SoLID Collaboration Meeting
Jefferson Lab, October 14, 2017

d/u at high x and the meaning of PDF errors

Wally Melnitchouk

Jefferson Lab

Outline

- Update on d/u ratio from global QCD analysis

- Why the need for a new global analysis paradigm?
  - $\longrightarrow$ Bayesian approach to fitting
  - $\longrightarrow$ single-fit (Hessian) vs. Monte Carlo approaches

- Incompatible data sets
  - $\longrightarrow$ "tolerance" factors

- Generalization to non-Gaussian likelihoods
  - $\longrightarrow$ disjoint probabilities, empirical Bayes, …

- Outlook

d/u ratio

CJ15 global PDF analysis

- NLO analysis of expanded set of proton & deuterium data

  $\longrightarrow$ include high-x region ($x > 0.5$)

- Analysis of high-x data requires careful treatment of (subleading) power corrections

  $\longrightarrow$ target mass corrections, higher twist effects

- Correct for nuclear effects in deuteron (binding + off-shell)

  $\longrightarrow$ binding + Fermi motion (well known), nucleon off-shell (less well known)

  $\longrightarrow$ impact on d/u ratio in large-x region

"CJ15" — Accardi, WM, Owens (2016)

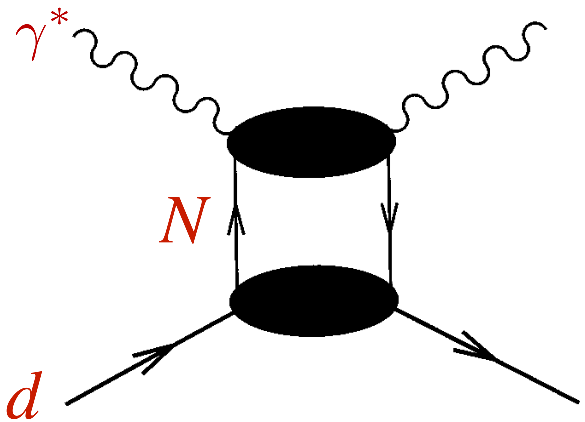
- data sets used in fit

Observable	Experiment	# points	χ^2					
			LO	NLO	NLO (OCS)	NLO (no nucl)	NLO (no nucl/D0)	
DIS F_2	BCDMS (p) [81]	351	430	**438**	436	440	427	
	BCDMS (d) [81]	254	297	**292**	289	301	301	
	SLAC (p) [82]	564	488	**434**	435	441	440	
	SLAC (d) [82]	582	396	**376**	380	507	466	
	NMC (p) [83]	275	431	**405**	404	405	403	
	NMC (d/p) [84]	189	179	**172**	173	174	173	
	HERMES (p) [86]	37	56	**42**	43	44	44	
	HERMES (d) [86]	37	51	**37**	38	36	37	
	Jefferson Lab (p) [87]	136	166	**166**	167	177	166	
	Jefferson Lab (d) [87]	136	131	**123**	124	126	130	
DIS F_2 tagged	Jefferson Lab (n/d) [21]	191	218	**214**	213	219	219	⟵ BONuS F_2^n/F_2^d
DIS σ	HERA (NC e^-p) [85]	159	325	**241**	240	247	244	
	HERA (NC e^+p 1) [85]	402	966	**580**	579	588	585	
	HERA (NC e^+p 2) [85]	75	184	**94**	94	94	93	
	HERA (NC e^+p 3) [85]	259	307	**249**	249	248	248	
	HERA (NC e^+p 4) [85]	209	348	**228**	228	228	228	
	HERA (CC e^-p) [85]	42	44	**48**	48	45	49	
	HERA (CC e^+p) [85]	39	56	**50**	50	51	51	
Drell-Yan	E866 (pp) [29]	121	148	**139**	139	145	143	
	E866 (pd) [29]	129	207	**145**	143	158	157	
W/charge asymmetry	CDF (e) [88]	11	11	**12**	12	13	14	
	DØ (μ) [17]	10	37	**20**	19	29	28	⟵ D0 A_l
	DØ (e) [18]	13	20	**29**	29	14	14	
	CDF (W) [89]	13	16	**16**	16	14	14	
	DØ (W) [19]	14	39	**14**	15	82	—	⟵ D0 A_W
Z rapidity	CDF (Z) [90]	28	100	**27**	27	26	26	
	DØ (Z) [91]	28	25	**16**	16	16	16	
jet	CDF (run 2) [92]	72	33	**15**	15	23	25	
	DØ (run 2) [93]	110	23	**21**	21	14	14	
γ+jet	DØ 1 [94]	16	17	**7**	7	7	7	
	DØ 2 [94]	16	34	**16**	16	17	17	
	DØ 3 [94]	12	34	**25**	25	24	25	
	DØ 4 [94]	12	76	**13**	13	13	13	
total		4542	5894	**4700**	4702	4964	4817	
total + norm			6022	**4708**	4710	4972	4826	
χ^2/datum			1.33	**1.04**	1.04	1.09	1.07	

~ 4500 data points, with χ^2 per datum = 1.04

Nuclear corrections

- Nuclear structure function at $x \gg 0$ dominated by incoherent scattering from individual nucleons



$$F_2^d(x,Q^2) = \int_x dy\ f(y,\gamma)\ F_2^N(x/y,Q^2)$$

$$+\ \delta^{(\mathrm{off})} F_2^d$$

nucleon momentum distribution in d (“smearing function”)

off-shell correction

$\longrightarrow$ y = momentum fraction of d carried by N

$\longrightarrow$ at finite Q^2, smearing function depends also on parameter

$$\gamma = |\mathbf{q}|/q_0 = \sqrt{1 + 4M^2x^2/Q^2}$$

Nuclear corrections

- Nucleon off-shell correction to quark PDF

$$\widetilde{q}(x,p^2) = q(x)\left[1 + \frac{(p^2 - M^2)}{M^2}\,\delta q(x)\right]$$

$$\delta q(x) = \left.\frac{\partial \log \widetilde{q}}{\partial \log p^2}\right|_{p^2=M^2}$$

$\longrightarrow$ quark "spectator" off-shell model

$$\widetilde{q}(x,p^2) = \int d\hat{p}^2\, \Phi_q(\hat{p}^2, \Lambda(p^2))$$

applied to $q_v,\ \bar{q}\ \&\ g$

momentum distribution of quarks with virtuality $\hat{p}^2$ in bound nucleon

$\longrightarrow$ scale parameter $\Lambda(p^2)$ suppresses large-p^2 contributions

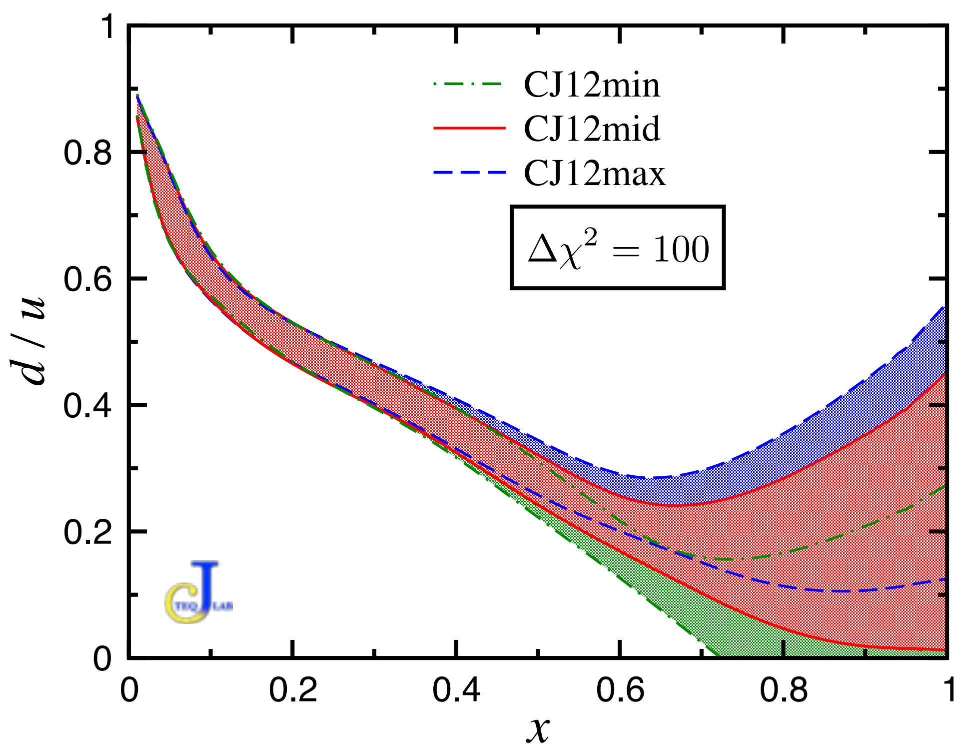
$\longrightarrow$ off-shell "rescaling" parameter $\lambda = \dfrac{\partial \log \Lambda^2}{\partial p^2}$ varied in fit to minimize χ^2

Kulagin, Petti (2006)
Owens, Accardi, WM (2013) — "CJ12"

Nuclear corrections

- In CJ12, considered 3 sets of PDFs corresponding to different amounts of nuclear corrections
  - CJ12min: WJC-1 + mild off-shell $(\lambda = 0.3\%)$
  - CJ12mid: AV18 + medium off-shell $(\lambda = 1.2\%)$
  - CJ12max: CD-Bonn + large off-shell $(\lambda = 2.1\%)$

off-shell parameter λ range motivated by Q^2 rescaling model of nuclear EMC effect
Close, Jaffe, Roberts, Ross (1988)



Owens, Accardi, WM (2013)

- with same functional form for $u\ \&\ d \sim x^{\alpha}(1-x)^{\beta}\,(1+\epsilon\sqrt{x}+\eta\,x)$ most PDF fits obtain either 0 or ∞ for $x \to 1$ limit

- more flexible parametrization for $x \to 1$ behavior

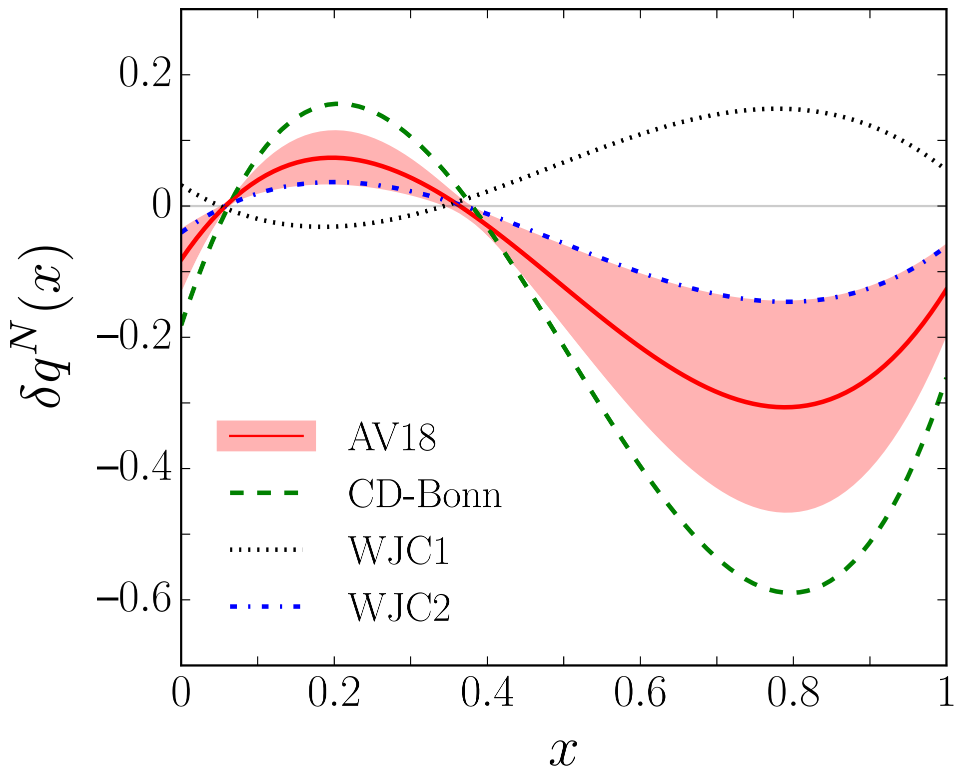
$$d \;\to\; d + a\,x^{b}\,u$$

allows finite, nonzero $x = 1$ limit

Nuclear corrections

- In CJ15, off-shell correction parametrized phenomenologically

$$\delta q^N = C_N(x - x_0)(x - x_1)(1 + x - x_0)$$

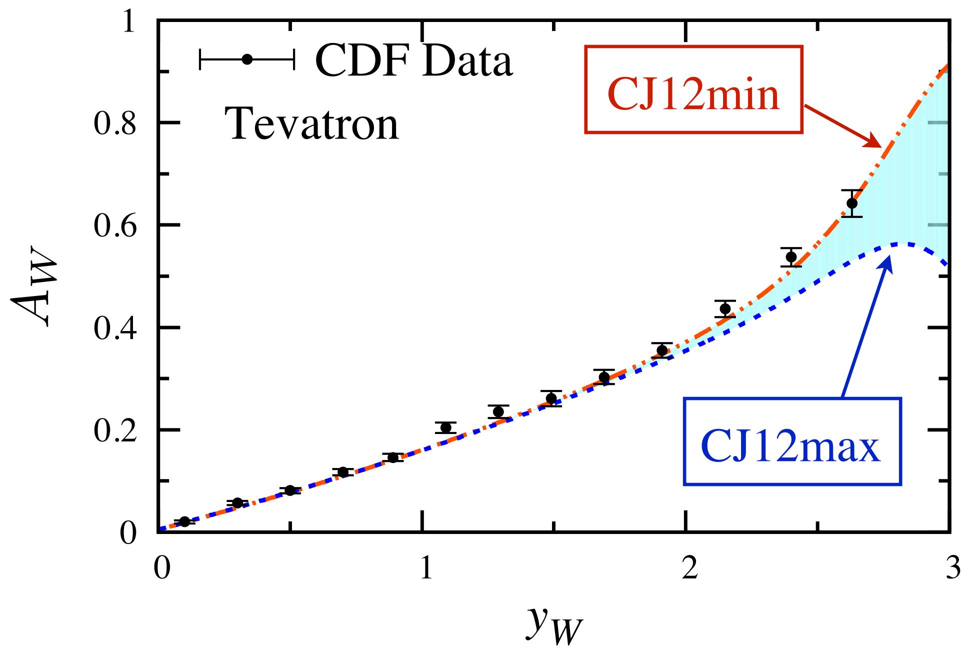


→ fitted off-shell corrections weakly dependent on deuteron wave function, except for WJC-1 (hardest momentum distribution – largest tail)

Constraints from W asymmetry

- $W^{\pm}$ asymmetry at large W-boson rapidity y_W is sensitive to d/u PDF ratio at high x

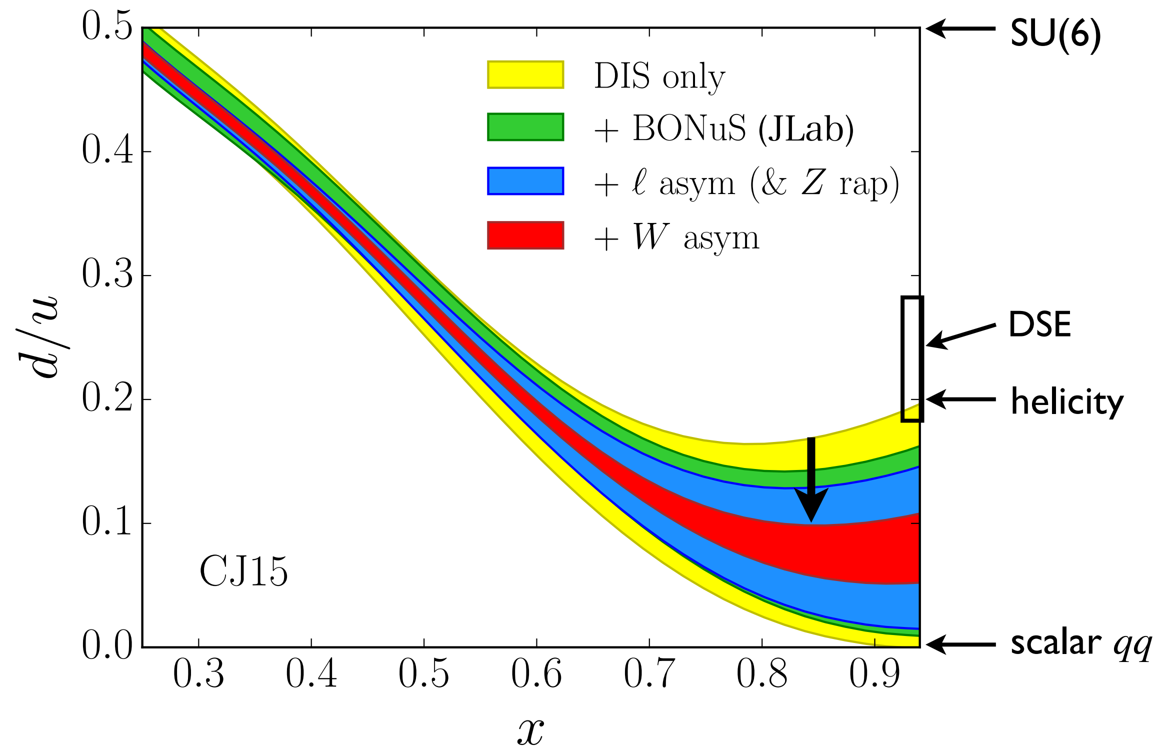
- Earlier CDF W-asymmetry data indicated preference for smaller nucleon off-shell corrections

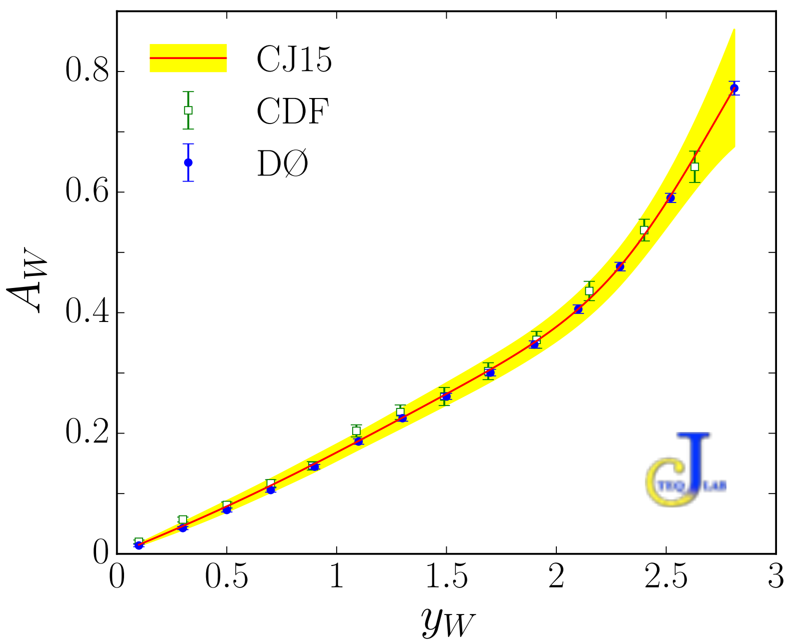


$\longrightarrow$ model dependence of W asymmetry extraction from measured lepton asymmetry??

Impact on d/u ratio

- Significant reduction of PDF errors with DØ W-asymmetry & BONuS data





→ extrapolated ratio at $x = 1$

$$d/u \rightarrow 0.09 \pm 0.03$$

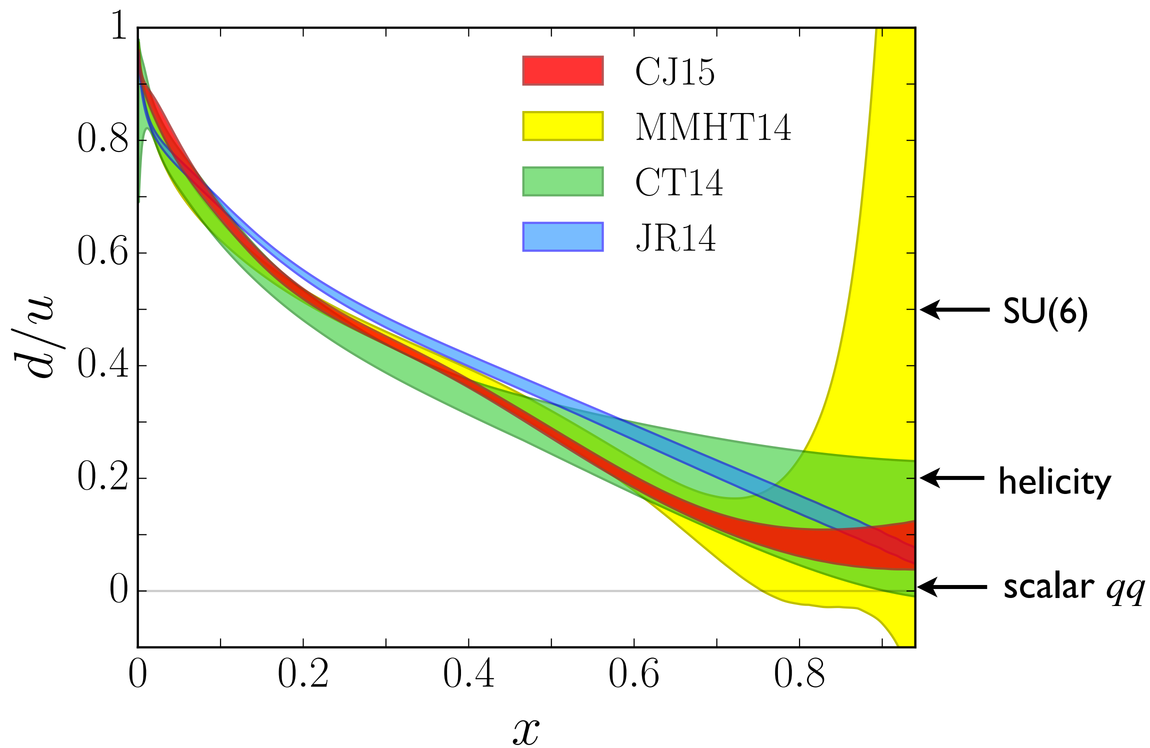
→ **note**: errors are 90% CL

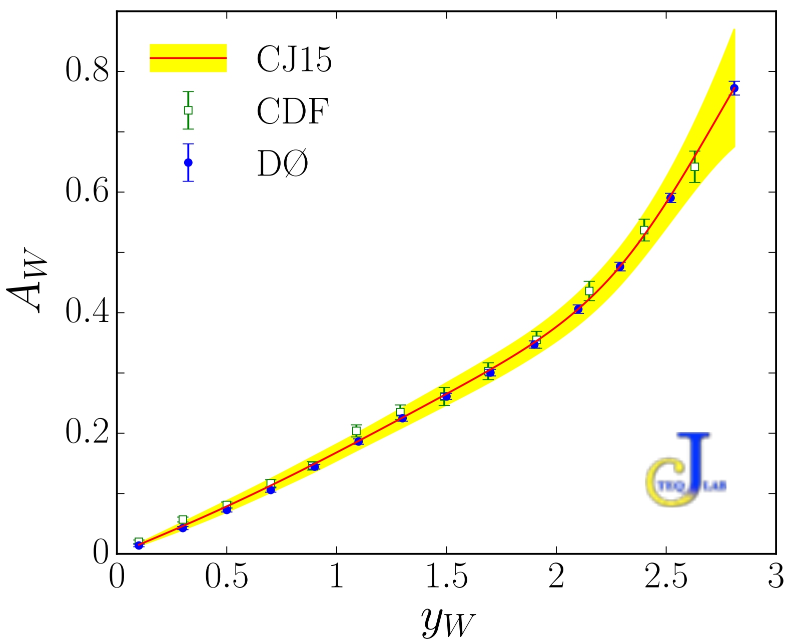
$$\Delta\chi^2 = 2.7$$

Impact on d/u ratio

- Different groups use different definitions of PDF uncertainties

  → cannot compare directly…





→ CJ15: $\Delta\chi^2 = 2.7$

→ MMHT: $\Delta\chi^2 \approx 25 - 100$

→ CT14: $\Delta\chi^2 \approx 100$

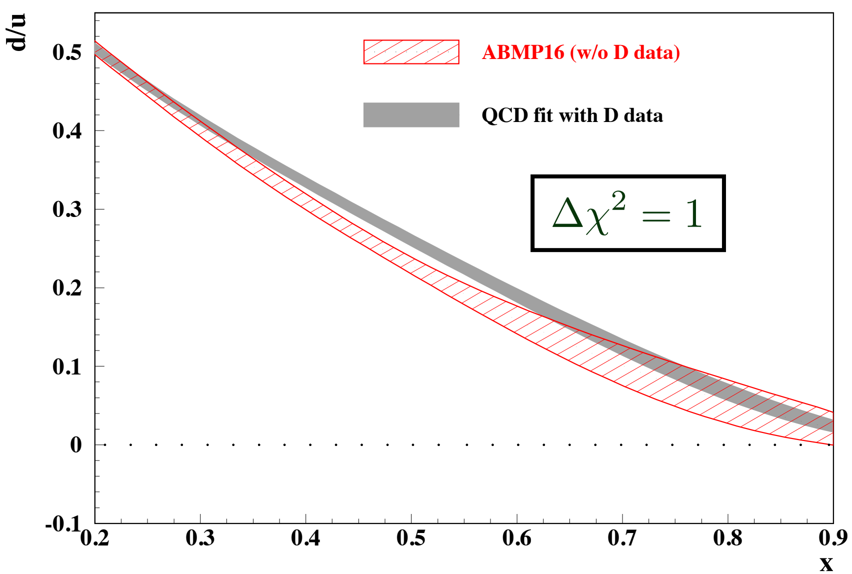
→ JR14: $\Delta\chi^2 = 1$

… is this a meaningful comparison?

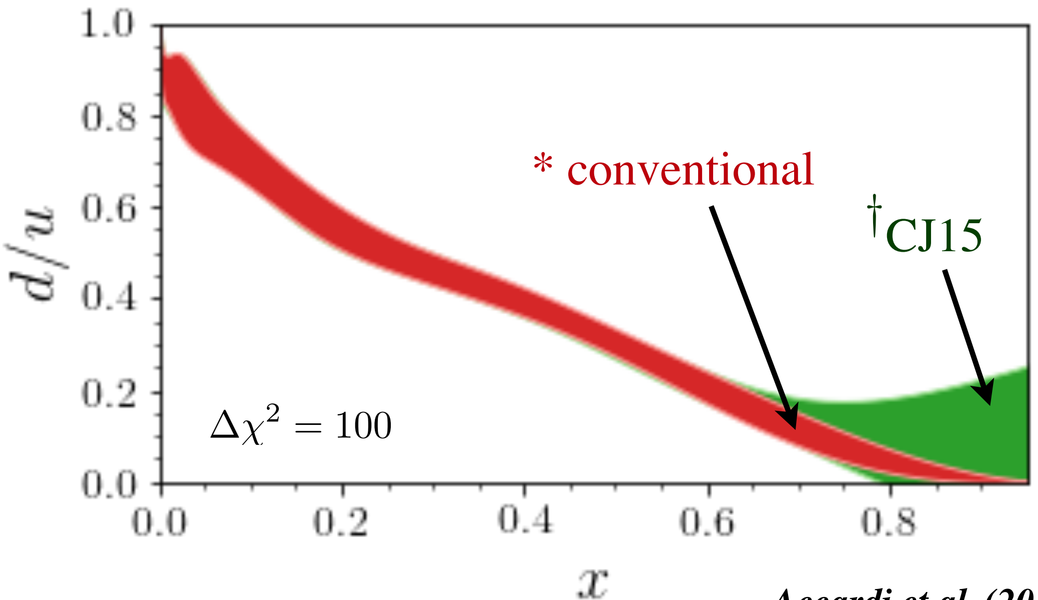
Impact on d/u ratio

- Dependence on PDF parametrization

  → recent analysis by AKP has tiny uncertainties, and $d/u \to 0$, which we (CJ) believe is simply parametrization bias!



Alekhin, Kulagin, Petti (2017)



Accardi et al. (2017)

* same functional form for u & d $\sim (1-x)^\beta$

† more flexible form $d \;\to\; d + a\,x^b\,u$

… <u>is there a more robust analysis</u>?

Need for new technology

- A major challenge has been to characterize PDF uncertainties — in a statistically meaningful way — in the presence of *tensions* among data sets

- Previous attempts sought to address tensions in data sets by introducing

  $\longrightarrow$ "tolerance" factors (artificially inflating PDF errors)

  $\longrightarrow$ "neural net" parametrization (instead of polynomial parametrization), together with MC techniques

- However, to address the problem in a more statistically rigorous way, one requires going *beyond* the standard χ^2 minimization paradigm

  $\longrightarrow$ utilize modern techniques based on Bayesian statistics!

Need for new technology

- In the near future, standard χ^2 minimization techniques will be unsuitable — even in the absence of tensions — *e.g.* for

  $\longrightarrow$ simultaneous analysis of collinear distributions (unpolarized & polarized PDFs, fragmentation functions)

  $\longrightarrow$ new types of observables — TMDs or GPDs — that will involve $> \mathcal{O}(10^5)$ data points, with $\mathcal{O}(10^3)$ parameters

Need for new technology

- Typically PDF parametrizations are nonlinear functions of the PDF parameters, $e.g.$

$$xf(x,\mu) = Nx^{\alpha}(1-x)^{\beta}\,P(x)$$

where P is a polynomial $e.g.$ $P(x) = 1 + \epsilon\sqrt{x} + \eta\, x$,
or Chebyshev, neural net, ...

$\longrightarrow$ have multiple local minima present in the χ^2 function

- Robust parameter estimation that thoroughly scans over a realistic parameter space, including multiple local minima, is only possible using MC methods

Bayesian approach to fitting

with Nobuo Sato

Connecticut / JLab

Bayesian approach to fitting

- Analysis of data requires estimating expectation values E and variances V of "observables" $\mathcal{O}$ (= PDFs, FFs) which are functions of parameters $\vec{a}$

$$\boxed{\begin{aligned} E[\mathcal{O}] &= \int d^n a\, \mathcal{P}(\vec{a}|\mathrm{data})\, \mathcal{O}(\vec{a}) \\ V[\mathcal{O}] &= \int d^n a\, \mathcal{P}(\vec{a}|\mathrm{data})\, \big[\mathcal{O}(\vec{a}) - E[\mathcal{O}]\big]^2 \end{aligned}}$$

"Bayesian master formulas"

- Using Bayes' theorem, probability distribution $\mathcal{P}$ given by

$$\mathcal{P}(\vec{a}|\mathrm{data}) = \frac{1}{Z}\,\mathcal{L}(\mathrm{data}|\vec{a})\,\pi(\vec{a})$$

in terms of the <u>likelihood function</u> $\mathcal{L}$

Bayesian approach to fitting

- Likelihood function

$$\mathcal{L}(\mathrm{data}|\vec{a}) = \exp\left(-\frac{1}{2}\chi^2(\vec{a})\right)$$

is a Gaussian form in the data, with χ^2 function

$$\chi^2(\vec{a}) = \sum_i \left(\frac{\mathrm{data}_{\,i} - \mathrm{theory}_i(\vec{a})}{\delta(\mathrm{data})}\right)^2$$

with priors $\pi(\vec{a})$ and "evidence" Z

$$Z = \int d^n a\, \mathcal{L}(\mathrm{data}|\vec{a})\, \pi(\vec{a})$$

$\longrightarrow$ Z tests if $e.g.$ an n-parameter fit is statistically different from $(n+1)$-parameter fit

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

<u>Maximum Likelihood</u>
(χ^2 minimization)

<u>Monte Carlo</u>

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood
(χ^2 minimization)

$\longrightarrow$ maximize probability distribution $\mathcal{P}$ by minimizing χ^2 for a set of best-fit parameters $\vec{a}_0$

$$E\left[\,\vec{a}\,\right] = \vec{a}_0$$

$\longrightarrow$ if $\mathcal{O}$ is $\approx$ linear in the parameters, and if probability is symmetric in all parameters

$$E\left[\,\mathcal{O}(\vec{a})\,\right] \approx \mathcal{O}(\vec{a}_0)$$

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood

(χ^2 minimization)

$\longrightarrow$ variance computed by expanding $\mathcal{O}(\vec{a})$ about $\vec{a}_0$

$e.g.$ in 1 dimension have "master formula"

$$V[\mathcal{O}] \approx \frac{1}{4}\Big[\mathcal{O}(a+\delta a) - \mathcal{O}(a-\delta a)\Big]^2$$

where

$$\delta a^2 = V[a]$$

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

<u>Maximum Likelihood</u>
(χ^2 minimization)

$\longrightarrow$ generalization to multiple dimensions via Hessian approach:

find set of (orthogonal) contours in parameter space around $\vec{a}_0$ such that $\mathcal{L}$ along each contour is parametrized by statistically independent parameters — directions of contours given by eigenvectors $\hat{e}_k$ of Hessian matrix H, with elements

$$H_{ij} = \frac{1}{2} \left. \frac{\partial^2 \chi^2(\vec{a})}{\partial a_i \partial a_j} \right|_{\vec{a}=\vec{a}_0}$$

and contours parametrized as $\Delta a^{(k)} = a^{(k)} - a_0 = t_k \dfrac{\hat{e}_k}{\sqrt{v_k}}$,
with v_k eigenvectors of H

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

<u>Maximum Likelihood</u>
(χ^2 minimization)

$\longrightarrow$ basic assumption: $\mathcal{P}$ factorizes along each eigendirection

$$\mathcal{P}(\Delta a) \approx \prod_k \mathcal{P}_k(t_k)$$

where

$$\mathcal{P}_k(t_k) = \mathcal{N}_k \exp\left[-\frac{1}{2}\chi^2\left(a_0 + t_k \frac{\hat{e}_k}{\sqrt{v_k}}\right)\right]$$

<u>note</u>: in quadratic approximation for χ^2, this becomes a normal distribution

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood

(χ^2 minimization)

$\longrightarrow$ uncertainties on $\mathcal{O}$ along each eigendirection
(assuming linear approximation)

$$(\Delta\mathcal{O}_k)^2 \approx \frac{1}{4}\left[\mathcal{O}\left(a_0 + T_k\frac{\hat{e}_k}{\sqrt{v_k}}\right) - \mathcal{O}\left(a_0 - T_k\frac{\hat{e}_k}{\sqrt{v_k}}\right)\right]^2$$

where T_k is finite step size in t_k , with total variance

$$V[\mathcal{O}] = \sum_k (\Delta\mathcal{O}_k)^2$$

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Monte Carlo

→ in practice, generally one has $E[\mathcal{O}(\vec{a})] \neq \mathcal{O}(E[\vec{a}])$
so the maximal likelihood method will sometimes fail

→ Monte Carlo approach samples parameter space and assigns weights w_k to each set of parameters a_k

→ expectation value and variance are then weighted averages

$$E[\mathcal{O}(\vec{a})] = \sum_k w_k\, \mathcal{O}(\vec{a}_k), \qquad V[\mathcal{O}(\vec{a})] = \sum_k w_k \left(\mathcal{O}(\vec{a}_k) - E[\mathcal{O}]\right)^2$$

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood
(χ^2 minimization)

- fast
- assumes Gaussianity
- no guarantee that global minimum has been found
- errors only characterize local geometry of χ^2 function

Monte Carlo

- slow
- does not rely on Gaussian assumptions
- includes all possible solutions
- accurate

Incompatible data sets

Incompatible data sets

- Incompatible data sets can arise because of errors in determining central values, or underestimation of systematic experimental uncertainties

  $\longrightarrow$ requires some sort of modification to standard statistics

- Often one modifies the master formula by introducing a "tolerance" factor T

$$V[\mathcal{O}] \;\rightarrow\; T^2 \, V[\mathcal{O}]$$

$e.g.$ for one dimension

$$V[\mathcal{O}] = \frac{T^2}{4}\Big[\mathcal{O}(a+\delta a) - \mathcal{O}(a-\delta a)\Big]^2$$

$\longrightarrow$ effectively modifies the likelihood function

Incompatible data sets

- Simple example: consider observable m, and two measurements

$$(m_1, \delta m_1), \quad (m_2, \delta m_2)$$

$\longrightarrow$ compute exactly the χ^2 function

$$\chi^2 = \left(\frac{m - m_1}{\delta m_1}\right)^2 + \left(\frac{m - m_2}{\delta m_2}\right)^2$$

and, from Bayesian master formula, the mean value

$$E[m] = \frac{m_1 \delta m_2^2 + m_2 \delta m_1^2}{\delta m_1^2 + \delta m_2^2}$$

and variance

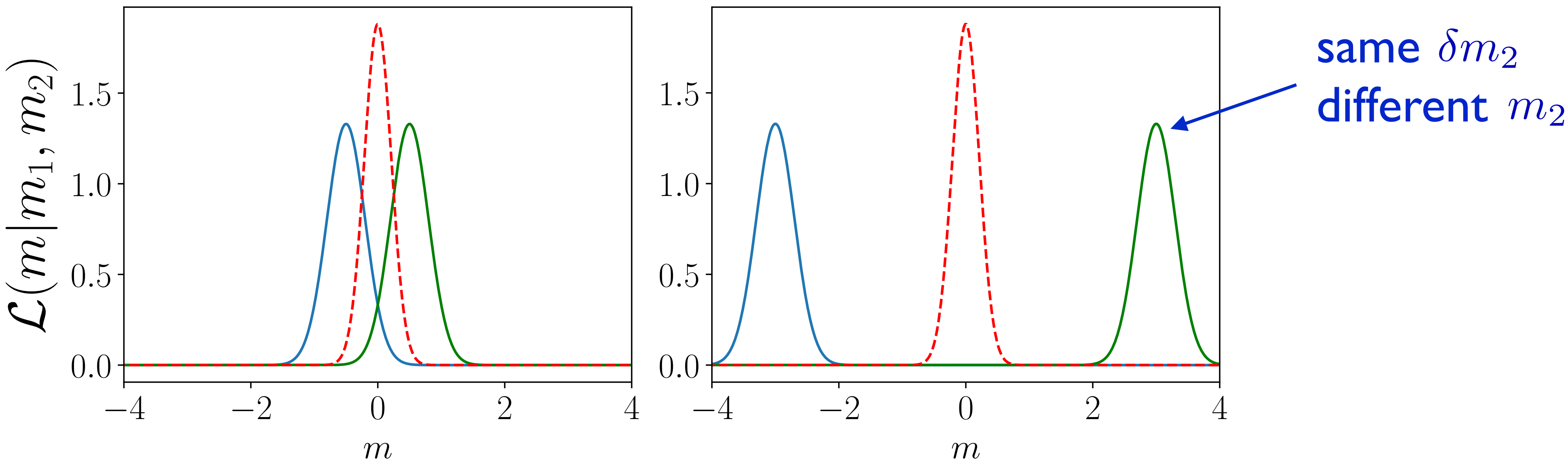
$$V[m] = H^{-1} = \frac{\delta m_1^2 \, \delta m_2^2}{\delta m_1^2 + \delta m_2^2}$$

does not depend on $m_1 - m_2$!

Incompatible data sets

- <u>Simple example:</u> consider observable m, and two measurements

$$(m_1, \delta m_1), \quad (m_2, \delta m_2)$$

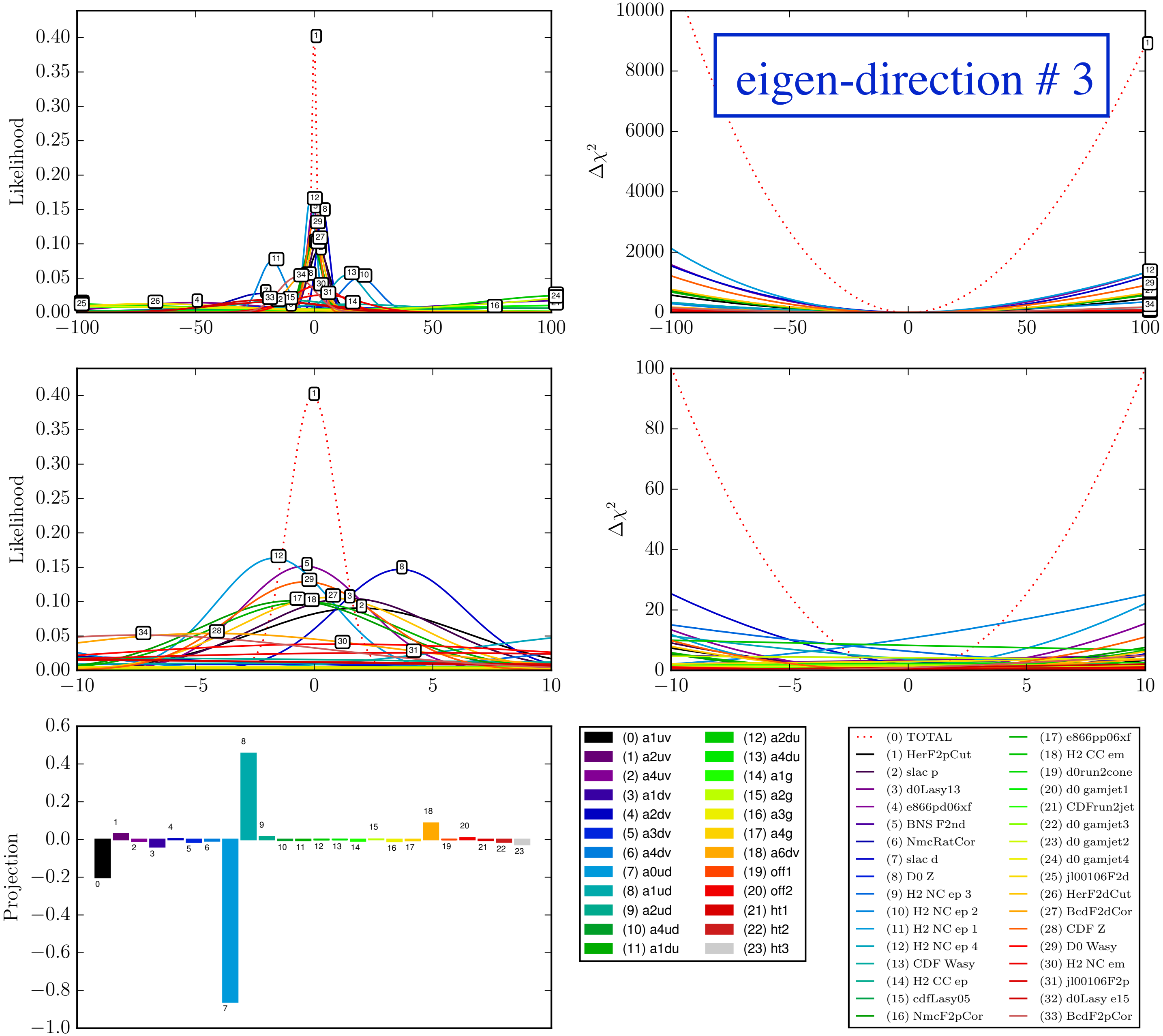


⟶ total uncertainty remains independent of degree of (in)compatibility of data

⟶ Gaussian likelihood gives unrealistic representation of true uncertainty

Incompatible data sets

- <u>Realistic example:</u> recent CJ (CTEQ-JLab) global PDF analysis

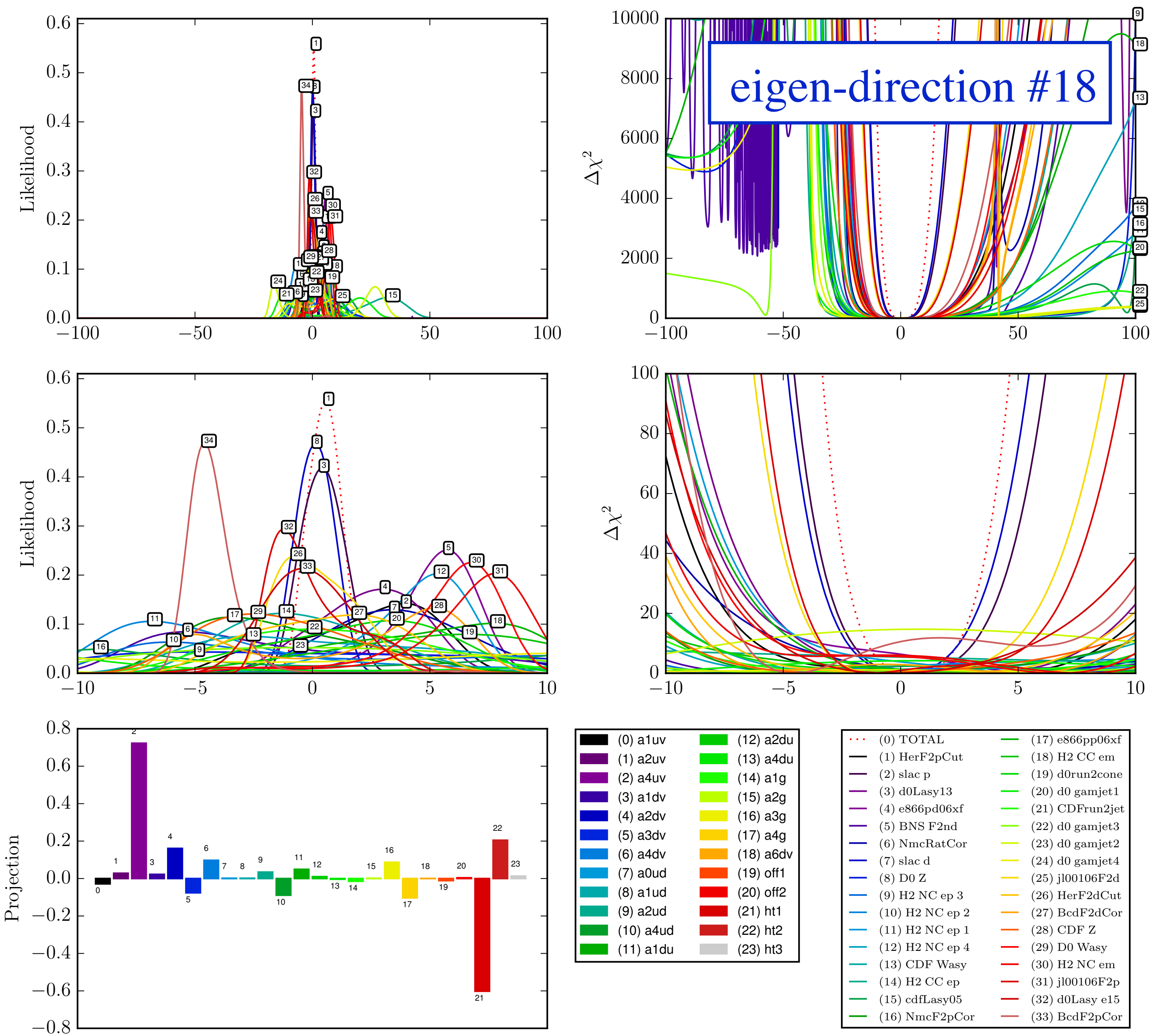


→ 24 parameters, 33 data sets

→ data sets compatible along this e-direction

Incompatible data sets

- <u>Realistic example:</u> recent CJ (CTEQ-JLab) global PDF analysis

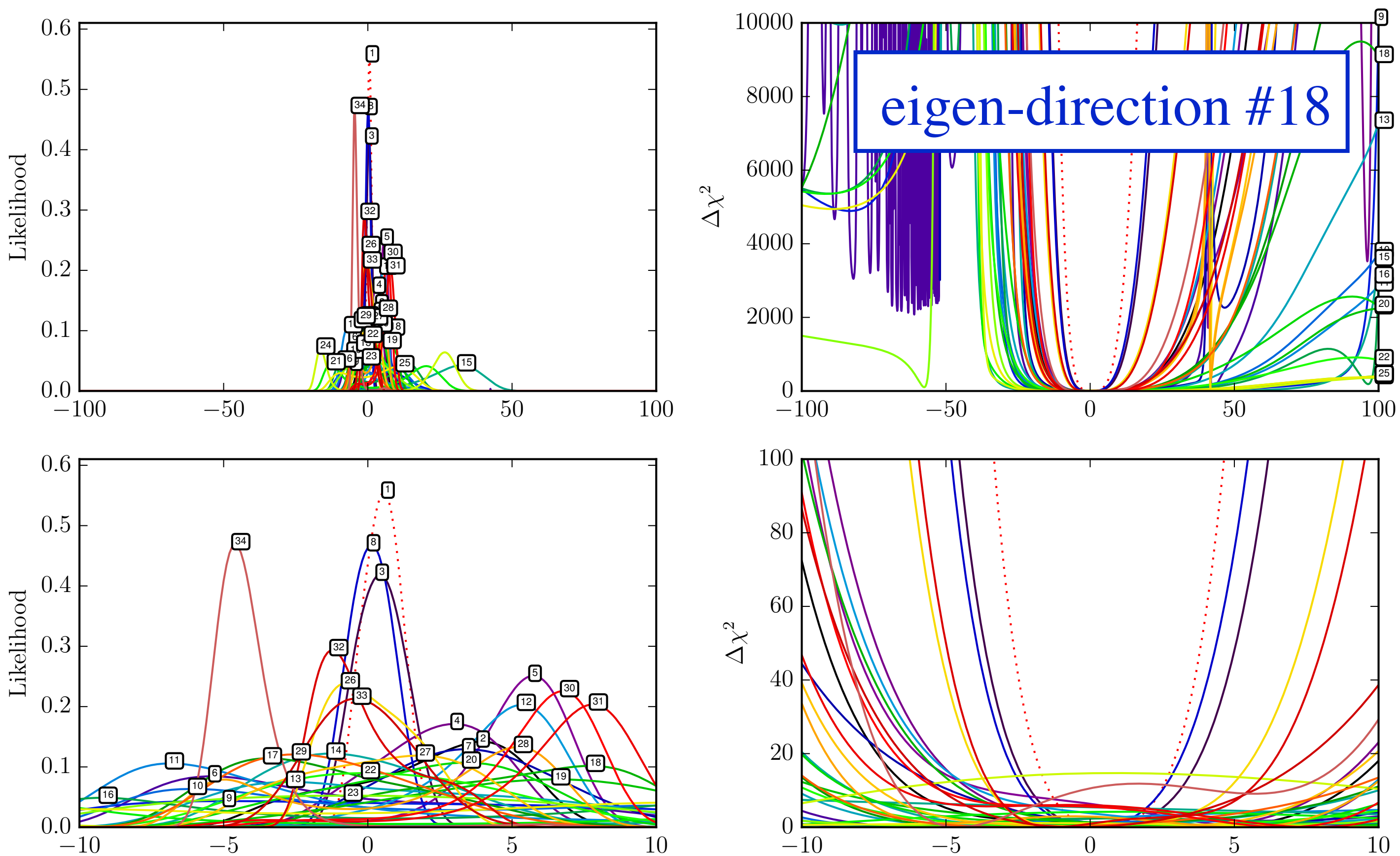


→ 24 parameters, 33 data sets

→ data sets <u>not</u> compatible along this e-direction

Incompatible data sets

- Realistic example: recent CJ (CTEQ-JLab) global PDF analysis



→ 24 parameters, 33 data sets

→ data sets not compatible along this e-direction

→ standard Gaussian likelihood incapable of accounting for underestimated individual errors (leading to incompatible data sets) — not designed for such scenarios!

Incompatible data sets

- Two ways in which tolerance factors usually implemented

$\longrightarrow$ CTEQ "tolerance criteria"
(variations adopted by other groups, *e.g.*, MMHT, CJ)

Pumplin, Stump, Huston, Lai, Nadolsky, Tung (2012)

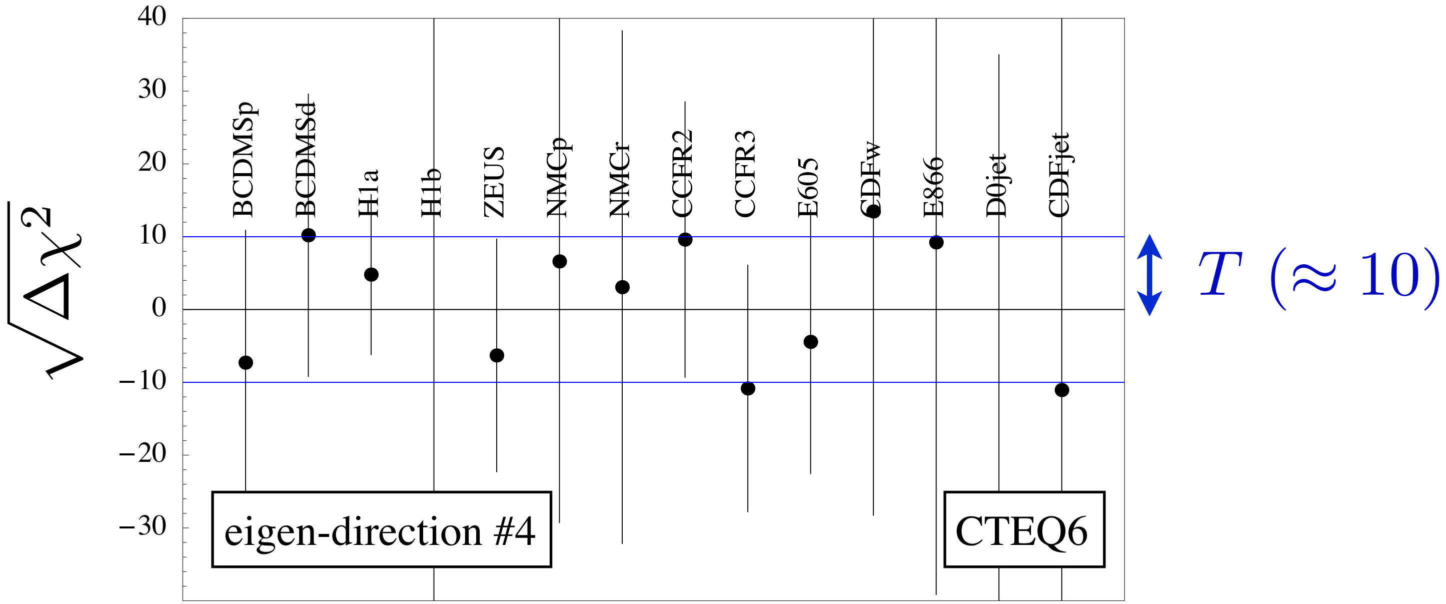
$\longrightarrow$ scaling of $\Delta\chi^2$ with number of parameters
(or number of degrees of freedom)

e.g. Brodsky, Gardner (2016)

JDHLM assess their PDF errors using a tolerance criteria of $\Delta\chi^2 = 1$ at 1σ; however, the actual value of $\Delta\chi^2$ to be employed depends on the number of parameters to be simultaneously determined in the fit. This is illustrated in Table 38.2 of Ref. [15] and is used broadly, noting, e.g., Refs. [16–19]. Ref. [7] employs the CT10 PDF analysis [20], so that it contains 25 parameters, plus one for intrinsic charm. Figure 38.2 of Ref. [15] then shows that $\Delta\chi^2 \approx 29$ at 1σ (68% CL), whereas $\Delta\chi^2 \approx 36$ at 90% CL. Ref. [7] uses the criterion $\Delta\chi^2 > 100$, determined on empirical grounds, to indicate a poor fit. JDHLM employs the framework of Ref. [21] which contains 25 parameters for the PDFs and 12 for the higher-twist contributions, so that a much larger tolerance than $\Delta\chi^2 = 1$ is warranted.

Incompatible data sets

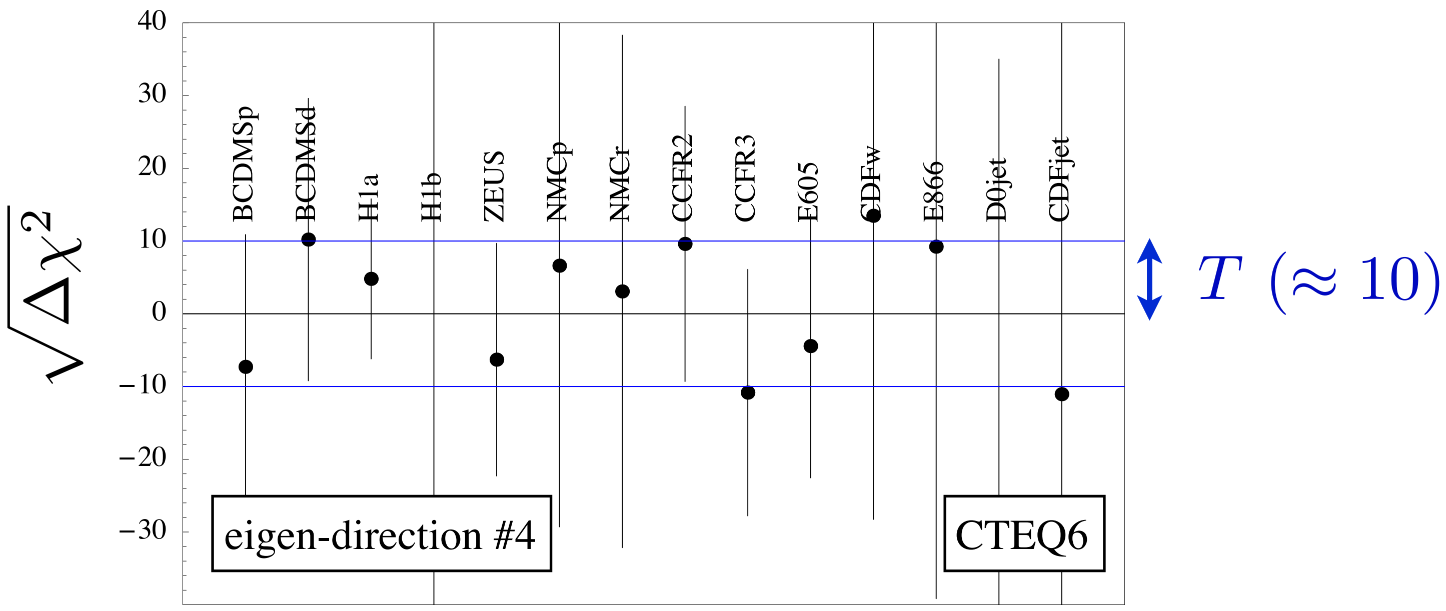
CTEQ tolerance criteria



- for each experiment, find minimum χ^2 along given e-direction
- from χ^2 distribution determine 90% CL for each experiment
- along each side of e-direction, determine maximum range $d_k^\pm$ allowed by the most constraining experiment
- T computed by averaging over all $d_k^\pm$ (typically $T \sim 5 - 10$)

Incompatible data sets

- **CTEQ tolerance criteria**



- This approach is *not consistent* with Gaussian likelihood

$\longrightarrow$ no clear Bayesian interpretation of uncertainties (ultimately, a prescription...)

Incompatible data sets

- Scaling of $\Delta\chi^2$ with # of parameters: "$\Delta\chi^2$ paradox"

- Simple example: two parameters $\theta_i \quad (i = 1, 2)$
  with mean values μ_i and standard deviation σ_i

$\longrightarrow$ joint probability distribution

$$\mathcal{P}(\theta_1, \theta_2) = \prod_{i=1,2} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2}\left(\frac{\theta_i - \mu_i}{\sigma_i}\right)^2\right]$$

$\longrightarrow$ change variables $\theta_i \;\to\; t_i = (\theta_i - \mu_i)/\sigma_i$ and use polar coordinates $r^2 = t_1^2 + t_2^2, \quad \phi = \tan^{-1}(t_2/t_1)$

$$d\theta_1 d\theta_2\, \mathcal{P}(\theta_1, \theta_2) = \frac{d\phi}{2\pi}\, r dr\, \exp\left[-\frac{1}{2}r^2\right]$$

Incompatible data sets

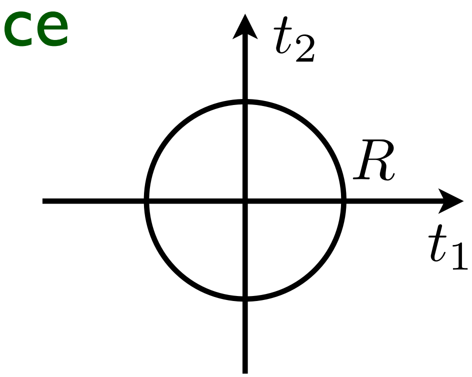
- Scaling of $\Delta\chi^2$ with # of parameters: "$\Delta\chi^2$ paradox"

$\longrightarrow$ confidence volume

$$\mathrm{CV} \equiv \int d\theta_1 d\theta_2\, \mathcal{P}(\theta_1,\theta_2) = \int_0^R dr\, r\, \exp\left[-\frac{1}{2}r^2\right]$$

$$= 68\% \text{ for } R = 2.279$$

$\longrightarrow$ note that $R^2 = t_1^2 + t_2^2 \equiv \chi^2$, so that confidence region for parameters $\max\,[t_i] = R$



$\longrightarrow$ implies that $\theta_i = \mu_i \pm \sigma_i\, R$, which contradicts original premise that $\theta_i = \mu_i \pm \sigma_i$!

Incompatible data sets

- Scaling of $\Delta\chi^2$ with # of parameters: "$\Delta\chi^2$ paradox"

$\longrightarrow$ to resolve paradox, use Bayesian master formulas

$$E[\theta_i] = \int_0^{2\pi} \frac{d\phi}{2\pi} \int_0^{\infty} dr\, \mathcal{P}(r,\phi)\, \theta_i$$

$$= \int_0^{2\pi} \frac{d\phi}{2\pi} \int_0^{\infty} dr\, r\, e^{-r^2/2}\, (\mu_i + t_i\, \sigma_i) \quad = \quad \mu_i \quad \checkmark$$

$$V[\theta_i] = \int_0^{2\pi} \frac{d\phi}{2\pi} \int_0^{\infty} dr\, \mathcal{P}(r,\phi)\, (\theta_i - \mu_i)^2$$

$$= \int_0^{2\pi} \frac{d\phi}{2\pi} \int_0^{\infty} dr\, r\, e^{-r^2/2}\, (t_i\, \sigma_i)^2 \quad = \quad \sigma_i^2 \quad \checkmark$$

Incompatible data sets

- Scaling of $\Delta\chi^2$ with # of parameters: "$\Delta\chi^2$ paradox"

  $\longrightarrow$ no paradox if use $\Delta\chi^2 = 1$ for *any number* of parameters to characterize the 1σ CL

  $\longrightarrow$ only consistent tolerance for Gaussian likelihood is $T = 1$

To summarize standard maximum likelihood method…

- Gradient search (in parameter space) depends how "good" the starting point is

  $\longrightarrow$ for ~ 30 parameters trying different starting points is impractical, if do not have some information about shape

- Common to free parameters initially, then freeze those not sensitive to data (χ^2 flat locally)

  $\longrightarrow$ introduces bias, does not guarantee that flat χ^2 globally

- Cannot guarantee solution is unique

- Error propagation characterized by quadratic χ^2 near minimum

  $\longrightarrow$ no guarantee this is quadratic globally ($e.g.$ Student t-distribution?)

- Introduction of tolerance modifies Gaussian statistics

Monte Carlo methods

Monte Carlo

- Designed to faithfully compute Bayesian master formulas

- Do not assume a single minimum, include all possible solutions (with appropriate weightings)

- Do not assume likelihood is Gaussian in parameters

- Allows likelihood analysis to be extended to address tensions among data sets via Bayesian inference

- More computationally demanding compared with Hessian method

Monte Carlo

- First group to use MC for global PDF analysis was NNPDF, using neural network to parametrize $P(x)$ in

Forte et al. (2002)

$$f(x) = N\, x^{\alpha} (1-x)^{\beta}\, P(x)$$

— α, β are fitted "preprocessing coefficients"

- Iterative Monte Carlo (IMC), developed by JAM Collaboration, variant of NNPDF, tailored to non-neutral net parametrizations

- Markov Chain MC (MCMC) / Hybid MC (HMC)
— recent "proof of principle" analysis, ideas from lattice QCD

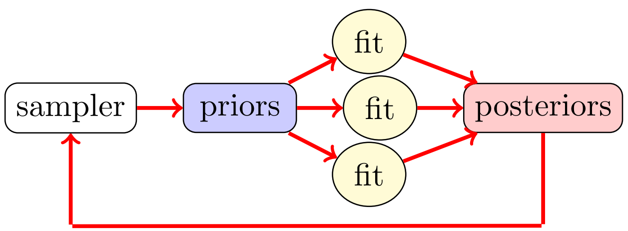
Gbedo, Mangin-Brinet (2017)

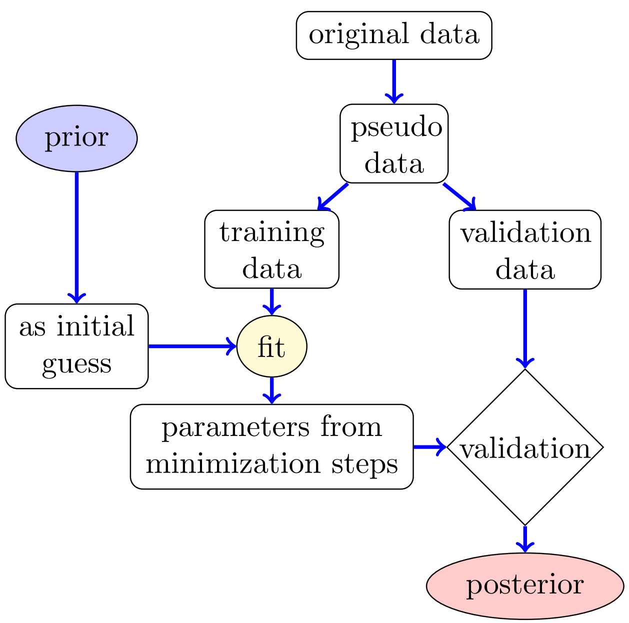
- Nested sampling (NS) — computes integrals in Bayesian master formulas (for $E,\ V,\ Z$) explicitly *Skilling (2004)*

Iterative Monte Carlo (IMC)

- Use traditional functional form for input distribution shape, but sample significantly larger parameter space than possible in single-fit analyses

Iterative Monte Carlo (IMC)





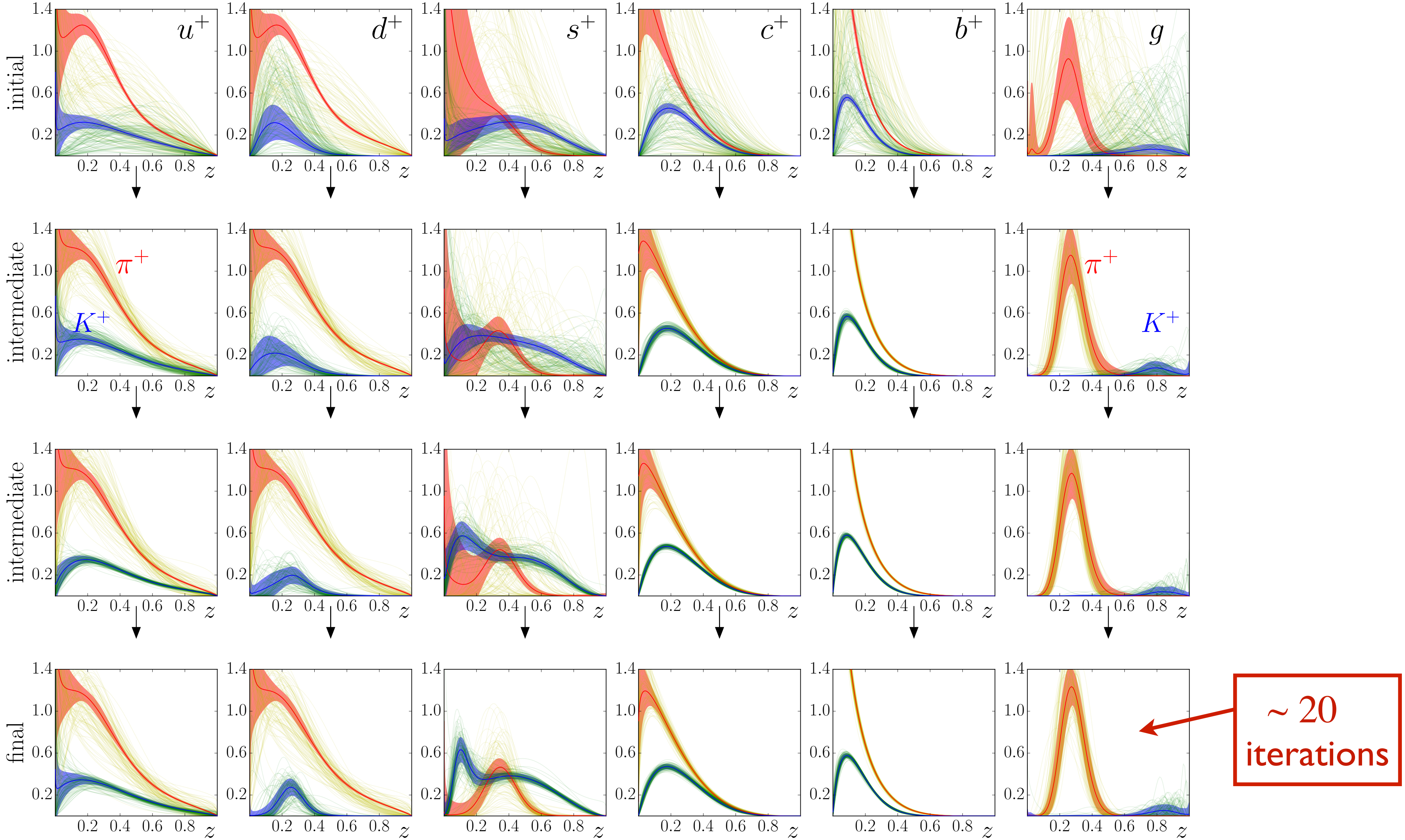
→ no assumptions for exponents

→ cross-validation to avoid overfitting

→ iterate until convergence criteria satisfied

Iterative Monte Carlo (IMC)

- *e.g.* of convergence (for fragmentation functions) in IMC



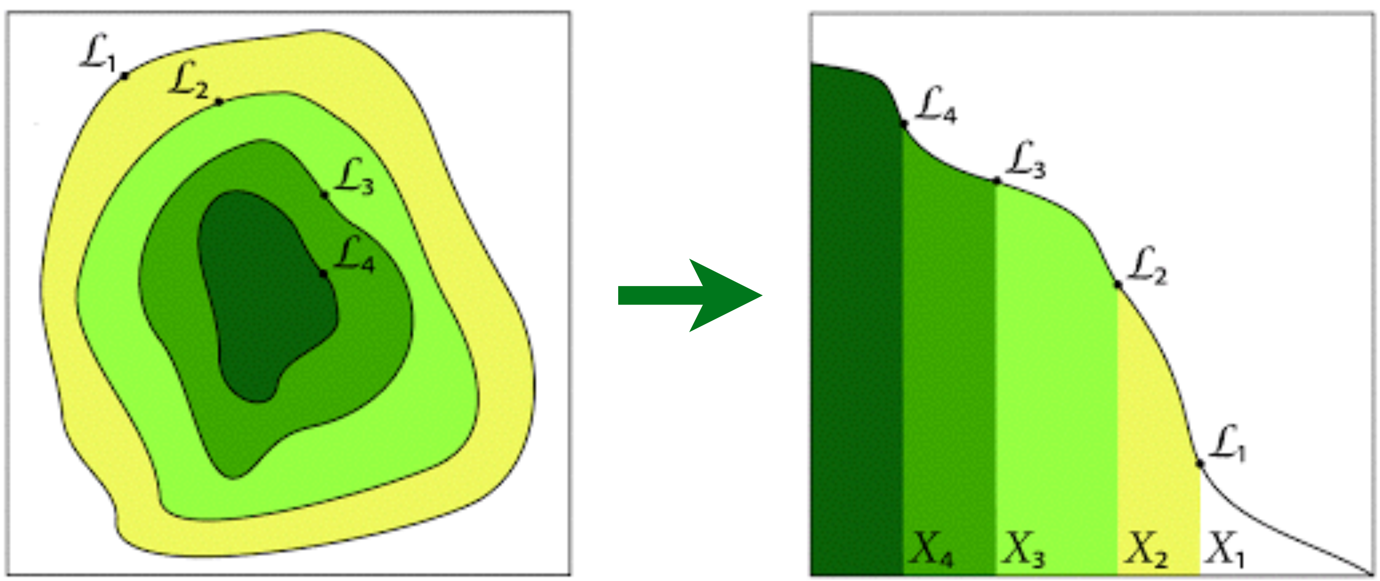
Sato et al. (2016)

Nested Sampling

- Basic idea: transform n-dimensional integral to 1-D integral

$$Z = \int d^n a \, \mathcal{L}(\text{data}|\vec{a}) \, \pi(\vec{a}) = \int_0^1 dX \, \mathcal{L}(X)$$

where *prior volume* $dX = \pi(\vec{a}) \, d^n a$



such that $0 < \cdots < X_2 < X_1 < X_0 = 1$

Feroz et al.
arXiv:1306.2144 [astro-ph]

Nested Sampling

- Approximate evidence by a weighted sum

$$Z \approx \sum_i \mathcal{L}_i \, w_i \qquad \text{with weights} \quad w_i = \frac{1}{2}(X_{i-1} - X_{i+1})$$

- Algorithm:

  → randomly select samples from full prior s.t. initial volume $X_0 = 1$

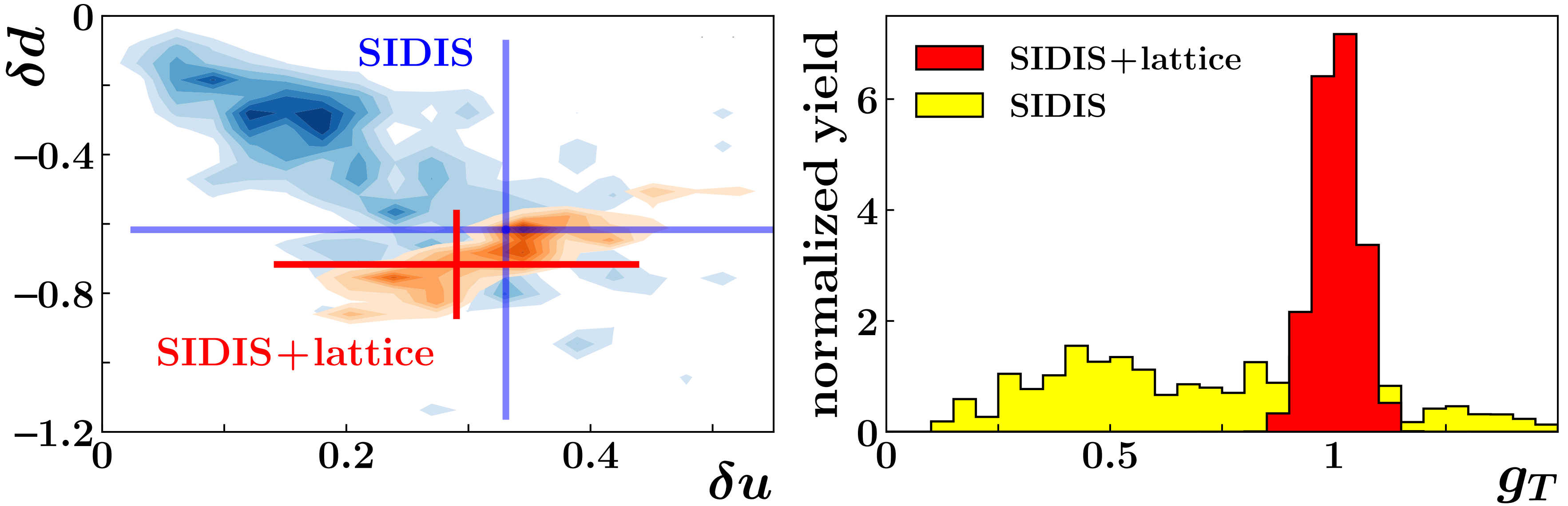
  → for each iteration, remove point with lowest $\mathcal{L}_i$, replacing it with point from prior with constraint that its $\mathcal{L} > \mathcal{L}_i$

  → repeat until entire prior volume has been traversed

  - can be parallelized
  - performs better than VEGAS for large dimensions
  - increasingly used in fields outside of (nuclear) analysis

Nested Sampling

- Recent application in global analysis of transversity TMD PDF
  (SIDIS data + lattice QCD constraint on isovector moment)



Lin, WM, Prokudin, Sato, Shows (2017)

→ distributions do not look very Gaussian!

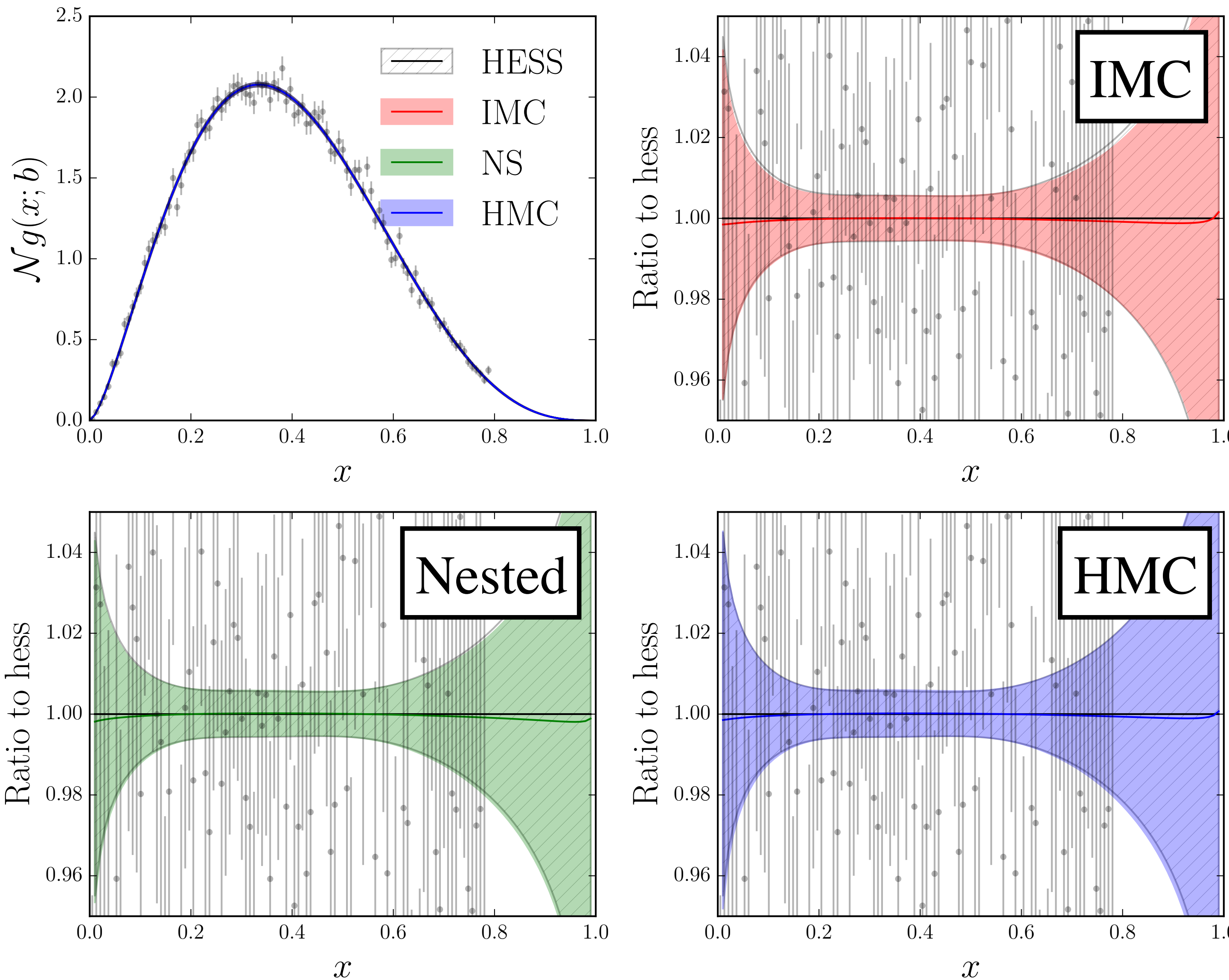
→ MC analysis gives $\delta u = 0.3 \pm 0.2,\ \delta d = -0.7 \pm 0.2$ → $g_T = 1.0 \pm 0.1$

→ maximum likelihood analysis would have given $g_T \approx 0.5$

MC Error Analysis

- Assuming a single minimum, a Hessian or MC analysis *must* give same results, if using same likelihood function

→ analysis of pseudodata, generated using Gaussian distribution

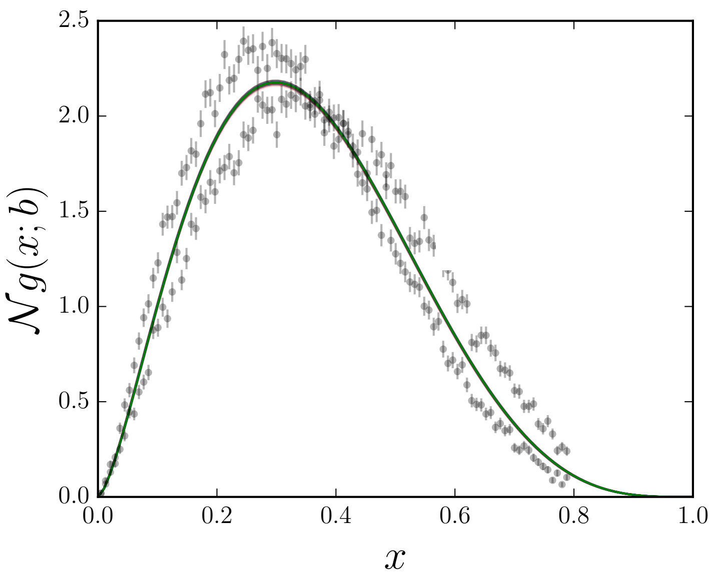


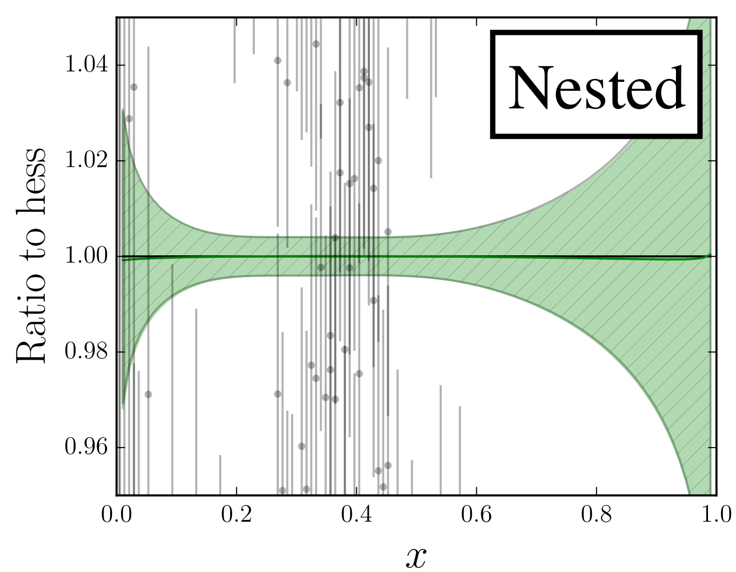
Sato et al. (2017)

MC Error Analysis

- Assuming a single minimum, a Hessian or MC analysis *must* give same results, if using same likelihood function

  $\longrightarrow$ also for discrepant data





$\longrightarrow$ almost identical uncertainty bands for Hessian and for MC!

MC Error Analysis

- Assuming a single minimum, a Hessian or MC analysis *must* give same results, if using same likelihood function

- Approaches that use Hessian + tolerance factor not consistent with Gaussian likelihood function

- NNPDF group claim that within their neural net MC methodology, no need for a tolerance factor, since uncertainties similar to other groups who use Hessian + tolerance

  $\longrightarrow$ how can this be?

- Assuming sufficient observables to determine PDFs, then PDF uncertainties cannot depend on parametrization!

Non-Gaussian likelihood

Incompatible data sets

- Rigorous (Bayesian) way to address incompatible data sets is to use generalization of Gaussian likelihood
  - joint vs. disjoint distributions
  - empirical Bayes
  - hierarchical Bayes
  - others, used in different fields

Disjoint distributions

- Instead of using total likelihood that is a <u>product</u> ("and") of individual likelihoods, $e.g.$ for simple example of two measurements

$$\mathcal{L}(m_1 m_2 | m; \delta m_1 \delta m_2) = \mathcal{L}(m_1 | m; \delta m_1) \times \mathcal{L}(m_2 | m; \delta m_2)$$

use instead <u>sum</u> ("or") of individual likelihoods

$$\mathcal{L}(m_1 m_2 | m; \delta m_1 \delta m_2) = \frac{1}{2}\Big[\mathcal{L}(m_1 | m; \delta m_1) + \mathcal{L}(m_2 | m; \delta m_2)\Big]$$

$\longrightarrow$ gives rather different expectation value and variance

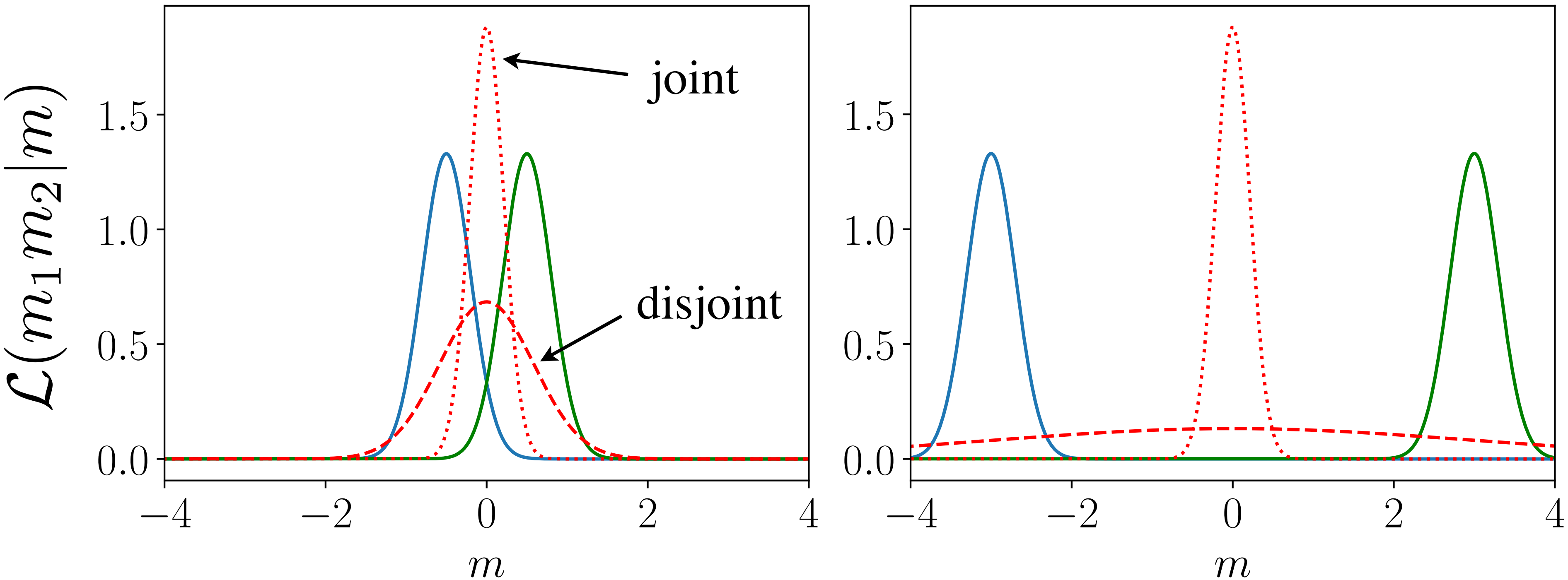
$$E[m] = \frac{1}{2}(m_1 + m_2)$$

$$V[m] = \frac{1}{2}(\delta m_1^2 + \delta m_2^2) + \left(\frac{m_1 - m_2}{2}\right)^2$$

depends on separation!

Disjoint distributions

- Symmetric uncertainties $\delta m_1 = \delta m_2$

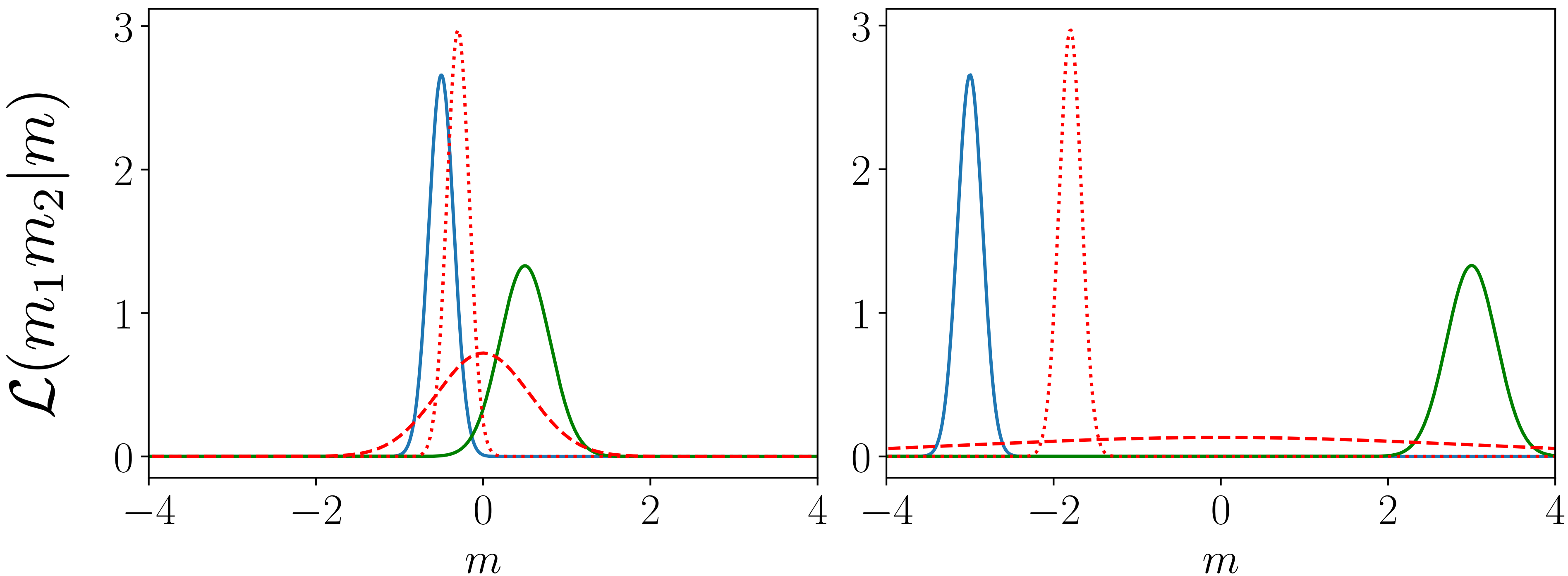


disjoint: $V[m] = \frac{1}{2}(\delta m_1^2 + \delta m_2^2) + \left(\frac{m_1 - m_2}{2}\right)^2$

joint: $V[m] = \frac{\delta m_1^2 \; \delta m_2^2}{\delta m_1^2 + \delta m_2^2}$

Disjoint distributions

- Asymmetric uncertainties $\delta m_1 \neq \delta m_2$



$\longrightarrow$ disjoint likelihood gives broader overall uncertainty, overlapping individual (discrepant) data

Empirical Bayes

- Shortcoming of conventional Bayesian — still assume prior distribution follows specific form (*e.g.* Gaussian)

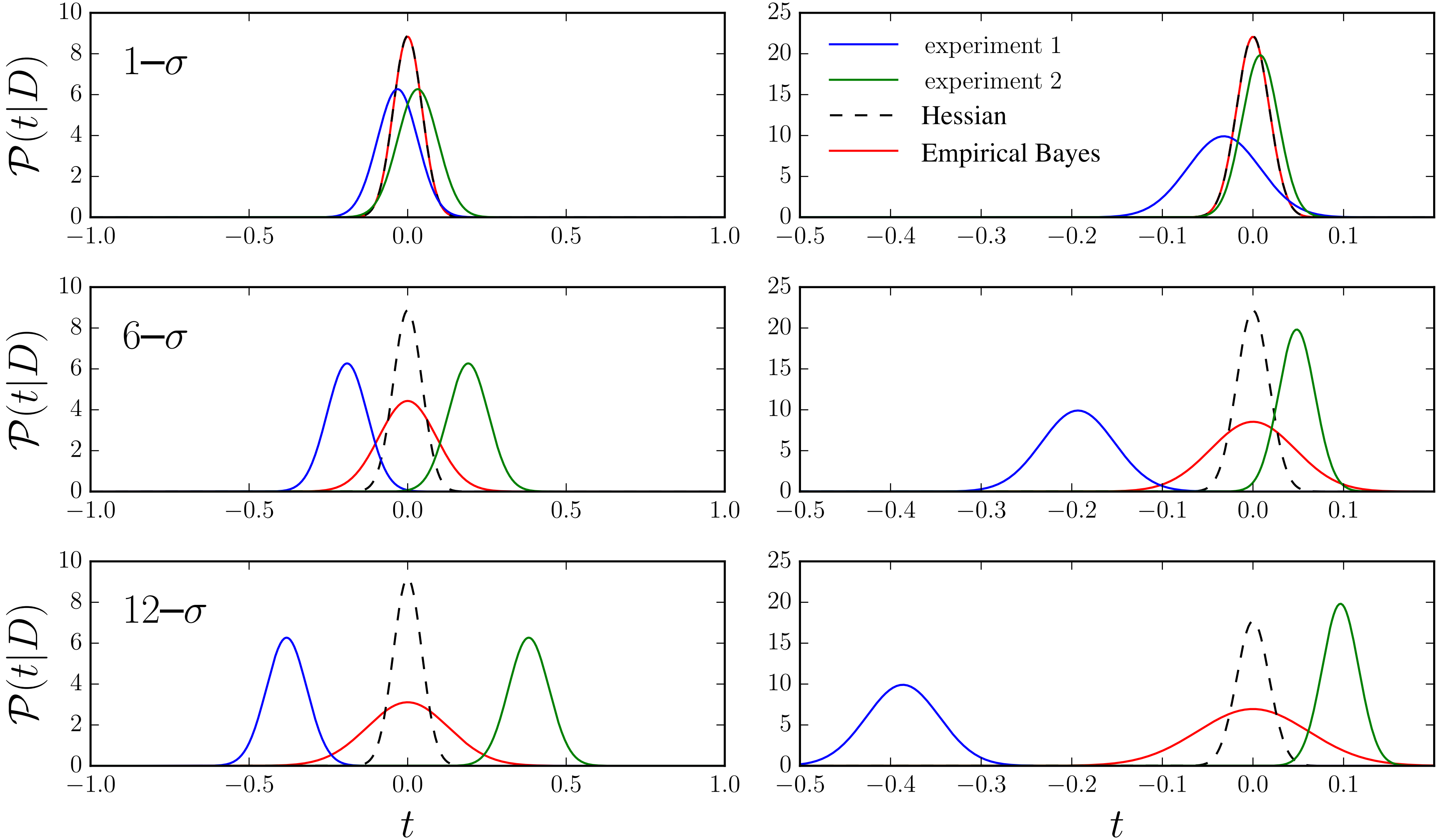
- Extend approach to more fully represent prior uncertainties, with final uncertainties that do not depend on initial choices

- In generalized approach, data uncertainties modified by distortion parameters, whose probability distributions given in terms of "hyperparameters" (or "nuisance parameters")

- Hyperparameters determined from data
  $\longrightarrow$ give posteriors for both PDF and hyperparameters

Empirical Bayes

- Simple example of EB for symmetric & asymmetric errors



Outlook

- SoLID data certainly still needed to probe d/u at high x, free of nuclear corrections!

- New approaches being developed for global QCD analysis
  — simultaneous determination of parton distributions using Monte Carlo sampling of parameter space

- Treatment of discrepant data sets needs serious attention
  — Bayesian perspective has clear merits

- Near−term future: "universal" QCD analysis of all observables sensitive to collinear (unpolarized & polarized) PDFs and FFs

- Longer−term: apply MC technology to global QCD analysis of transverse momentum dependent (TMD) PDFs and FFs

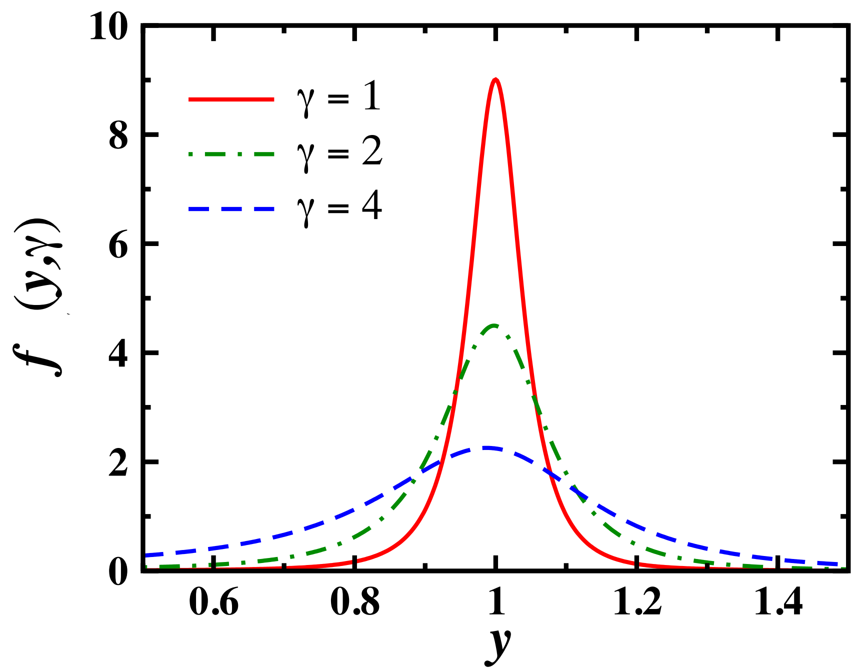
Nuclear corrections

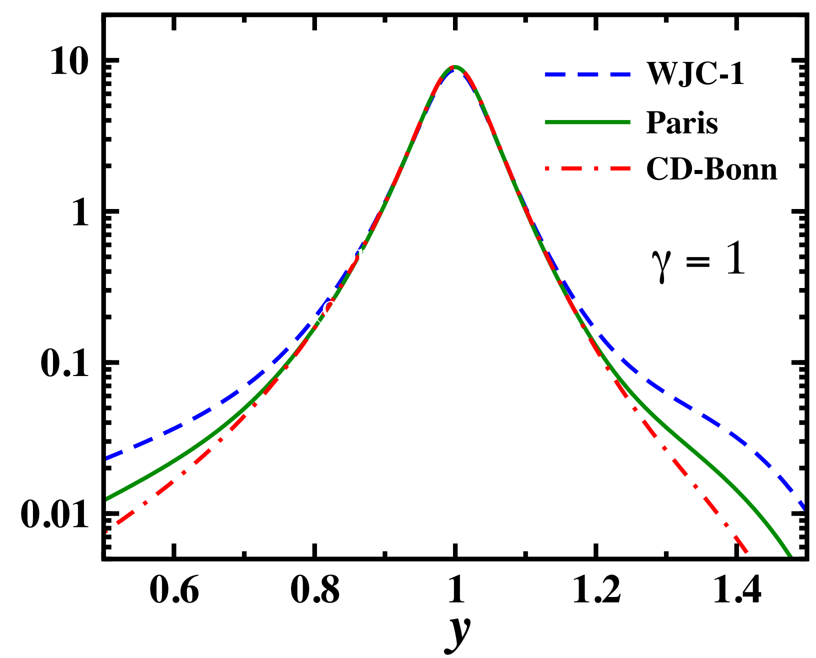
- Smearing function in the deuteron computed in "weak binding approximation" – expand in powers of $\vec{p}^{\,2}/M^2$

$$f(y,\gamma) = \int \frac{d^3p}{(2\pi)^3}\, |\psi_d(p)|^2\, \delta\Big(y - 1 - \frac{\varepsilon + \gamma p_z}{M}\Big) \times \frac{1}{\gamma^2}\Big[1 + \frac{\gamma^2 - 1}{y^2}\Big(1 + \frac{2\varepsilon}{M} + \frac{\vec{p}^{\,2}}{2M^2}(1 - 3\hat{p}_z^2)\Big)\Big]$$

$\psi_d(p) = d$ wave function

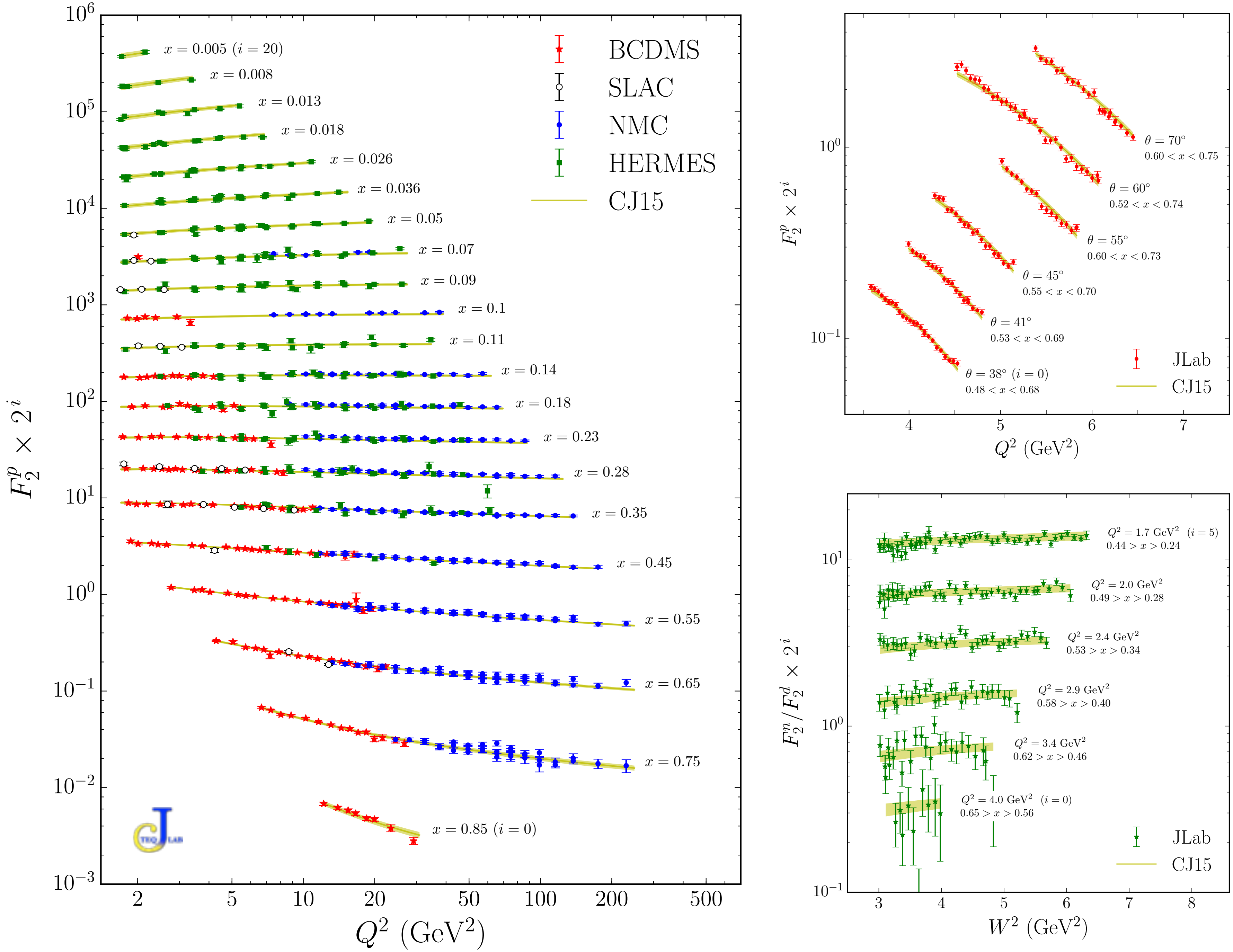
$$\varepsilon = \varepsilon_d - \frac{\vec{p}^{\,2}}{2M}$$





$\longrightarrow$ effectively more smearing for larger x and lower Q^2

$\longrightarrow$ greater wave function dependence at large y $(\longrightarrow$ large $x)$



$\longrightarrow$ excellent description over orders of magnitude in x and Q^2

Empirical Bayes

- Standard mean and variance that characterize data

$$\theta = \mu + \sigma \quad \longrightarrow \quad f(\mu) + g(\sigma)$$

where $f(\mu), g(\sigma)$ are unknown functions that account for faulty measurements

- Simple choice is

$$(\mu, \sigma) \to (\zeta_1\, \mu + \zeta_2,\ \zeta_3\, \sigma)$$

where $\zeta_{1,2,3}$ are distortion parameters, with prob. dists. described by hyperparameters $\phi_{1,2,3}$

- Likelihood function is then

$$\mathcal{L}(\text{data}|\vec{a}, \zeta_{1,2,3}) \ \sim \ \exp\left[-\frac{1}{2}\sum_i \left(\frac{d_1 - f(\mu_i(\vec{a}, \zeta_{1,2}))}{g(\sigma, \zeta_3)}\right)^2\right] \pi_1(\zeta_1|\phi_1)\pi_2(\zeta_1|\phi_2)\pi_3(\zeta_1|\phi_3)$$