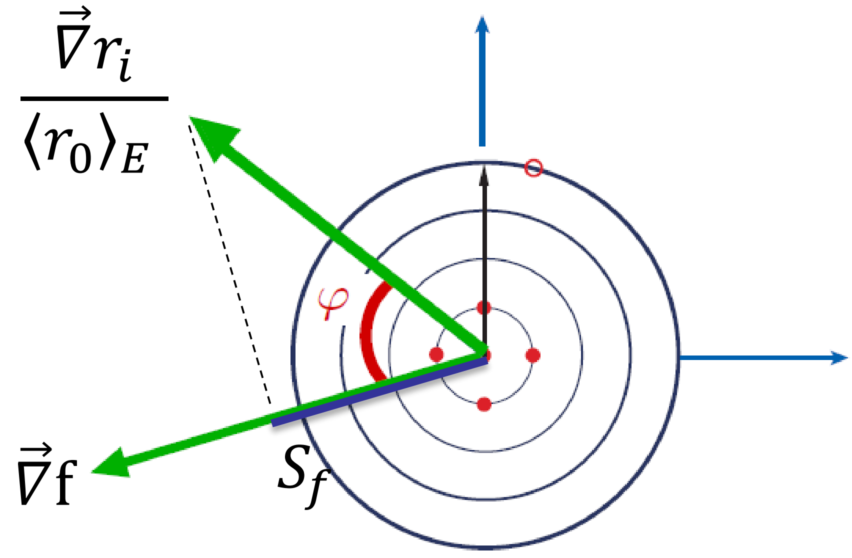


$$\langle r_0 \rangle_E \equiv \sqrt{\frac{1}{N_d}\sum_{i=1}^{N_d} r_{0,i}^2} \approx \sqrt{\frac{\chi_E^2}{N_d}}$$

$\langle r_0 \rangle_E \approx 1$ in a good fit to E

Correlation C_f and sensitivity S_f

Given $2N + 1$ Hessian vectors for the shifted residual r_i and PDF-dependent quantity f, **define**:



- $C_f \equiv \mathrm{Corr}[\rho_i(\vec{a})), f(\vec{a})] = cos\varphi$

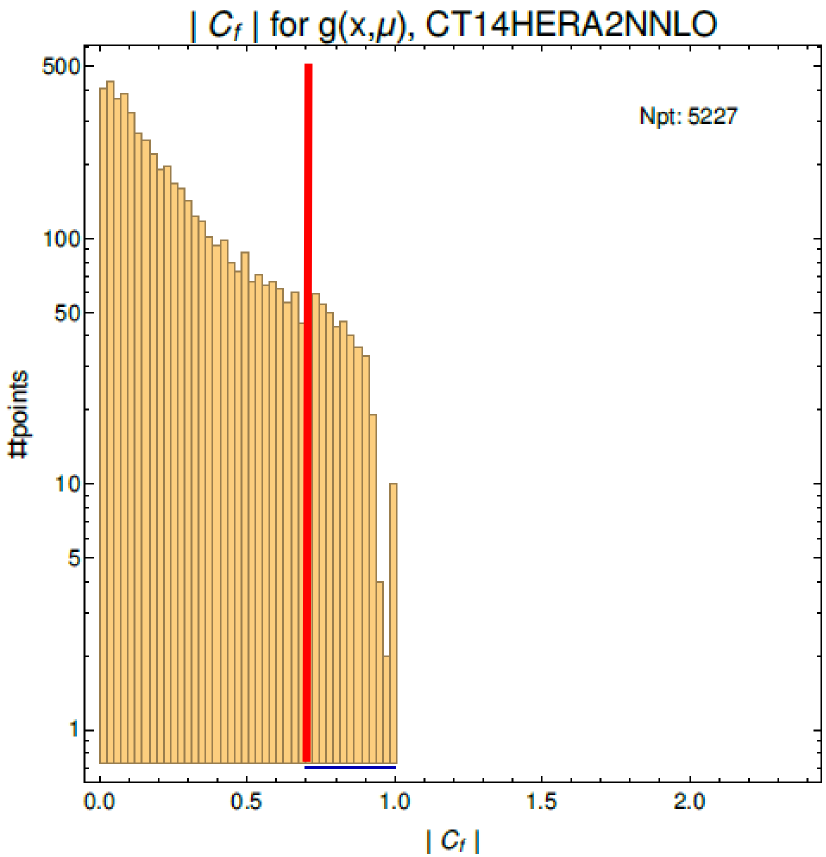
C_f is **independent** of the experimental and PDF uncertainties

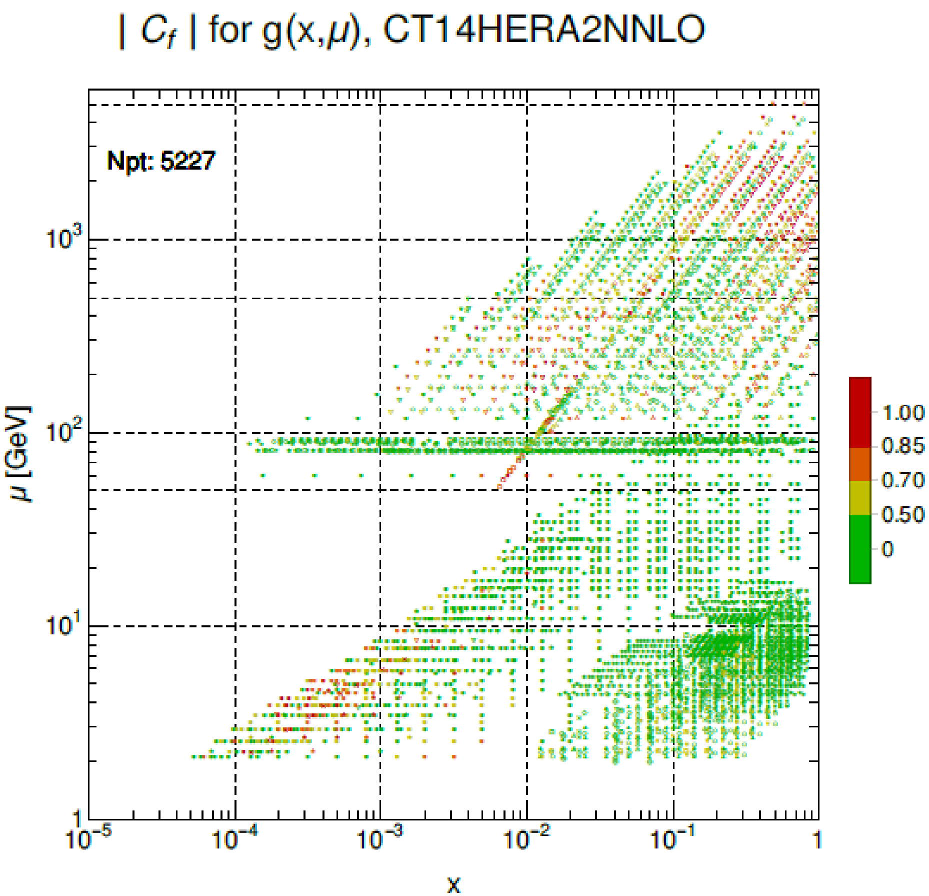
- $S_f \equiv |\vec{\rho}_i| cos\varphi = C_f \frac{\Delta r_i}{\langle r_0 \rangle_E}$ -- projection of $\vec{\rho}_i(\vec{a})$ on $\vec{\nabla} f$

S_f is proportional to $\cos\varphi$ and the ratio of the PDF uncertainty to the experimental uncertainty. We can sum $|S_f|$.

Absolute correlation C_f for $f = g(x_i, \mu_i)$

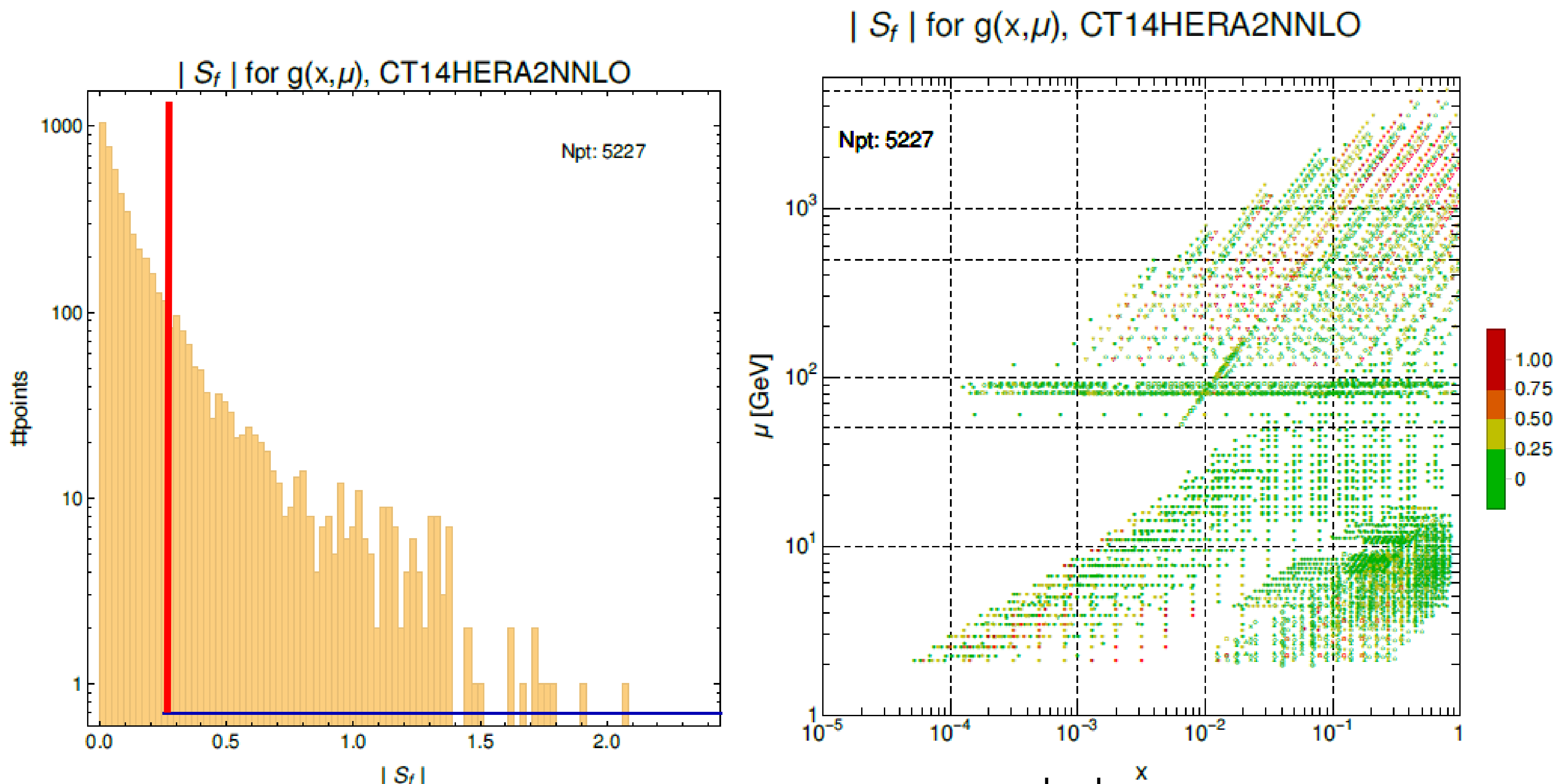
The PDFs are evaluated at the same $\{x, \mu\}$ as each data point





443 data points have a strong correlation, taken to be $|C_f| > 0.7$.

Absolute correlation S_f for $f = g(x_i, \mu_i)$

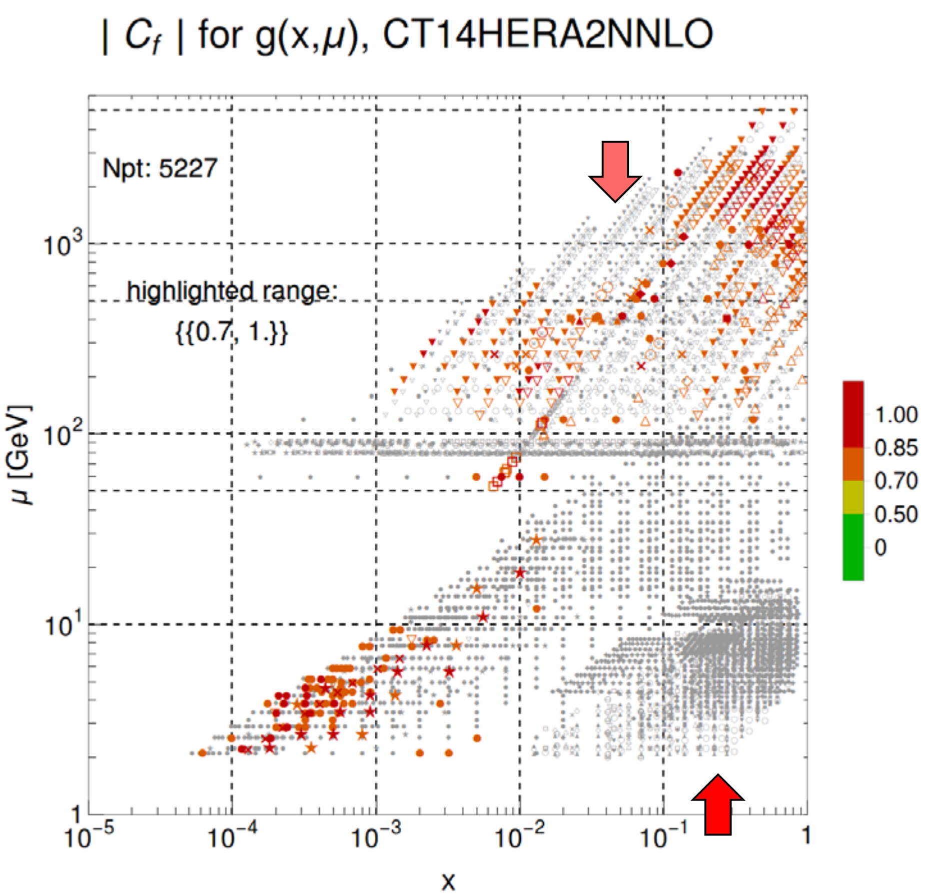


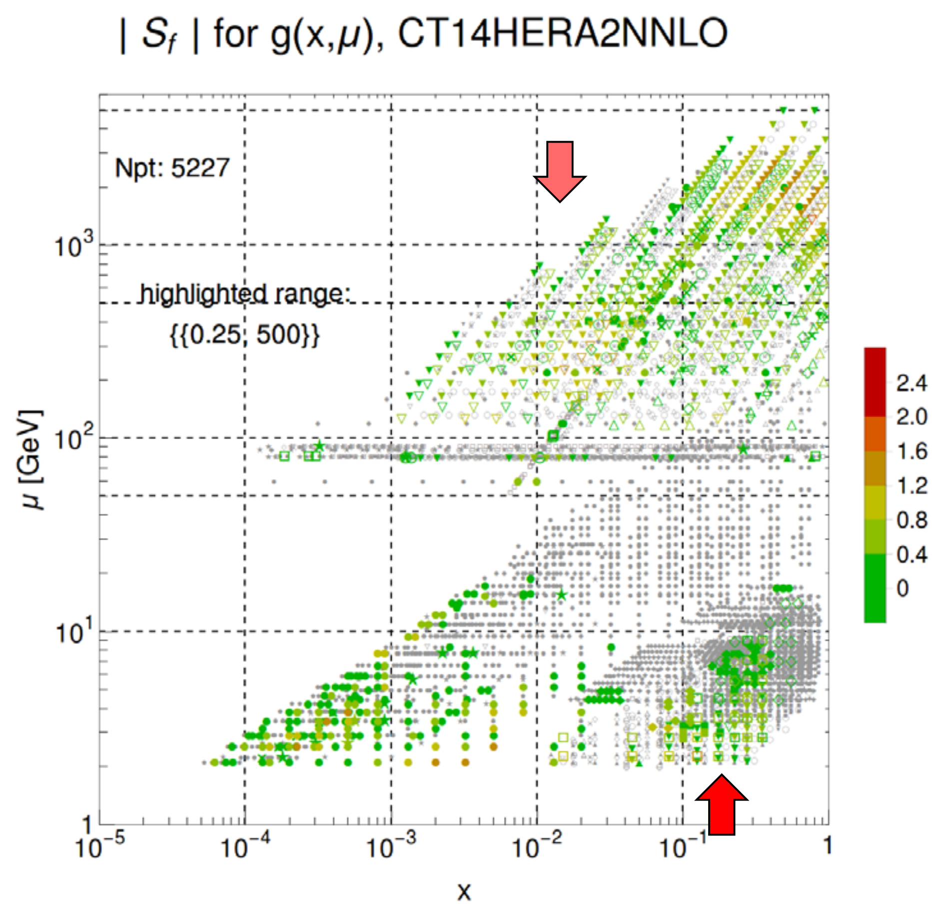
990 data points have a strong sensitivity, taken to be $|S_f| > 0.25$ and contributing $\Delta\chi^2 \gtrsim 0.0625$. These points likely contribute to the constraint.

Only showing $|C_f| > 0.7$, $|S_f| > 0.25$

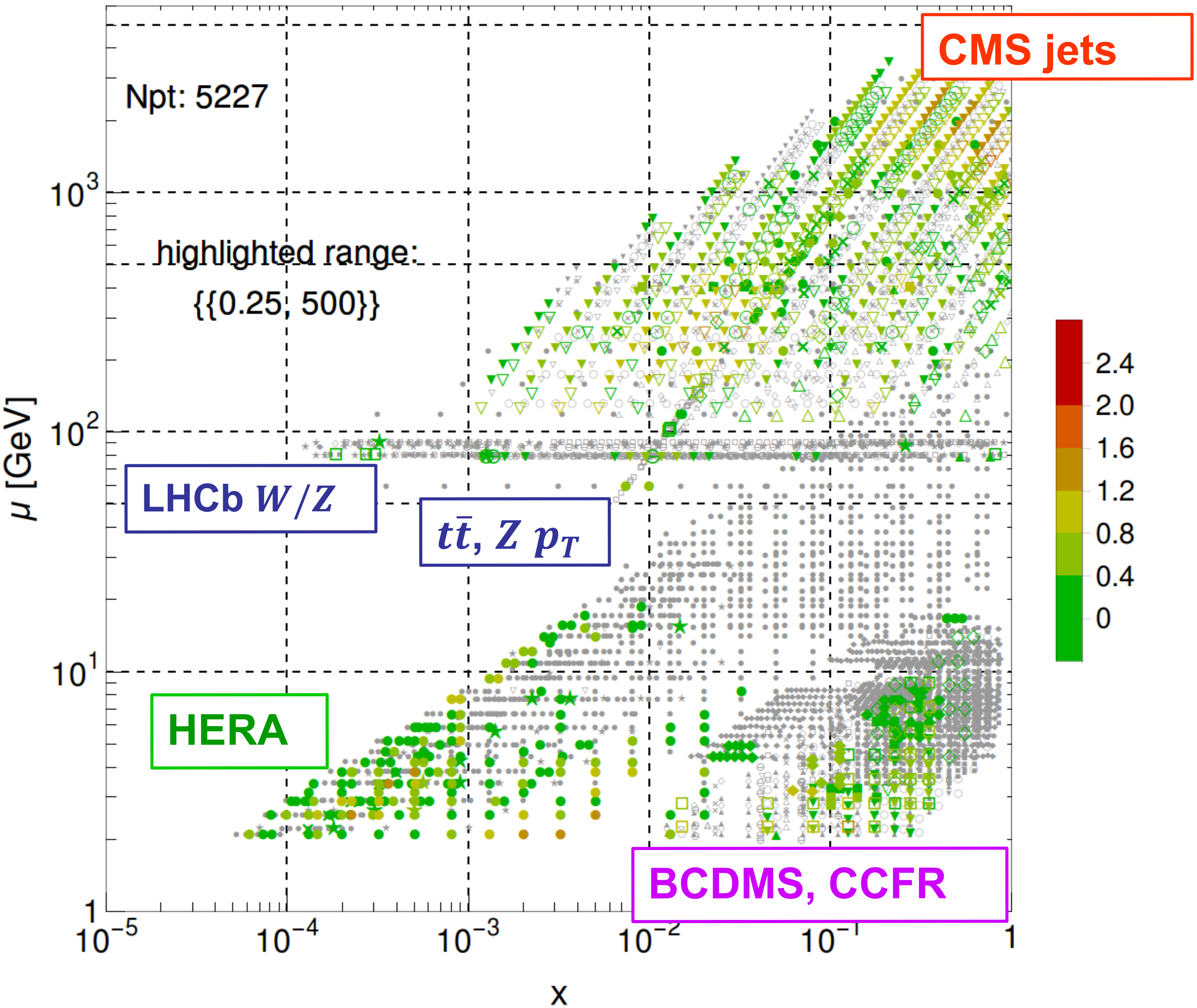
C_f does not identify fixed-target DIS points sensitive to $g(x_i, \mu_i)$. But S_f does.

Corr. syst. errors smear S_f over many data points for jet production, etc.

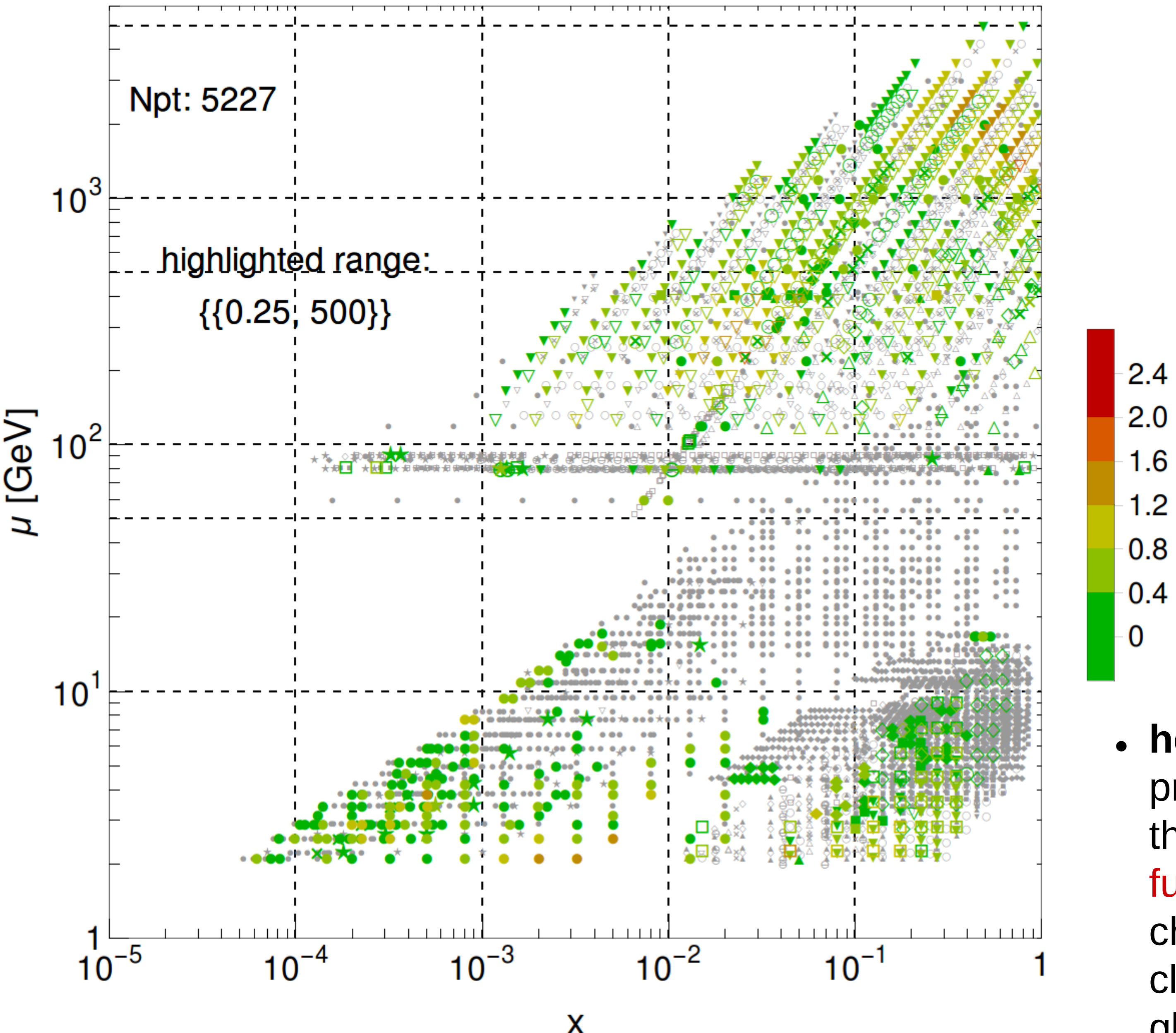


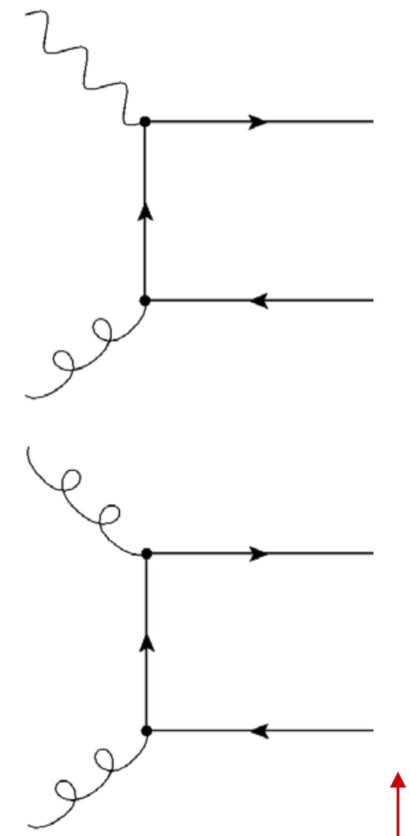


| S_f | for g(x,μ), CT14HERA2NNLO



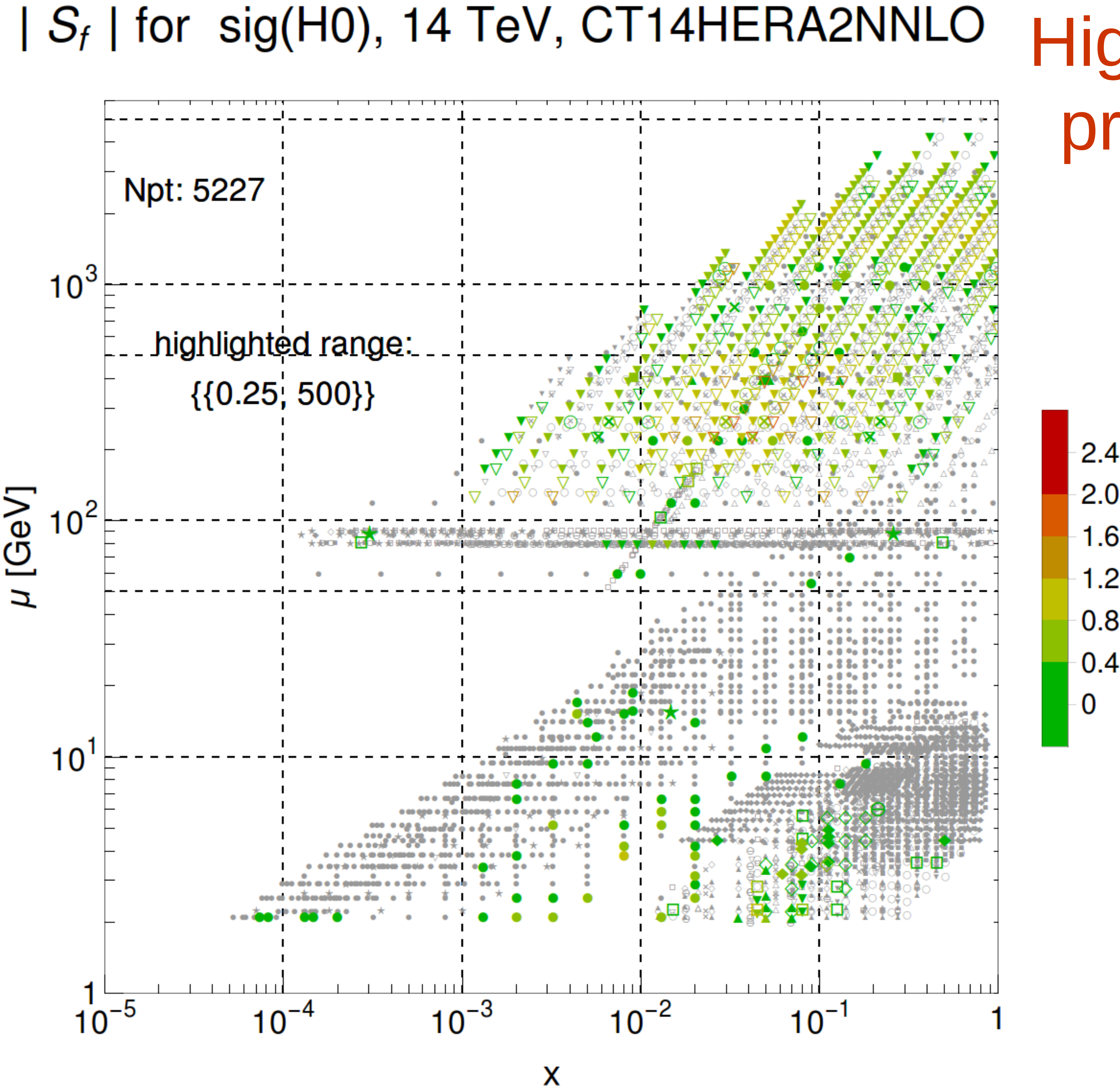
$| S_f |$ for c(x,μ), CT14HERA2NNLO





- **heavy quark** production proceeds through boson fusion diagrams: charm sensitivities closely track the gluon plot

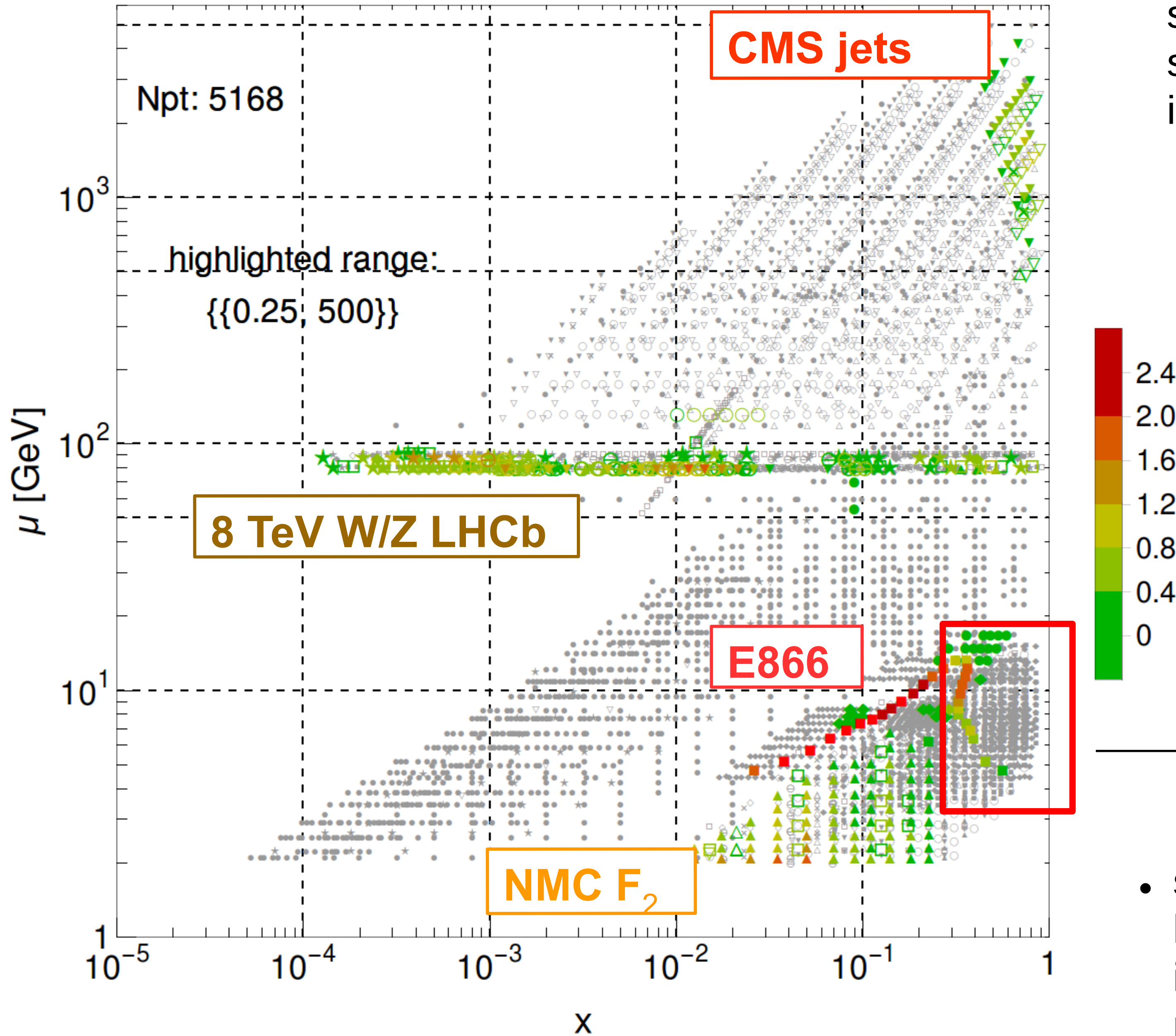
Higgs boson production



We find CMS/ATLAS **inclusive jet production** to have the dominant sensitivity!

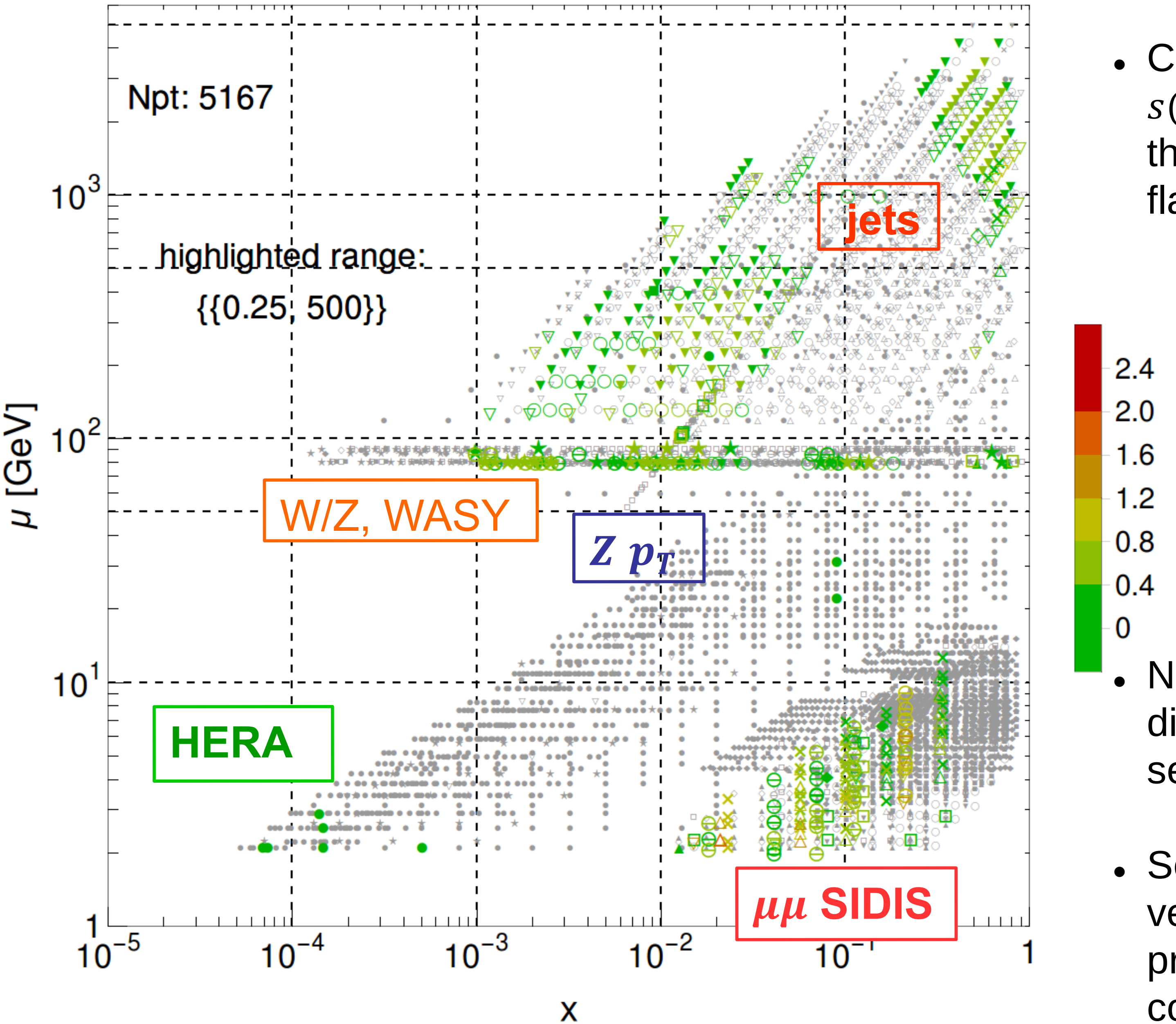
Good correlations with some points in HERA DIS, E866, CCFR, $Z\ p_T$ and $t\bar{t}$ production; but fewer points, larger exp. errors in these processes

| S_f | for $\bar{d}/\bar{u}(x,\mu)$, CT14HERA2NNLO



- PDF ratio is sensitive to flavor symmetry breaking in the light quark sea
  - the large E866 pd/pp sensitivity *degrades* at larger x
  - this is a prime motivation for higher x DY measurements at **E906** (SeaQuest)
- some contribution at high x from CMS inclusive jet production

| S_f | for s(x,μ), CT14HERA2NNLO



- Constraints on $s(x,\mu)$ are weaker than on the other flavors
- NuTeV, CCFR dimuon SIDIS most sensitive
- Sensitivities of vector boson production, jets are comparable

Summary tables: sample ranking

No.	Exp. ID	N_d	$\sum_f \lvert S_f^E\rvert$	$\langle\sum_f \lvert S_f^E\rvert\rangle$	$\lvert S_{\bar d}^E\rvert$	$\langle\lvert S_{\bar d}^E\rvert\rangle$	$\lvert S_{\bar u}^E\rvert$	$\langle\lvert S_{\bar u}^E\rvert\rangle$	$\lvert S_g^E\rvert$	$\langle\lvert S_g^E\rvert\rangle$	$\lvert S_u^E\rvert$	$\langle\lvert S_u^E\rvert\rangle$	$\lvert S_d^E\rvert$	$\langle\lvert S_d^E\rvert\rangle$	$\lvert S_s^E\rvert$	$\langle\lvert S_s^E\rvert\rangle$
1	160	1120.	620.	**HERA**			**A**	3	**A**	3	**A**	3	B		C	
2	**545**	185	260.	**CMS jets 8**		3	C	3	**A**	**1**				3	C	3
3	111	86	218.	CCF3 F3p		**1**	C	**1**		3	B	**1**	C	2		
4	**542**	158	206.	**CMS jets 7**		3	C	3	B	**1**					C	3
5	101	337	184.	BCDMS F2p			C		C		B	3	C			
6	104	123	169.	NMC		2					C	2	B	2		
7	102	250	141.	BCDMS F2d					C	3	C	3	C	3		
8	109	96	115.	CDHSW		2	C	2		3	C	2	C	3		
9	201	119	113.	E605		2	C	2				3				

Experiments are listed in the descending order of the summed sensitivities to $\bar d, \bar u, g, u, d, s$

For each flavor, A and 1 indicate the strongest total sensitivity and strongest sensitivity per point

C and 3 indicate marginal sensitivities; low sensitivities are not shown

40	**247**	8	5.84	Z pT 7 TeV	3	3	3	3	3	
41	169	9	3.99	HERA F_L			$\boxed{2}$			
42	**567**	7	3.9	$t\bar t$			$\boxed{2}$			
43	227	11	3.7	CDF WASY (2005)					3	
44	**568**	5	3.4				$\boxed{2}$			
45	**566**	5	3.19	$t\bar t$			$\boxed{2}$			
46	145	10	1.14	HERA b						

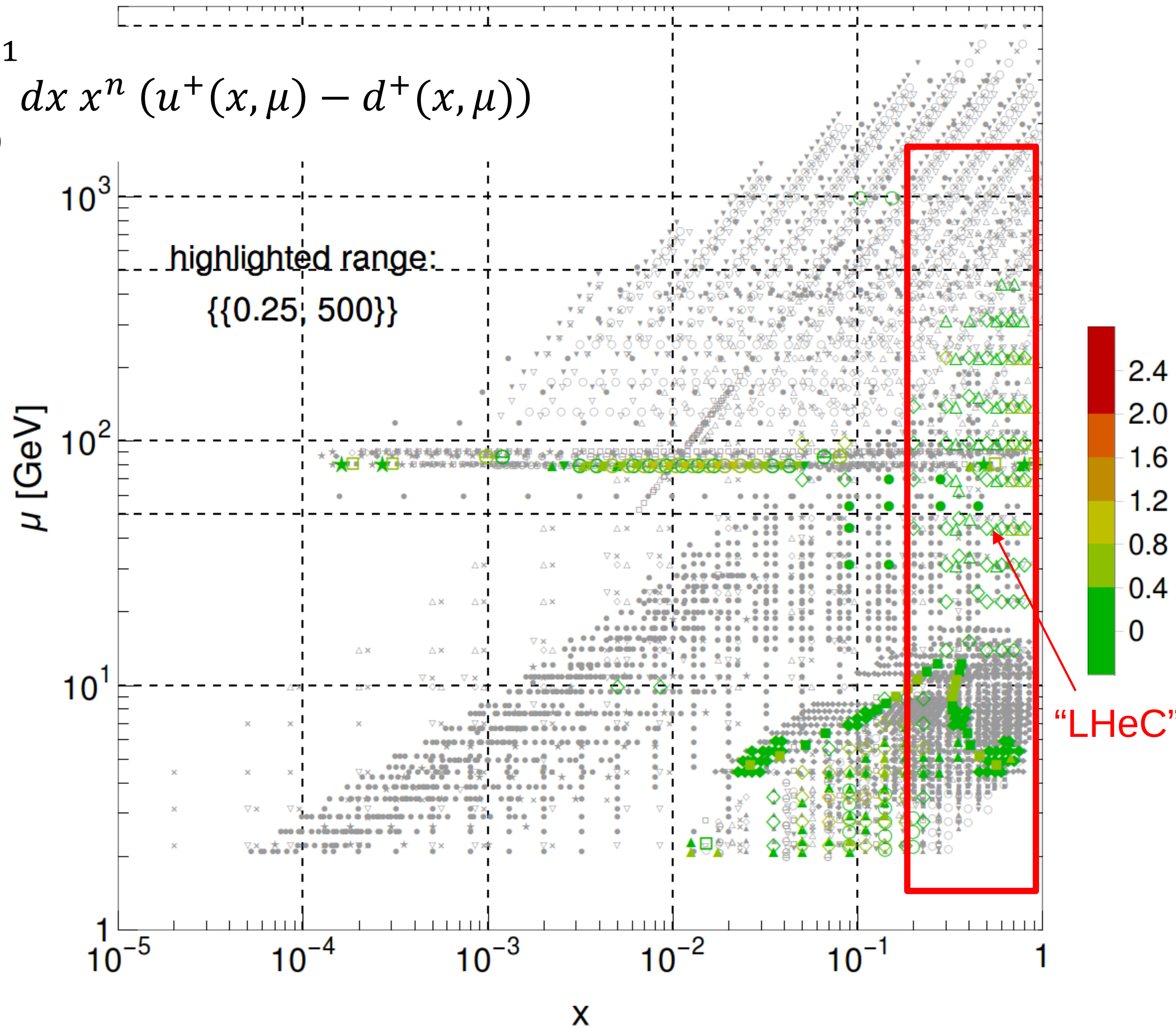
Good per-point S_f , small N_d

| S_f | for <x^1>u+−d+, CT14HERA2NNLO

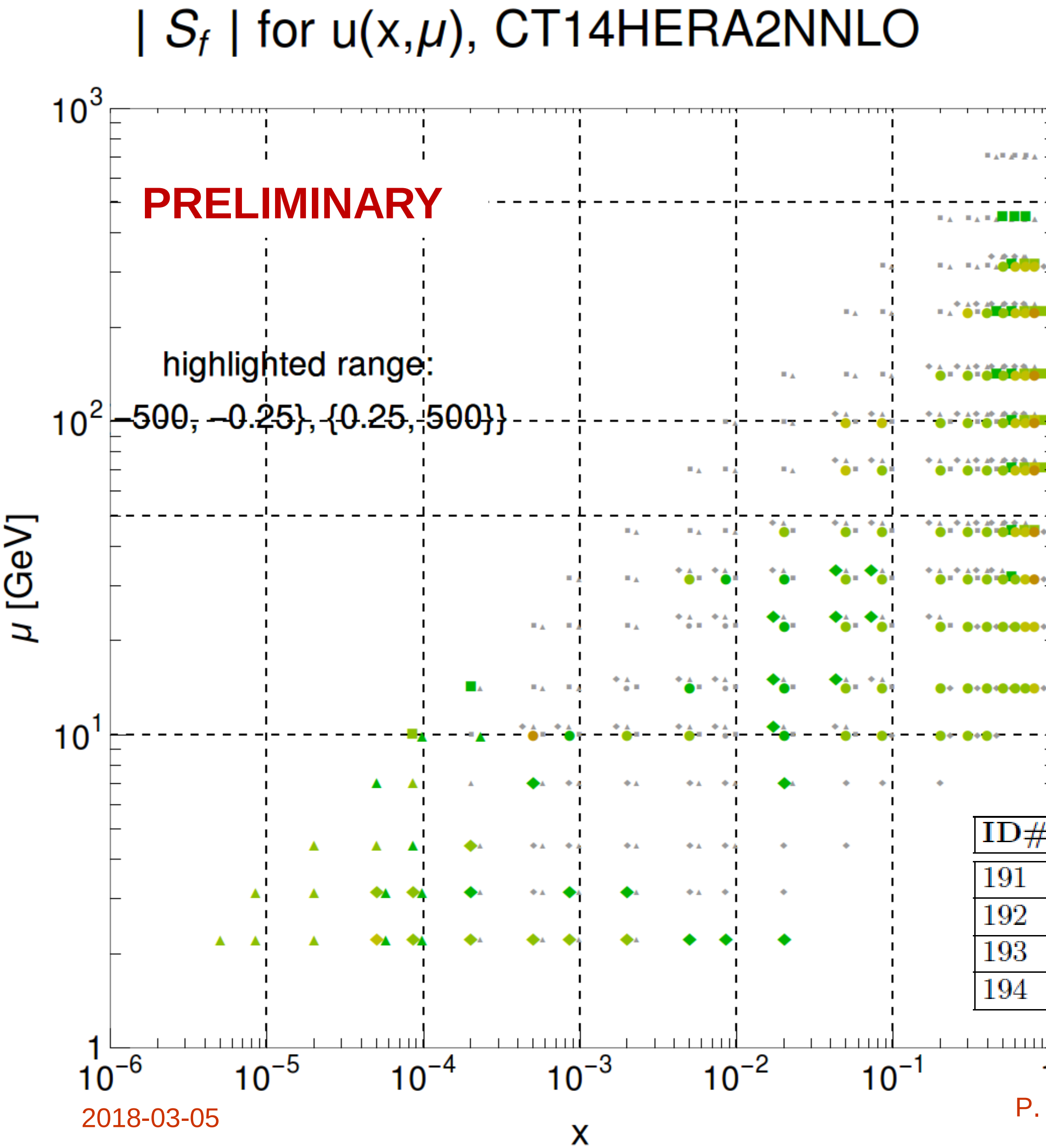
$$\langle x^n \rangle_{u^+-d^+} \equiv \int_0^1 dx\, x^n \left(u^+(x,\mu) - d^+(x,\mu)\right)$$

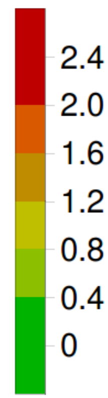
- Several 1st PDF moments can be computed on the **lattice**; S_f suggests $\langle x \rangle_{u^+-d^+}$ is constrained weakly, primarily by p, d fixed-target DIS and DY

- higher moments shift sensitivities *rightward*



Visualizing LHeC constraints





LHeC pseudodata from M. Klein and V. Radescu ($\sqrt{s} = 0.49, 1.3$ TeV; $L = 0.1\ ab^{-1}$; 1 year of running)

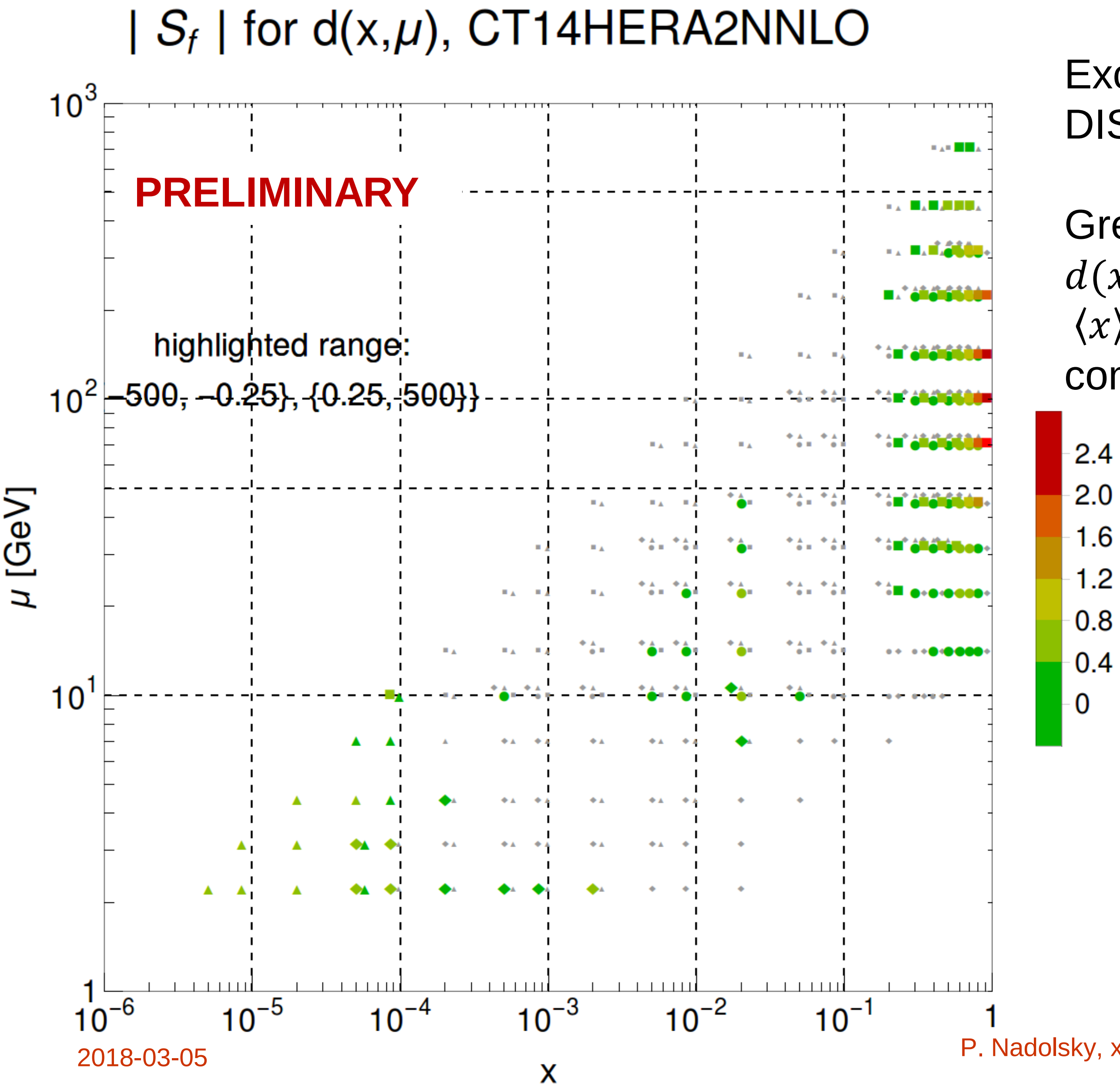
For all flavors:

NC DIS: good S_f for $x < 0.005$; Lower S_f at $x > 10^{-2}$, large y

CC DIS: **great** S_f at $x > 10^{-2}$, medium y

ID#	LHeC pseudodata set		N_{pt}	Fig. 8 symbol
191	CC e^-p	lower	93	circle
192	CC e^+p		109	square
193	NC e^-p	upper	128	diamond
194	NC e^+p		148	triangle

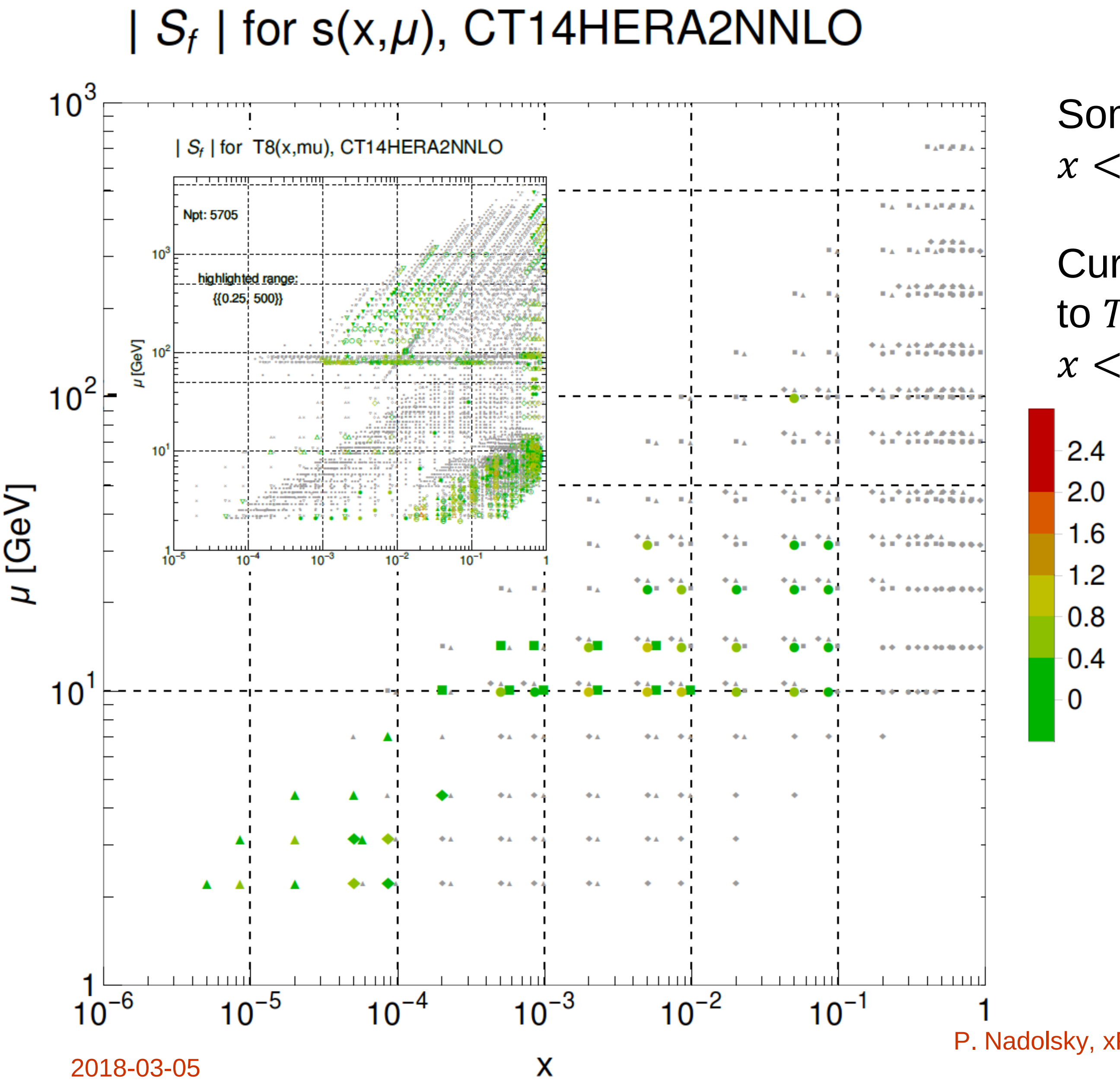
Visualizing LHeC constraints



Exceptional sensitivity in CC DIS to $d(x)$ at $x \to 1$!

Great potential to constrain $d(x)/u(x)$ at $x \to 1$, $\langle x \rangle_{u^+ - d^+}$ for lattice QCD comparisons

Visualizing LHeC constraints

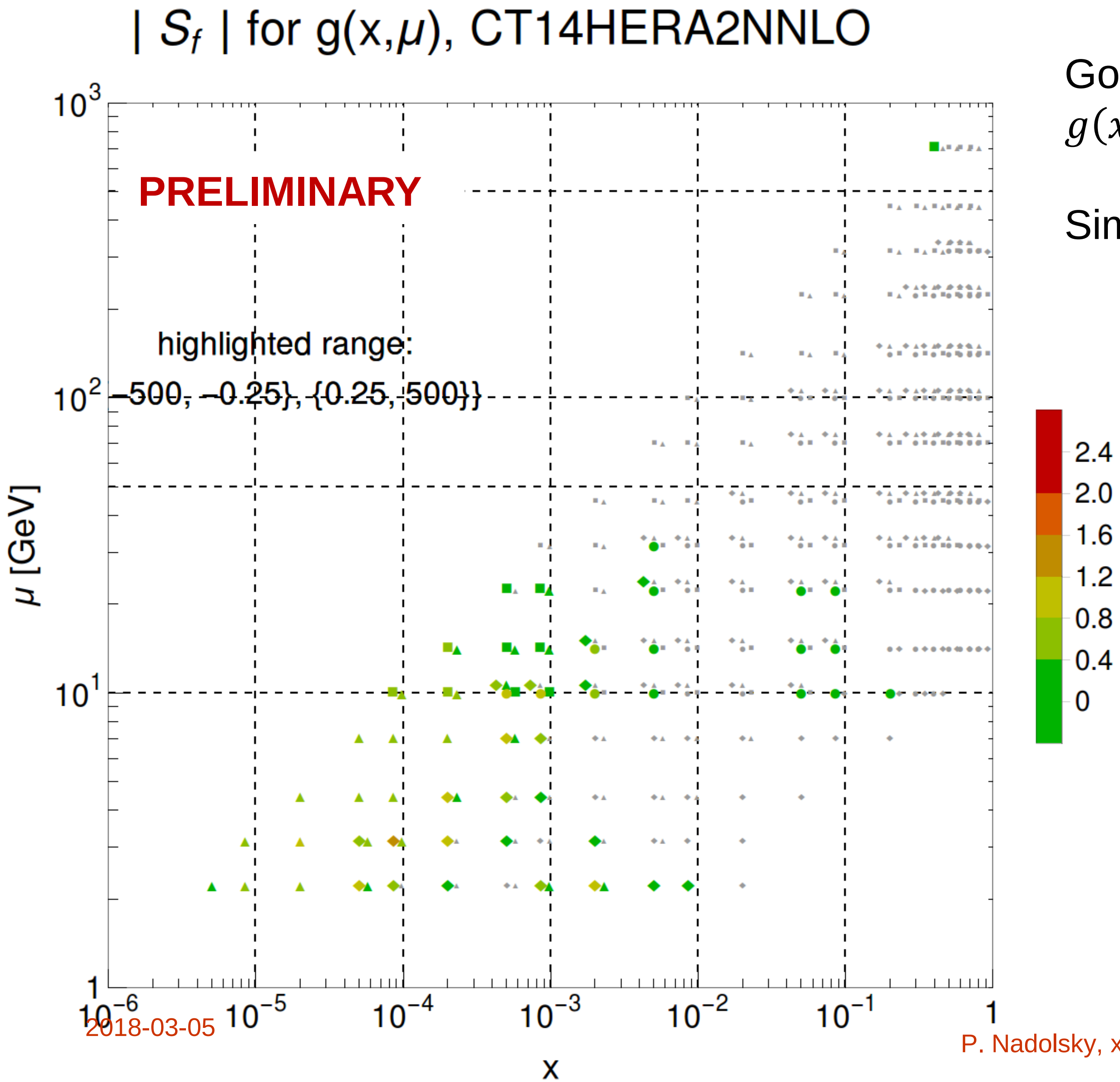


Some sensitivity to $s(x,\mu)$ at $x < 0.1$, $\mu < 40$ GeV

Current data are insensitive to $T_3(x,\mu)$ and $T_8(x,\mu)$ at $x < 10^{-3}$, $Q < 10$ GeV

$\Rightarrow$ All u, d, s (anti)quark S_f are roughly the same at such $\{x, Q\}$

Visualizing LHeC constraints



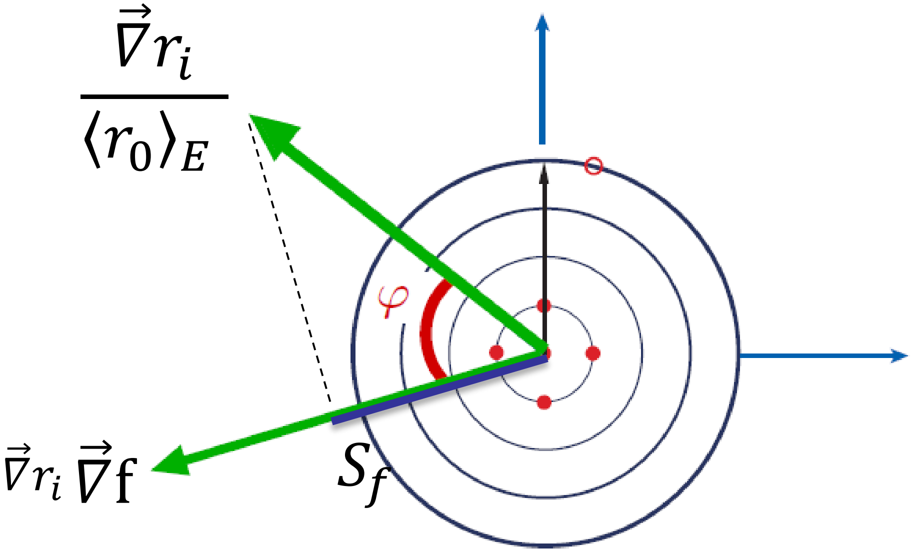
Good sensitivity to $g(x, Q)$ at $x < 10^{-2}$!

Similar $\{x, \mu\}$ maps for $\bar{u}$, $\bar{d}$

The underlying geometrical picture

S_f is a projection of a residual gradient $\vec{\nabla} r_i$ onto the direction for $\vec{\nabla} f$.

S_f captures a small part of information about the 28- (or 56)-dimensional population of $\vec{\nabla} r_i$ for 4000 data points

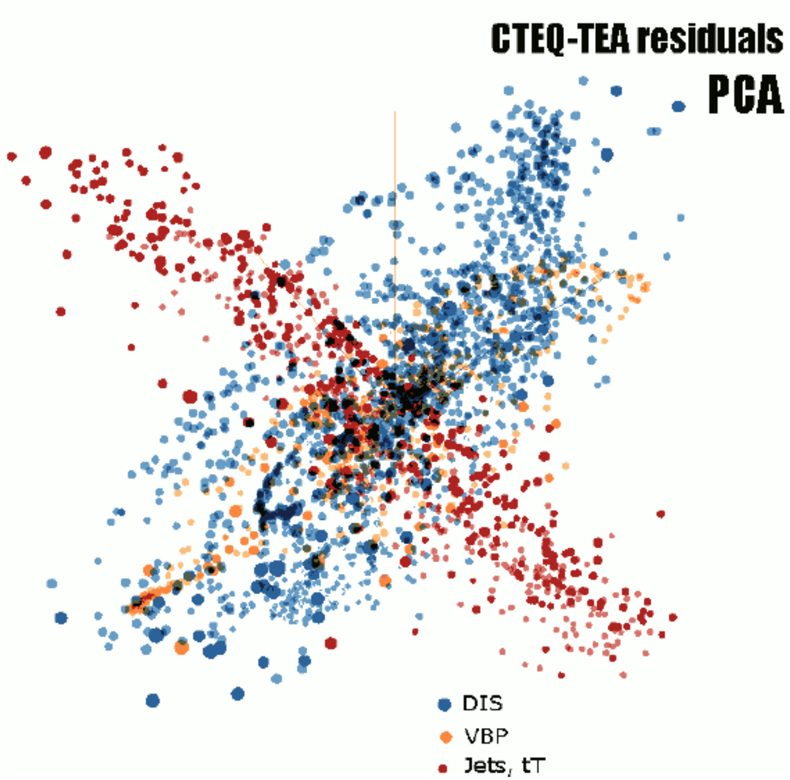


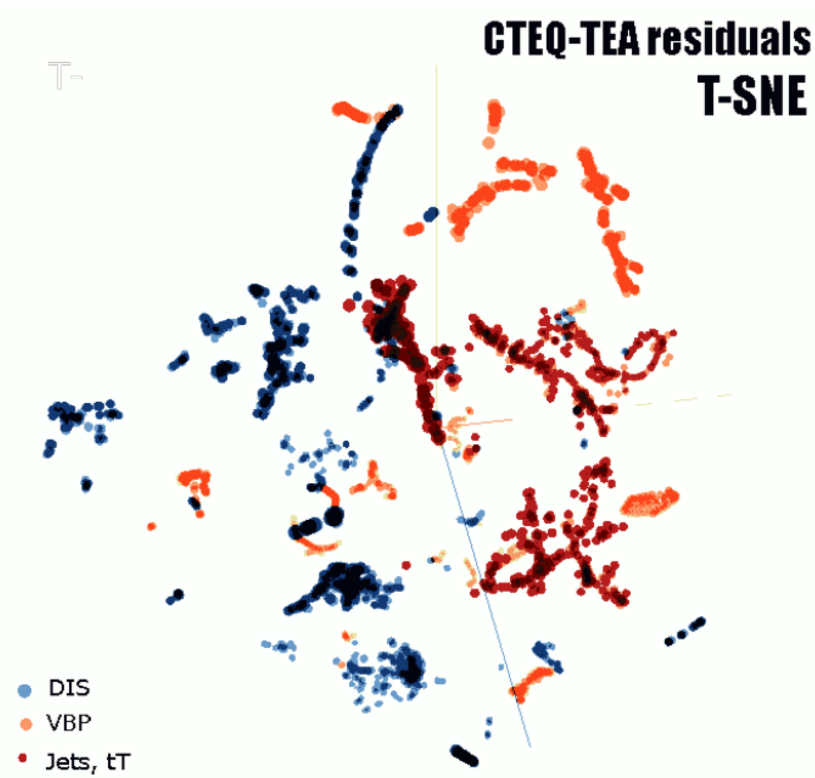
PDFSense can export the normalized $\vec{\nabla} r_i$ vectors (specifically, $(r_{i,L} - r_{i,0})/\langle r_0 \rangle_E$ for i $= 1,4000$; $L = 1,56$) in the .TSV format to analyze the whole population of $\vec{\nabla} r_i$ using machine learning tools

TensorFlow Embedding Projector

http://projector.tensorflow.org

Reads 2 .tsv files with $\vec{\nabla} r_i / \langle r_0 \rangle_E$ vectors and metadata (descriptions of data points)

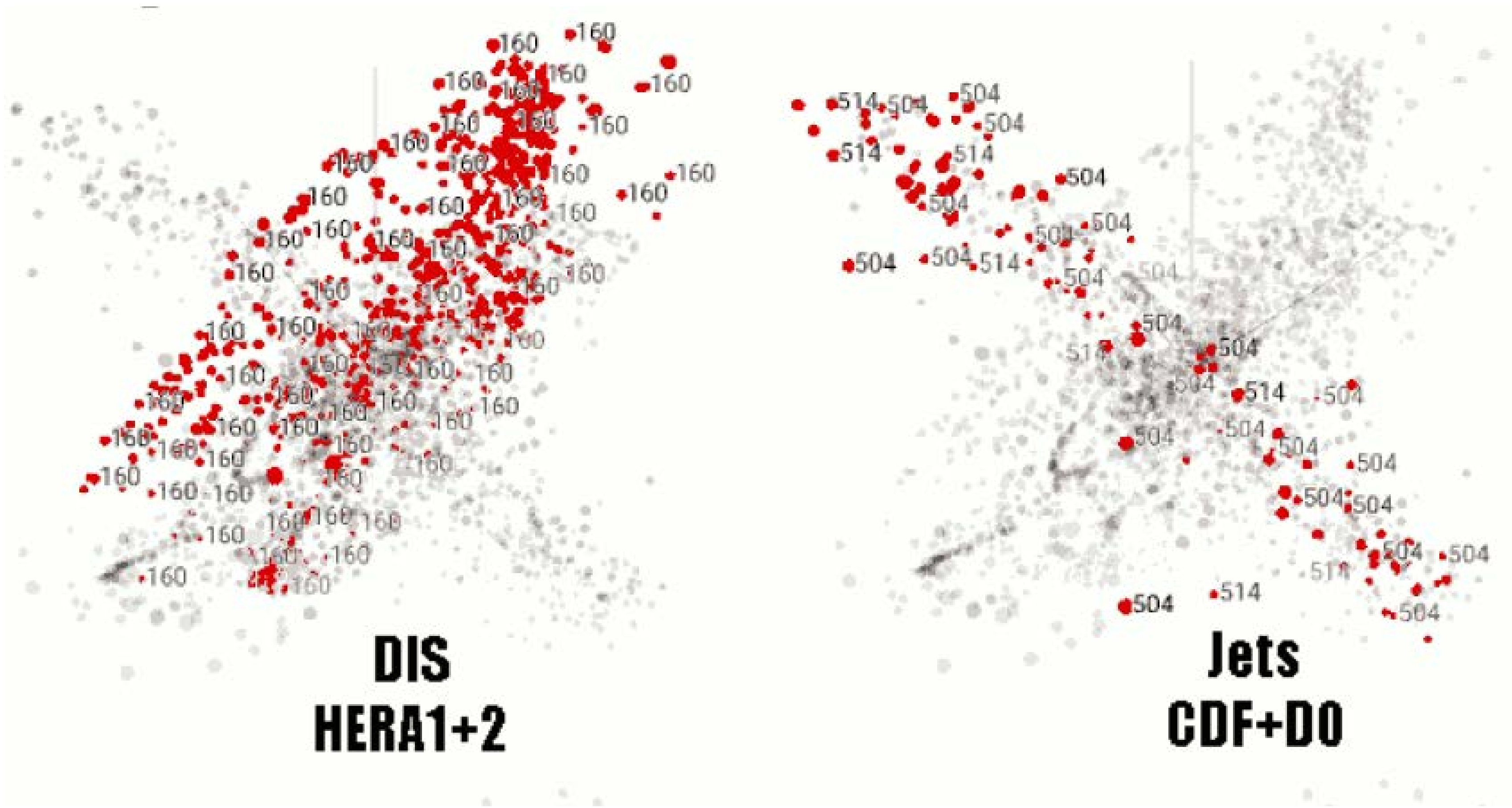




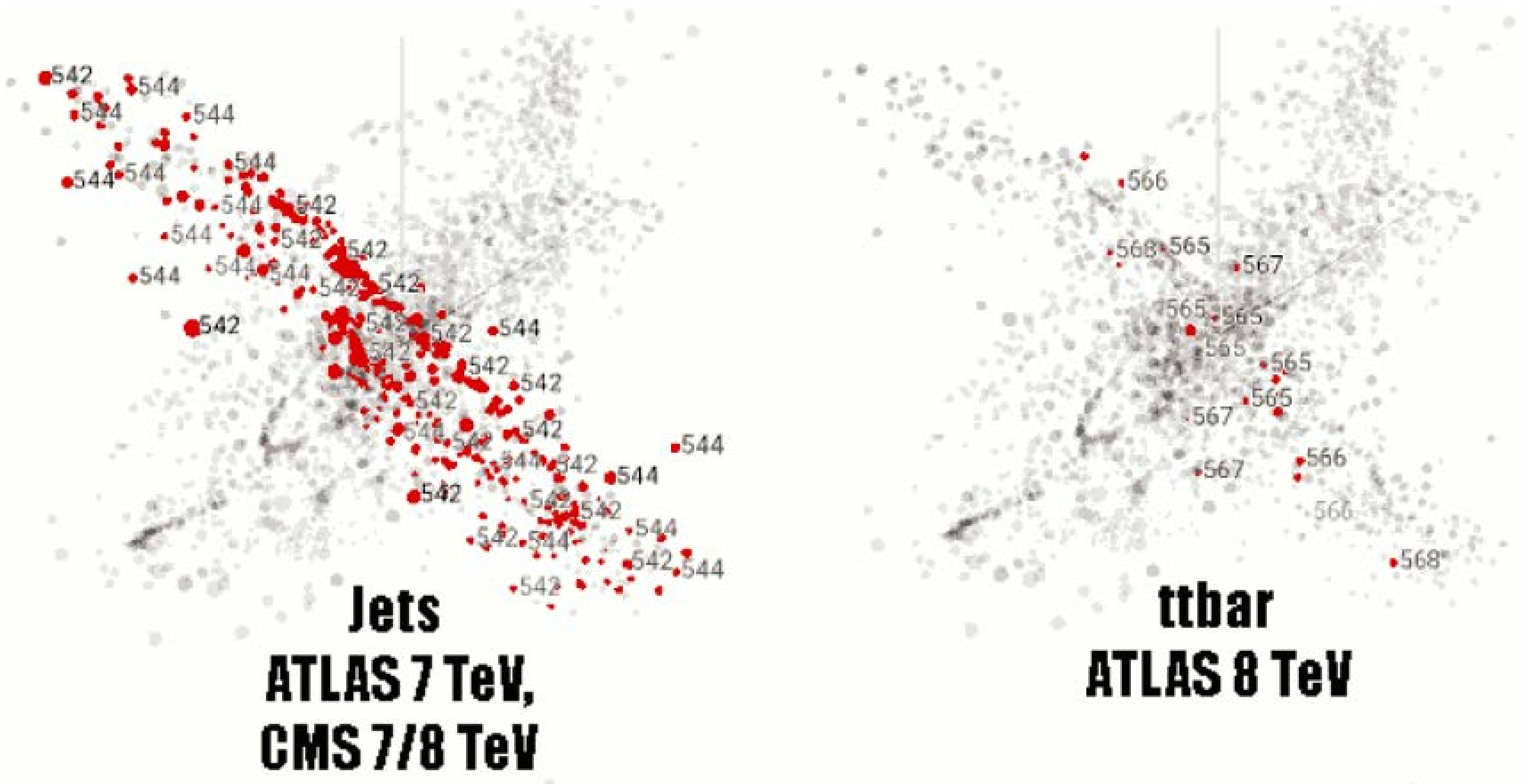
Principal Component Analysis (PCA) visualizes the 56-dim. manifold by reducing it to 10 dimensions (à la META PDFs)

t-distributed stochastic neighbor embedding (**t-SNE**) sorts $\vec{\nabla} r_i / \langle r_0 \rangle_E$ vectors according to their similarity

CTEQ-TEA residuals PCA



CTEQ-TEA residuals PCA



CTEQ-TEA residuals PCA